



Contents

Editorial

Guest Editorial: Real-Time Image processing with deep neural networks and optimization algorithms
Guest Editorial: Interactive and Innovative Technologies for Smart Education

Papers

- 1075 Blockchain-based Raw Material Shipping with PoC in Hyperledger Composer
Hemraj Saini, Satyabrata Dash, Subhendu Pani, Maria José Sousa, Alvaro Rocha
- 1093 An Innovative Quality Lane Change Evaluation Scheme based on Reliable Crowd-ratings
Konstantinos Psarrafis, Theodoros Anagnostopoulos, Klimis Ntalianis
- 1115 COVID-19 Datasets: A Brief Overview
Ke Sun, Wuyang Li, Vidya Saikrishna, Mehmood Chadhar, Feng Xia
- 1133 Development of Recommendation Systems Using Game Theoretic Techniques
Evangelos Sofikitis, Christos Makris
- 1155 Effective Methods for Email Classification: Is it a Business or Personal Email?
Milena Sošić, Jelena Graovac
- 1177 Re-evaluation of the CNN-based State-of-the-art Crowd-counting Methods with Enhancements
Matija Teršek, Maša Kljun, Peter Peer, Žiga Emeršič
- 1199 A novel Approach for sEMG Gesture Recognition using Resource-constrained Hardware Platforms
Matias J. Micheletto, Carlos I. Chesñevar, Rodrigo M. Santos
- 1213 Fabric-GC: A Blockchain-based Gantt Chart System for Cross-organizational Project Management
Dun Li, Dezhi Han, Benhui Xia, Tien-Hsiung Weng, Arcangelo Castiglione, Kuan-Ching Li
- 1241 Efficient Generative Transfer Learning Framework for the Detection of COVID-19
J. Bhuvana, T. T. Mirmalinee, B. Bharathi, Infant Sneha
- 1261 Federating Digital Contact Tracing using Structured Overlay Networks
Silvia Ghilezan, Simona Kašterović, Luigi Liquori, Bojan Marinković, Zoran Ognjanović, Tamara Stefanović
- 1283 Nearest Close Friend Query in Road-Social Networks
Zijun Chen, Ruoyu Jiang, Wenyuan Liu

Special Section: Real-Time Image processing with deep neural networks and optimization algorithms

- 1305 Crowdsourcing Platform for QoE Evaluation for Cloud Multimedia Services
Asif Ali Laghari, Hui He, Asiya Khan, Rashid Ali Laghari, Shoulin Yin, Jiachi Wang
- 1329 A Novel Motion Recognition Method Based on Improved Two-stream Convolutional Neural Network and Sparse Feature Fusion
Chen Chen
- 1349 A New Frog Leaping Algorithm-oriented Fully Convolutional Neural Network for Dance Motion Object Saliency Detection
Yin Lyu, Chen Zhang
- 1371 A Novel Art Gesture Recognition Model Based on Two Channel Region-Based Convolution Neural Network for Explainable Human-computer Interaction Understanding
Pingping Li, Lu Zhao
- 1389 Adaptive Wavelet Transform Based on Artificial Fish Swarm Optimization and Fuzzy C-means Method for Noisy Image Segmentation
Rui Yang, Dahai Li
- 1409 BiSeNet-oriented Context Attention Model for Image Semantic Segmentation
Lin Teng, Yulong Qiao
- 1427 DRN-SEAM: A Deep Residual Network Based on Squeeze-and-Excitation Attention Mechanism for Motion Recognition in Education
Xinxiang Hua
- 1445 Human Action Recognition Using a Depth Sequence Key-frames Based on Discriminative Collaborative Representation Classifier for Healthcare Analytics
Yuhang Wang, Tao Feng, Yi Zheng
- 1463 A Novel Deep LeNet-5 Convolutional Neural Network Model for Image Recognition
Jingsi Zhang, Xiaosheng Yu, Xiaoliang Lei, Chengdong Wu

Special Section: Interactive and Innovative Technologies for Smart Education

- 1481 Application of Wearable Motion Sensor in Business English Teaching
Dan Lu, Fen Guo
- 1499 Construction of Innovative Thinking Training System for Computer Majors under the Background of New Engineering Subject
Guoxun Zheng, Xiaoxian Zhang, Ruojin Wang, Liang Zhao, Chengming Wang, Chunlai Wang
- 1517 Multimedia Teaching System Based on Art Interaction Technology
Xiaozhong Chen
- 1533 The Research and Implementation Feasibility Analysis of an Intelligent Robot for Simulating Navigational English Dialogue under the Background of Artificial Intelligence
Wei Sun
- 1549 Interactive and Innovative Technologies for Smart Education
Babatomiwa Omonayajo, Fadi Al-Turjman, Nadire Cavus
- 1565 The Impact of Digital Transformation in Teachers' Professional Development During The COVID-19 Pandemic
Ayden Kahraman, Huseyin Bicen
- 1583 Data Mining Technology in the Analysis of College Students' Psychological Problems
Jia Yu, JingJing Lin



Computer Science and Information Systems

Published by ComSIS Consortium

Volume 19, Number 3
September 2022

Volume 19, Number 3, 2022
Novi Sad

Computer Science and Information Systems

ISSN: 1820-0214 (Print) 2406-1018 (Online)

The ComSIS journal is sponsored by:

Ministry of Education, Science and Technological Development of the Republic of Serbia
<http://www.mpn.gov.rs/>



ComSIS Computer Science and Information Systems

AIMS AND SCOPE

Computer Science and Information Systems (ComSIS) is an international refereed journal, published in Serbia. The objective of ComSIS is to communicate important research and development results in the areas of computer science, software engineering, and information systems.

We publish original papers of lasting value covering both theoretical foundations of computer science and commercial, industrial, or educational aspects that provide new insights into design and implementation of software and information systems. In addition to wide-scope regular issues, ComSIS also includes special issues covering specific topics in all areas of computer science and information systems.

ComSIS publishes invited and regular papers in English. Papers that pass a strict reviewing procedure are accepted for publishing. ComSIS is published semiannually.

Indexing Information

ComSIS is covered or selected for coverage in the following:

- Science Citation Index (also known as SciSearch) and Journal Citation Reports / Science Edition by Thomson Reuters, with 2021 two-year impact factor 1.170,
- Computer Science Bibliography, University of Trier (DBLP),
- EMBASE (Elsevier),
- Scopus (Elsevier),
- Summon (Serials Solutions),
- EBSCO bibliographic databases,
- IET bibliographic database Inspec,
- FIZ Karlsruhe bibliographic database io-port,
- Index of Information Systems Journals (Deakin University, Australia),
- Directory of Open Access Journals (DOAJ),
- Google Scholar,
- Journal Bibliometric Report of the Center for Evaluation in Education and Science (CEON/CEES) in cooperation with the National Library of Serbia, for the Serbian Ministry of Education and Science,
- Serbian Citation Index (SCIndeks),
- doiSerbia.

Information for Contributors

The Editors will be pleased to receive contributions from all parts of the world. An electronic version (LaTeX), or three hard-copies of the manuscript written in English, intended for publication and prepared as described in "Manuscript Requirements" (which may be downloaded from <http://www.comsis.org>), along with a cover letter containing the corresponding author's details should be sent to official journal e-mail.

Criteria for Acceptance

Criteria for acceptance will be appropriateness to the field of Journal, as described in the Aims and Scope, taking into account the merit of the content and presentation. The number of pages of submitted articles is limited to 20 (using the appropriate LaTeX template).

Manuscripts will be refereed in the manner customary with scientific journals before being accepted for publication.

Copyright and Use Agreement

All authors are requested to sign the "Transfer of Copyright" agreement before the paper may be published. The copyright transfer covers the exclusive rights to reproduce and distribute the paper, including reprints, photographic reproductions, microform, electronic form, or any other reproductions of similar nature and translations. Authors are responsible for obtaining from the copyright holder permission to reproduce the paper or any part of it, for which copyright exists.

Computer Science and Information Systems

Volume 19, Number 3, September 2022

CONTENTS

Editorial

Guest Editorial: Real-Time Image processing with deep neural networks and optimization algorithms

Guest Editorial: Interactive and Innovative Technologies for Smart Education

Papers

- 1075 Blockchain-based Raw Material Shipping with PoC in Hyperledger Composer**
Composer
Hemraj Saini, Satyabrata Dash, Subhendu Pani, Maria José Sousa, Álvaro Rocha
- 1093 An Innovative Quality Lane Change Evaluation Scheme based on Reliable Crowd-ratings**
Konstantinos Psaraftis, Theodoros Anagnostopoulos, Klimis Ntalianis
- 1115 COVID-19 Datasets: A Brief Overview**
Ke Sun, Wuyang Li, Vidya Saikrishna, Mehmood Chadhar, Feng Xia
- 1133 Development of Recommendation Systems Using Game Theoretic Techniques**
Evangelos Sofikitis, Christos Makris
- 1155 Effective Methods for Email Classification: Is it a Business or Personal Email?**
Milena Šošić, Jelena Graovac
- 1177 Re-evaluation of the CNN-based State-of-the-art Crowd-counting Methods with Enhancements**
Matija Teršek, Maša Kljun, Peter Peer, Žiga Emeršič
- 1199 A novel Approach for sEMG Gesture Recognition using Resource-constrained Hardware Platforms**
Matías J. Micheletto, Carlos I. Chesñevar, Rodrigo M. Santos
- 1213 Fabric-GC: A Blockchain-based Gantt Chart System for Cross-organizational Project Management**
Dun Li, Dezhi Han, Benhui Xia, Tien-Hsiung Weng, Arcangelo Castiglione, Kuan-Ching Li
- 1241 Efficient Generative Transfer Learning Framework for the Detection of COVID-19**
J. Bhuvana, T. T. Mirnalinee, B. Bharathi, Infant Sneha
- 1261 Federating Digital Contact Tracing using Structured Overlay Networks**

Silvia Ghilezan, Simona Kašterović, Luigi Liquori, Bojan Marinković, Zoran Ognjanović, Tamara Stefanović

- 1283 Nearest Close Friend Query in Road-Social Networks**
Zijun Chen, Ruoyu Jiang, Wenyuan Liu

Special Section: Real-Time Image processing with deep neural networks and optimization algorithms

- 1305 Crowdsourcing Platform for QoE Evaluation for Cloud Multimedia Services**
Asif Ali Laghari, Hui He, Asiya Khan, Rashid Ali Laghari, Shoulin Yin, Jiachi Wang
- 1329 A Novel Motion Recognition Method Based on Improved Two-stream Convolutional Neural Network and Sparse Feature Fusion**
Chen Chen
- 1349 A New Frog Leaping Algorithm-oriented Fully Convolutional Neural Network for Dance Motion Object Saliency Detection**
Yin Lyu, Chen Zhang
- 1371 A Novel Art Gesture Recognition Model Based on Two Channel Region-Based Convolution Neural Network for Explainable Human-computer Interaction Understanding**
Pingping Li, Lu Zhao
- 1389 Adaptive Wavelet Transform Based on Artificial Fish Swarm Optimization and Fuzzy C-means Method for Noisy Image Segmentation**
Rui Yang, Dahai Li
- 1409 BiSeNet-oriented Context Attention Model for Image Semantic Segmentation**
Lin Teng, Yulong Qiao
- 1427 DRN-SEAM: A Deep Residual Network Based on Squeeze-and-Excitation Attention Mechanism for Motion Recognition in Education**
Xinxiang Hua
- 1445 Human Action Recognition Using a Depth Sequence Key-frames Based on Discriminative Collaborative Representation Classifier for Healthcare Analytics**
Yuhang Wang, Tao Feng, Yi Zheng
- 1463 A Novel Deep LeNet-5 Convolutional Neural Network Model for Image Recognition**
Jingsi Zhang, Xiaosheng Yu, Xiaoliang Lei, Chengdong Wu

Special Section: Interactive and Innovative Technologies for Smart Education

- 1481 Application of Wearable Motion Sensor in Business English Teaching**
Dan Lu, Fen Guo
- 1499 Construction of Innovative Thinking Training System for Computer Majors under the Background of New Engineering Subject**
Guoxun Zheng, Xiaoxian Zhang, Ruojin Wang, Liang Zhao, Chengming Wang, Chunlai Wang
- 1517 Multimedia Teaching System Based on Art Interaction Technology**
Xiaozhong Chen
- 1533 The Research and Implementation Feasibility Analysis of an Intelligent Robot for Simulating Navigational English Dialogue under the Background of Artificial Intelligence**
Wei Sun
- 1549 Interactive and Innovative Technologies for Smart Education**
Babatomiwa Omonayajo, Fadi Al-Turjman, Nadire Cavus
- 1565 The Impact of Digital Transformation in Teachers' Professional Development During The COVID-19 Pandemic**
Ayden Kahraman, Huseyin Bicen
- 1583 Data Mining Technology in the Analysis of College Students' Psychological Problems**
Jia Yu, JingJing Lin

Editorial

Mirjana Ivanović, Miloš Radovanović, and Vladimir Kurbalija

University of Novi Sad, Faculty of Sciences
Novi Sad, Serbia
{mira,radacha,kurba}@dmi.uns.ac.rs

This third issue Computer Science and Information Systems in 2022 consists of 11 regular articles and two special sections: “Real-Time Image Processing with Deep Neural Networks and Optimization Algorithms” (9 articles) and “Interactive and Innovative Technologies for Smart Education” (7 articles). We are grateful for the hard work and enthusiasm of our authors, reviewers, and guest editors, without whom the current issue and the publication of the journal itself would not have been possible.

In the first regular article, “Blockchain-based Raw Material Shipping with PoC in Hyperledger Composer,” Hemraj Saini et al. discuss the problems facing the shipping process of raw materials from providers to industry end-users via various intermediaries, and propose a framework based on Blockchain that provides integrity and tamper resistance in the shipping chain. A proof-of-concept in Hyperledger Composed is described, together with a performance evaluation.

The second regular article, “An Innovative Quality Lane Change Evaluation Scheme based on Reliable Crowd-ratings” by Konstantinos Psaraftis et al. tackles the problem of mitigating bias and malicious activity in crowdsourced data in the domain of intelligent transportation systems, when no auxiliary information is available at the individual level. The case study involves a crowdsourced database of lane change evaluations, on which the proposed algorithm is applied to negate the noisy ground truth and improve overall quality.

Ke Sun et al., in their article “COVID-19 Datasets: A Brief Overview” organise the numerous COVID-19 available three categories: time-series, knowledge base, and media-based datasets, thus assisting researchers in focusing on methodology rather than the datasets themselves. Problems pertaining to privacy and quality, as well as the potential of COVID-19 datasets are also discussed.

In “Development of Recommendation Systems Using Game Theoretic Techniques,” Evangelos Sofikitis and Christos Makris use game theory in the design of recommender systems on three levels: (1) interaction of the two aspects, query reformulation and relevance estimation, is modeled as a cooperative game where the two players have a common utility, to supply optimal recommendations, (2) three basic recommendation methods developed using the previous approach (collaborative filtering, content based filtering, and demographic filtering) are combined into hybrid systems using game-theoretic techniques, and (3) methods are combined with the use of a genetic algorithm where game theory is used for the parent selection process.

The article “Effective methods for Email Classification: Is it a Business or Personal Email?” by Milena Šošić and Jelena Graovac presents a comprehensive set of experiments has been deep-learning and classical machine-learning algorithms to differentiate between personal and official written e-mail conversations. A notable contribution of the article

is the extraction of a large number of additional lexical, conversational, expressional, emotional, and moral features, which proved to be very useful for the given task.

“Re-evaluation of the CNN-based State-of-The-Art Crowd-counting Methods with Enhancements” authored by Matija Teršek et al., compares five deep-learning-based approaches to crowd counting, reevaluates them, and presents a novel CSRNet-based approach that introduces a Bayesian crowd counting loss function and pixel modeling. The results show that models based on SFA-Net and DM-Count outperform state-of-the-art when trained and evaluated on similar data, and the proposed extended model outperforms the base model with the same backbone when trained and evaluated on significantly different data.

Matías J. Micheletto et al., in “A Novel Approach for sEMG Gesture Recognition Using Resource-Constrained Hardware Platforms” introduce a novel approach for human gesture classification using surface electromyographic sensors (sEMG) in which two different strategies are combined: (1) A technique based on autoencoders is used to perform feature extraction, and (2) Two alternative machine learning algorithms (namely J48 and K*) are then used for the classification stage. Experimental results show that for limited computing power platforms the approach outperforms the alternative methodologies.

“Fabric-GC: A Blockchain-based Gantt Chart System for Cross-organizational Project Management” by Dun Li et al. propose Fabric-GC, a Gantt chart system based on Blockchain which facilitates secure and effective cross-organizational cooperation for project management, providing access control to multiple parties for project visualization. Experimental results show that Fabric-GC achieves stable performance in large-scale request and processing distributed environments, where the data synchronization speed of the consortium chain is several times faster than that of a public chain.

In “Efficient Generative Transfer Learning Framework for the Detection of COVID-19,” J. Bhuvana et al. address the problem of the lack of annotated COVID-19 data by using Deep Convolutional Generative Adversarial Networks (DCGAN) to generate synthetic data, and applying Densenet-201, as well as conventional machine learning approaches such as SVM, Random Forest and Convolutional Neural Network (CNN) to detect COVID-19 from X-ray images. Experiments demonstrate that the proposed transfer learning approach based on DenseNet-201 along with DCGAN based augmentation outperforms the state-of-the-art approaches like ResNet50, CNN, and VGG-16.

Silvia Ghilezan et al., in “Federating Digital Contact Tracing using Structured Overlay Networks,” present a comprehensive, yet simple, extension to the existing systems used for digital contact tracing in the COVID-19 pandemic, which enables those systems, regardless of their underlying protocol, to enhance their sets of traced contacts and improve the global fight against the pandemic during the phase of opening borders.

Finally, “Nearest Close Friend Query in Road-Social Networks” authored by Zijun Chen et al. addresses the realization of nearest close friend queries ($k\ell$ -NCF) in geo-social networks, which aim to find the k nearest user objects from among the ℓ -hop friends of the query user. Existing efforts on $k\ell$ -NCF find the user objects in the Euclidean space, while this article studies this problem on road-social networks. Two methods are proposed, one based on Dijkstra’s algorithm, and the other on IS-Label. Experiments on real and synthetic datasets demonstrate the efficiency of the proposed methods.

Guest Editorial – Real-Time Image processing with deep neural networks and optimization algorithms

Shoulin Yin¹ and Mirjana Ivanović²

¹ Software College, Shenyang Normal University
253 Huanghe Bei Dajie, Huanggu District, Liaoning Province, 110034 China
yslin@hit.edu.cn

² University of Novi Sad, Faculty of Sciences
Novi Sad, Serbia
mira@dmi.uns.ac.rs

Real-time images are often captured and processed without any buffer delays. Since most real-time images are captured from many sources, the quality of the image resolution may vary. However, due to recent advancements in image processing, there are various types of real-time image processing techniques. Real-time image processing may lead to high computational overhead and delays in the transmission of the images, and to overcome these limitations, deep neural networks techniques (DNNs) and optimization algorithms (OAs) may be an asset moving forward. Deep neural networks (DNNs) approach is very popular due to big data support and automatically features selection, this will reduce the workload of scientists, and also convolution neural networks (CNNs) techniques will be used to increase accuracy as compared to machine learning methods. DNNs, like convolution neural networks (CNN), Deep adversarial network (DAN), long short-term memory (LSTM), autoencoder, and deep belief networks have been used to provide real-time image processing.

Using deep neural networks, various hidden layers within them will capture important features of an image or a frame. When the image is captured on a real-time basis, it can be processed by deep neural networks more efficiently and effectively. However, there may be significant performance pressure on the processing and evaluation of real-time high resolution and multi-resolution images. This special section provides an exemplary forum for researchers to discuss theories and ideas associated with real-time image processing using deep neural networks methods and optimization algorithms. Also, this special section discusses all the challenges and limitations of using deep neural networks models in real-time image processing.

This special section aims to receive high-quality papers that extend the current state of the art with innovative ideas and solutions in the broad area of utilization of deep neural networks in real-time image processing. Contributions may present and solve open research problems, integrate efficient novel solutions, present performance evaluations, and compare new methods with existing solutions. Theoretical as well as experimental studies for typical and newly emerging convergence technologies and use cases enabled by recent advances are encouraged. For this special section potential topics that were suggested for potential authors included but were not limited to the following:

- DNNs/OAs-based real-time image processing techniques
- Intelligent learning algorithms for real-time image reconstruction and processing
- Real-time image security and privacy using DNNs/OAs
- Federated learning methodologies used in real-time image processing

- processing of real-time images in remote sensing applications using DNNs/OAs techniques
- Quality of Experience (QoE) and Quality of Service (QoS) for real-time image processing
- DNNs/OAs pattern recognition in real-time image processing and processing
- Evaluation of enhanced real-time images using DNNs/OAs methods
- Computational-based DNNs/OAs models for detection of abnormalities in real-time captured images
- New objective functions of DNNs/OAs for real-time image reconstruction
- Performance analysis of semantic segmentation of images using DNNs/OAs algorithms
- Limitations of DNNs/OAs and hybrid models for real-time image processing
- Sports and arts image processing by DNNs/OAs algorithms
- Machine learning models, big data/cloud computing/fog computing etc, for real-time image processing

This special section received 42 submissions where the corresponding authors were majorly counted by the deadline for manuscript submission with an open call-for-paper period of 6 months. All these submissions are considered significant in the field, but however, only one-third of them passed the pre-screening by guest editors. The qualified papers then went through double-blinded peer review based on a strict and rigorous review policy. After a totally three-round review, 9 papers were accepted for publication. We believe that this special section brings challenging research papers and novel approaches that will be interesting and useful for readers.

A brief overview to the papers in this section can be revealed below, and we expect the content may draw attention from public readers, and furthermore, prompt the society development.

The first paper entitled “Crowdsourcing Platform for QoE Evaluation for Cloud Multimedia Services” by Asif Ali Laghari et al. presents a novel web-based crowdsourcing platform for the assessment of the subjective and objective quality of experience (QoE) of the video service in the cloud-server environment. The user has the option to enter subjective QoE data for video service by filling out a web questionnaire. The objective QoE data of the cloud-server, network condition, and the user device is automatically captured by the crowdsourcing platform. The proposed system collects both objective and subjective QoE simultaneously in real-time. The paper presents the key technologies used in the development of the platform and describes the functional requirements and design ideas of the system in detail. The system collects real-time comprehensive data to enhance the quality of the user experience to provide a valuable reference.

The second paper entitled “A Novel Motion Recognition Method Based on Improved Two-stream Convolutional Neural Network and Sparse Feature Fusion” by Chen Chen et al. proposes a novel motion recognition method based on an improved two-stream convolutional neural network and sparse feature fusion. In the low-rank space, because sparse features can effectively capture the information of motion objects in the video, meanwhile, they supplement the network input data, in view of the lack of information interaction in the network, they fuse the high-level semantic information and low-level detail information to recognize the motions, which makes the performance of the two-stream convolutional neural network have more advantages.

The third paper entitled “A New Frog Leaping Algorithm-oriented Fully Convolutional Neural Network for Dance Motion Object Saliency Detection” by Yin Lyu et al. proposes a new frog leaping algorithm-oriented fully convolutional neural network for dance motion object saliency detection.

The fourth paper entitled “A Novel Art Gesture Recognition Model Based on Two Channel Region-Based Convolution Neural Network for Explainable Human-computer Interaction Understanding” by Pingping Li et al. proposes a novel gesture recognition based on two channel region-based convolution neural network for explainable human-computer interaction understanding. The input gesture image is extracted through two mutually independent channels. The two channels have convolution kernel with different scales, which can extract the features of different scales in the input image, and then carry out feature fusion at the fully connection layer. Finally, it is classified by the softmax classifier. The two-channel convolutional neural network model is proposed to solve the problem of insufficient feature extraction by the convolution kernel. Experimental results of gesture recognition on public data sets NTU and VIVA show that the proposed algorithm can effectively avoid the over-fitting problem of training models and has higher recognition accuracy and stronger robustness than traditional algorithms.

The fifth paper entitled “Adaptive Wavelet Transform Based on Artificial Fish Swarm Optimization and Fuzzy C-means Method for Noisy Image Segmentation” by Rui Yang et al. proposes a noisy image segmentation method based on FCM wavelet domain feature enhancement. Firstly, the noise image is decomposed by two-dimensional wavelet. Secondly, the edge enhancement of the approximate coefficient is carried out, and the artificial fish swarm (AFS) optimization algorithm is used to process the threshold value of the detail coefficient, and the processed coefficient is reconstructed by wavelet transform. Finally, the reconstructed image is segmented by FCM algorithm.

The sixth paper entitled “BiSeNet-oriented context attention model for image semantic segmentation” by Lin Teng et al. proposes a BiSeNet-oriented context attention model for image semantic segmentation. In the BiSeNet, the spatial path is utilized to extract more low-level features to solve the problem of information loss in deep network layers. Context attention mechanism is used to mine high-level implied semantic features of images. Meanwhile, the focus loss is used as the loss function to improve the final segmentation effect by reducing the internal weighting.

The seventh paper entitled “DRN-SEAM: A Deep Residual Network Based on Squeeze-and-Excitation Attention Mechanism for Motion Recognition in Education” by Xinxiang Hua et al. proposes a residual network based on Squeeze-and-Excitation attention mechanism. Deep residual network is widely used in various fields due to the high recognition accuracy.

The eighth paper entitled “Human Action Recognition Using a Depth Sequence Key-frames Based on Discriminative Collaborative Representation Classifier for Healthcare Analytics” by Yuhang Wang et al. proposes a new deep map sequence feature expression method based on discriminative collaborative representation classifier, which highlights the time sequence of human action features. In this paper, the energy field is established according to the shape and action characteristics of human body to obtain the energy information of human body. Then the energy information is projected onto three orthogonal axes to obtain deep spatial-temporal energy map. Meanwhile, in order to solve the problem of high misclassification probability of similar samples by collaborative representa-

tion classifier (CRC), a discriminative CRC (DCRC) is proposed. The classifier takes into account the influence of all training samples and each kind of samples on the collaborative representation coefficient, it obtains the highly discriminative collaborative representation coefficient, and improves the discriminability of similar samples.

The last paper entitled “A Novel Deep LeNet-5 Convolutional Neural Network Model for Image Recognition” by Jingsi Zhang et al. proposes a novel deep LeNet-5 convolutional neural network model for image recognition. On the basis of Lenet-5 model with the guaranteed recognition rate, the network structure is simplified, and the training speed is improved. Meanwhile, they modify the Logarithmic Rectified Linear Unit (L ReLU) of the activation function. The method showed the better effect.

We would also like to thank Prof. Mirjana Ivanović the editor-in chief of ComSIS, for her support during the preparation of this special section in ComSIS journal. Finally, we gratefully acknowledge all the hard work and enthusiasm of authors, reviewers, and core editorial team without whom the special section would not have been possible.

Guest Editorial – Interactive and Innovative Technologies for Smart Education

Fadi Al-Turjman^{1,2} and Mirjana Ivanović³

¹ Artificial Intelligence Engineering Dept., AI and Robotics Institute
Near East University, Nicosia, Mersin 10, Turkey
Fadi.alturjman@neu.edu.tr

² Research Center for AI and IoT, Faculty of Engineering
University of Kyrenia, Kyrenia, Mersin 10, Turkey
Fadi.alturjman@kyrenia.edu.tr

³ University of Novi Sad, Faculty of Sciences
Novi Sad, Serbia
mira@dmi.uns.ac.rs

Smart education is the key technology that facilitates a major shift from the traditional practice of face-to-face learning methodologies into distance learning paradigms. Due to its distributed nature, a broad range of disruptive technologies such as cloud computing, virtual reality, and augmented reality have been increasingly adopted to enhance the process of the smart education system. As a matter of fact, communication forms a vital part of the educational forum, which becomes extremely complex when dealing with online learning paradigms such as smart education. It has been a long way that smart education has come into existence, but it still needs to improve its interface and interactive methodologies to leverage its full potential towards society. In the present scenario, dramatic improvements in e-learning methodologies such as virtual learning and online interactive learning solve typical intractable problems and offer sophisticated learning experiences to the end-users in a more convenient way. Virtual classrooms have attracted increased attention in recent times as it provides the most reasonable educational services. It efficiently visualizes the data from hundreds of computer networks simultaneously, enabling the users to offer the most relevant and actionable response to the virtual environment using headphones. In addition, the use of advanced network technologies such as 5G makes this process take place in a much faster way, with much more data flow facilities. In summary, smart education in the digital era of computing technology is actually a cool upcoming trend, where the vast amount of individuals from the educational society may unimaginably gain benefit from this learning method. However, emphasizing the need for interactive and innovative technologies for smart education also acquire greater importance as it forms the key requirement of the present-day smart education environment.

As a result, this special section aims to explore more deep insights into interactive and innovative technologies for smart education from various perspectives of the teaching and learning methodologies. It consists of seven articles from different areas of smart education selected among more than 30 submitted papers. Each paper was reviewed by three reviewers. We are grateful for the hard work and enthusiasm of authors and reviewers, without which the current special section would not have been possible.

The first article, “Application of Wearable Motion Sensor in Business English Teaching” by Fen Guo conducts an empirical analysis on motion sensors application in business English teaching. It mainly collects speech information through wearable motion sensor,

analyzes the reading correctness of students through speech recognition, so as to promote students to learn business English better. Firstly, the wearable sensor is used to collect and preprocess the speech information of students' business English reading as the input of speech recognition. Secondly, the linear predictive cepstrum coefficient (LPCC) and Meier frequency cepstrum coefficient (MFCC) of students' business English reading speech are extracted, and the mixed parameters of LPCC and MFCC are taken as speech features. Finally, the correctness of reading speech is recognized by combining HMM and WNN. Through the simulation analysis of students' reading speech recognition, it is shown that the speech recognition based on wearable motion sensor is feasible and the recognition method has good performance. In addition, the feasibility of wearable motion sensor in business English teaching is verified by the establishment of experimental class, which can promote students' English learning better.

The second article titled "Construction of Innovative Thinking Training System for Computer Majors under the Background of New Engineering Subject" by Guoxun Zheng, Xiaoxian Zhang, Ruojin Wang, Liang Zhao, Chengming Wang, Chunlai Wang is focused on cultivation of talents in computer major and cooperation between industry and university. Computer major has trained a large number of computer related talents for the society. The graduates of this major are an important force of social development, and also make a significant contribution to the development of the national economy. Paying attention to the new demand of social development for high-quality computer talents, targeted training is the key to the development of scientific and technological innovation. Firstly, this article points out the main problems affecting the cultivation of talents in this major. Then, based on the basic idea of a new engineering subject, it discusses how to renew the basic educational concept of computer major, strengthen the cooperation between industry and university, reform according to the requirements of new engineering subject, and realize incremental optimization, stock adjustment and cross-integration from various aspects.

In the third article with the title "Multimedia Teaching System Based on Art Interaction Technology" by Xiaozhong Chen aim is at improving the Massive Open Online Course (MOOC) platform, which is the largest application of hybrid learning. It integrates animation technology and multimedia technology, and designs a multimedia-teaching platform based on art interaction technology, which effectively improves the attraction of MOOC platform to learners. Firstly, this paper introduces multimedia, animation and interactive animation technologies in detail, and applies them to MOOC platform. Secondly, according to the analysis of the research results of teaching platform requirements, the design principles and system framework of this paper are given. Finally, the information processing system of B/S architecture mode is built to make the improved platform have high response speed and data processing ability. In addition, this paper constructs a small-scale multimedia hybrid learning platform for testing and finds that the multimedia teaching platform based on art interactive technology designed in this paper can well promote students' autonomous learning and improve the effect of students' learning.

The fourth article titled "The Research and Implementation Feasibility Analysis of an Intelligent Robot for Simulating Navigational English Dialogue Under the Background of Artificial Intelligence" by Wei Sun uses the test data set to test the analytical model of navigational English dialogue instructions. The experimental results show that the conditional random field (CRF) + domain dictionary + ambiguity resolution method has the

highest segmentation effect. The calculated percentages of the analytical model are correct rate: 76.85%; recall rate: 80.36%; F-value: 88.46%. This paper implements a robot teaching and reproduction method based on simulated navigational English conversation and human-computer interaction under the background of artificial intelligence, and designs robot motion realization experiments and speech recognition experiments. The three-dimensional error after fine-tuning the voice is between 1.6798mm and 2.9968mm. This article constructs a simulation navigational English dialogue robot system. The FAQ component has up to 79.2%; others have a lower accuracy rate of only 59.03%.

In the fifth article with the title “Interactive and Innovative Technologies for Smart Education” by Babatomiwa Abdulazeez Omonayajo, Fadi Al-Turjman, Nadire Cavus the focus is on new concepts and ideas that have been recently emerged in the process of obtaining and disseminating cognitive, ethical, and public knowledge. In the current state of education, a learner, tutor, and the knowledge being transferred are all present, and smart education has made learning more flexible and convenient. This concept is accomplished through the use of smart devices and technologies that are interconnected to access digital resources. Smart education refers to a new way of learning that has gotten a lot of attention, notably during the 2020 COVID-19 Pandemic. This article examines the technologies that have aided smart education in achieving its educational goals. With smart technological solutions, modern technologies are enhancing the teaching - learning process in today’s education. Smart education, with the help of sophisticated technology, simplifies the activities of teaching, learning, networking, and cooperating, as well as making speedy alerts more productive.

The next article titled “The Impact of Digital Transformation in Teachers’ Professional Development During the COVID-19 Pandemic” by Ayden Kahraman, Huseyin Bicen presents a study conducted to reveal the positive and negative aspects of professional development programs applied to teachers involved in distance education during the COVID-19 process and investigate whether they contributed to the digital transformation with the competencies they acquired through these programs. A total of 30 teachers participated in the study voluntarily. The research has been carried out as a case study. In order to ensure the validity, reliability, and consistency of the data, the mixed research method consisting of the qualitative and quantitative phases has been used for acquiring data. Once the teacher-oriented professional development program was completed, the teachers were subjected to an achievement test and a self-assessment questionnaire. A focus group interview was conducted to collect various views of 18 teachers regarding the program. This study also reveals that teacher-oriented professional development programs can be applied efficiently through online education and have a crucial role in strengthening and enhancing the technical competencies of the teachers involved in distance education.

The last article titled “Data mining technology in the analysis of college students’ psychological problems”, by Jia Yu, and JingJing Lin expounds on the research status of data mining and the status quo of college students’ psychological health problems, deeply analyzing the feasibility of introducing data mining technology into the analysis of college students’ psychological health. After studying and analyzing the decision tree technology of data mining, and taking the psychological health problem data of the students in a university in 2021 as the research object, this paper uses the decision tree to analyze the psychological health problem data. The main work includes the following: determining the mining object and mining target; preprocessing the original data; and according to

the characteristics of the data used, choosing the decision tree algorithm to construct the decision tree of the students. Finally, based on the analysis and comparison of the decision tree model before and after pruning, classification rules are extracted from the optimal decision tree model, thus providing a scientific decision-making basis for mental health education in colleges and universities.

Acknowledgments. The guest editors are thankful to the anonymous reviewers for their effort in reviewing the manuscripts. We are also thankful to the Edit-in-Chief, Prof. Mirjana Ivanović for the supportive guidance during the entire process.

Prof. Dr. Fadi Al-Turjman received his Ph.D. in computer science from Queen's University, Canada, in 2011. He is the associate dean for research and the founding director of the International Research Center for AI and IoT at Near East University, Nicosia, Cyprus. Prof. Al-Turjman is the head of Artificial Intelligence Engineering Dept., and a leading authority in the areas of smart/intelligent IoT systems, wireless, and mobile networks' architectures, protocols, deployments, and performance evaluation in Artificial Intelligence of Things (AIoT). His publication history spans over 400 SCI/E publications, and more than 11000 citations, in addition to numerous keynotes and plenary talks at flagship venues. He has authored and edited more than 40 books about cognition, security, and wireless sensor networks' deployments in smart IoT environments, which have been published by well-reputed publishers such as Taylor and Francis, Elsevier, IET, and Springer. He has received several recognitions and best papers' awards at top international conferences. He also received the prestigious Best Research Paper Award from Elsevier Computer Communications Journal for the period 2015-2018, in addition to the Top Researcher Award for 2018 at Antalya Bilim University, Turkey and the Lifetime Golden-award of Dr. Suat Günsel from Near East University, Cyprus in 2022. Prof. Al-Turjman has led a number of international symposia and workshops in flagship communication society conferences. Currently, he serves as book series editor and the lead guest/associate editor for several top tier journals, including the IEEE Communications Surveys and Tutorials (IF 23.9) and the Elsevier Sustainable Cities and Society (IF 7.58), in addition to organizing international conferences and symposiums on the most up to date research topics in AI and IoT.

Mirjana Ivanović holds the position of Full Professor at Faculty of Sciences, University of Novi Sad, Serbia. She is a member of National Scientific Committee for Electronics, Telecommunication and Informatics within Ministry of Education, Science and Technological Development, Republic of Serbia. She was a member of University Council for Informatics for more than 12 years. Prof. Ivanovic is author or co-author of 14 textbooks, several international monographs and more than 450 research papers, most of which are published in international journals and conferences. Her research interests include agent technologies, intelligent techniques, applications of data mining and machine learning techniques in medical domains and technology enhanced learning. She is member of Program Committees of more than 300 international conferences, Program/General Chair of several international conferences, and leader of numerous international research projects. Mirjana Ivanovic delivered several keynote speeches at international conferences and visited numerous academic institutions all over the world as visiting researcher (Germany, Slovenia, Australia, China, Korea). Currently she is Editor-in-Chief of the Computer Science and Information Systems journal.

Blockchain-based Raw Material Shipping with PoC in Hyperledger Composer

Hemraj Saini¹, Satyabrata Dash², Subhendu Kumar Pani³, Maria José Sousa⁴ and Álvaro Rocha⁵

¹ School of Computing

DIT University, Dehradun-248009, India

hemraj1977@yahoo.co.in

² Department of Computer Science and Engineering

Ramachandra College of Engineering, Eluru, Andhra Pradesh, India

dash_satyabrata@yahoo.co.in

³ Krupajal Engineering College, Bhubaneswar, Odisha, India

skpani.india@gmail.com

⁴ Instituto Universitário de Lisboa, Portugal

maria.jose.sousa@iscte-iul.pt

⁵ ISEG - Universidade de Lisboa, Portugal

amrrocha@gmail.com

Abstract. In today's world, a lot of various kinds of raw materials are shipped from one place to another as per the requirement of industries. This shipping process involved many multiple levels with multiple personalities or authorities. The intermediates may be influenced by some illegal external factors and there may be some theft or modification in the raw material which is in the shipping process. This generates a significant loss if the material is of high cost. Presently, the advancements in information technology precede a method to restrict this loss and it is blockchain. Blockchain technology is an essential feature in enabling a comprehensive view of events back to origination. The shipping chain of raw materials that provides integrity and tamper resistance for raw materials in the shipping process is proposed in the manuscript. we have also provided proof of concept (PoC) in Hyperledger Composer with performance evaluation.

Keywords: Blockchain, SCRM, PoC, Hyperledger, Raw Material.

1. Introduction

Presently, the ever-growing industrial world needs a rapid consumption of various kinds of raw materials. In the shipping process, this raw material can be modified or partially theft or mixing can be possible at some level during the shipping process. Therefore, the digital evidence of such illegal activities is significant during the investigation. In the investigation process, the integrity and originality of the evidence are ensured throughout the life cycle of the investigations [3], [17]. The multilevel life cycle of shipping raw materials involved multiple personalities with multiple identities. If there is an unwanted activity with the raw materials during the shipping process by any of them due to some influencing personalities or factors then with the physical evidence it is difficult to identify

the culprit. In such cases, digital evidence will play an important role, so, by completing the investigation process, this digital evidence must be secured.

The shipping chain of raw materials (SCRM) is defined as a process to maintain the integrity of raw materials and keep all the documentation of interaction or raw materials with different identities and their work with all history in digital form. SCRM plays an important role in Digital investigation, if requires, as it maintains all minute-by-minute records of events in digital documentation. Different kinds of raw materials pass through different hierarchies starting from mining or production to delivery and verification of receiving, and in this process, digital records are the important weapons to record the whole process. SCRM logs the information such as how, when and what digital evidence of events are recorded and preserved transparently. This evidence will be important in any kind of investigation if needed.

A system that guarantees the four facts like transparency [2], [6], authentication [23], [18], security [12], [1] and auditability [7], [14] for ensuring the integrity of recorded digital events in blockchain and it is achieved by SCRM. As per the Gartner Hype Cycle for Supply Chain Strategy, 2020, as shown in Figure 1 there are a healthy number of capabilities to the left of the peak, which reflects the many emerging capabilities that supply chain organizations are exploring. On the right are the capabilities that companies should be actively adopting at scale to optimize their performance. So, blockchain is also aligned to potentially fulfil critical and long-standing challenges presented across dynamic and complex global supply chains that traditionally have held centralized governance models. Current capabilities offered by blockchain solutions for supply chains include a loose portfolio of technologies and processes that spans middleware, database, verification, security, analytics, and contractual and identity management concepts. Blockchain is a linked representation of blocks and a block contains immutable information of the events. In this way, it generates a distributed system of linked representation. Using SCRM a real-time audit can be performed without bias and with accuracy.

In the case of shipping raw materials, there are three major concerns including document workflow management, financial processes, and device connectivity. These concerns are having a problem of adaption due to lack of trust, information security, miscommunication, different accounting system usage, and various digitized formats. With the different processes involved in the above-mentioned activities, there is a fair possibility of biasing unknowingly or knowingly which leads to a significant loss in the process of shipping raw materials. Besides, it is also difficult to track back the transactions in the traditional systems of shipping raw materials.

Therefore, to overcome the above-mentioned problem it is quite necessary to identify an automated solution which maintains all the artefacts during the whole shipping process. This record keeping should be secure, transparent and scalable and it is SCRM chain proposed in the paper.

The rest of the paper is written in different other sections including section 2 gives a brief idea about the process of shipping raw materials, tools required to generate blockchain, discussion of the feasibility test of blockchain for shipping raw materials. Section 3 provides motivation for the work. Section 4 presents the proposed shipping chain of raw material with PoC in Hyperledger Composer. Section 5 illustrates the performance analysis of the proposed blockchain framework. Section 6 gives the conclusion of the paper and at the end, a list of references is provided.

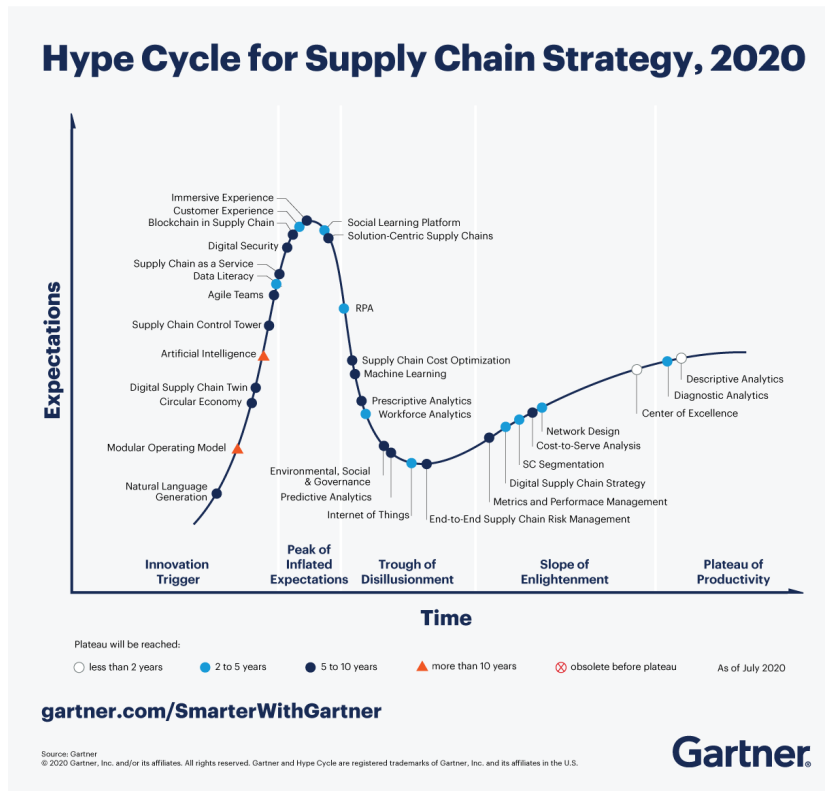


Fig. 1. Gartner Hype Cycle for Supply Chain Strategy, 2020

2. Shipping Process of Raw Materials

The shipping process includes many actors throughout the process like Importer, Exporter, Bank, Insurance Company, Freight Forwarder, Shipping Company, Customs House Agent (CHA), Customs Authorities, Port Authorities, and Intermodal Transport Providers. Information exchange happens when the raw material moves from one actor to another. In this case, willful illegal information exchange may lead to some morphing in raw material. Therefore, a secure chain of information exchange has to be maintained and that has to be immutable and transparent to all the actors. In this paper, this chain is proposed by referring to SCRM. As SCRM is a blockchain-based chain and it has all the information exchange records in the form of blocks which are immutable, secure and transparent to all the actors involved in the process [9], [15]. In addition, all the digital records remain integrated in SCRM. Figure 2 depicts the detailed shipping process of raw materials starting from the origin to the destination after including all the actors involved in the whole process. Further, the shipping process is depicted in figure 3 in the form of SCRM, where recordkeeping of all the steps including identifying the raw material to reporting at different actors is displayed.

Hyperledger Composer [5] in an open blockchain application development environ-

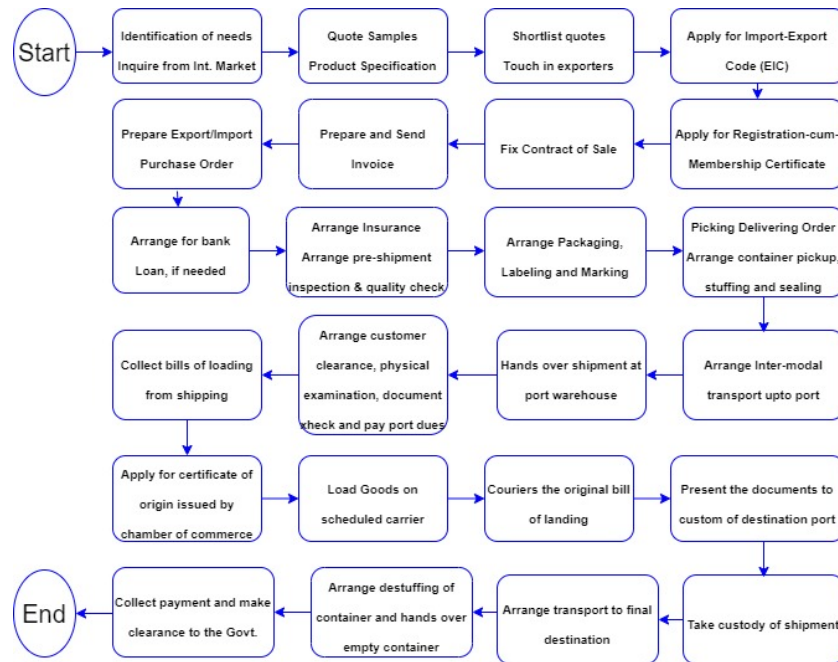


Fig. 2. Shipping Process

ment and toolset. It is used to develop blockchain applications rapidly. Integration of data with blockchain applications can also be done by the use of Hyperledger Composer. Blockchain business application networks contain assets, participants, and transactions across multiple blockchain networks that have to be recorded by the Hyperledger Composer. The transaction can be done by various stakeholders as per their roles such as buyers, sellers, and viewers in the blockchain. After doing registration of these parties in the blockchain they can do the transaction. Figure 4 shows a systematic framework of Hyperledger composer framework.

Hyperledger composer framework works in three layers [10]. In the first layer, the composer is used to create a business network definition comprised of the model, script ACL, and query files. The second layer packages up the business network definition and export it as an archive which is ready to deploy anywhere. Finally, the third layer uses ID cards to deploy business network definitions to a distributed ledger.

2.1. Tool for performance evaluation of blockchain

The performance of the blockchain is major concern before it is deployed anywhere and for this, it needs a tool. Here, we have used Hyperledger Caliper [8] to evaluate the performance of our proposed blockchain SCRM with predefined use cases. Hyperledger Caliper provides a number of reports for different performance indicators including time per transaction (tps), latency during transactions, utilization of a resource etc. On the basis of these reports, the suitability of the blockchain is decided as per the user's requirement.

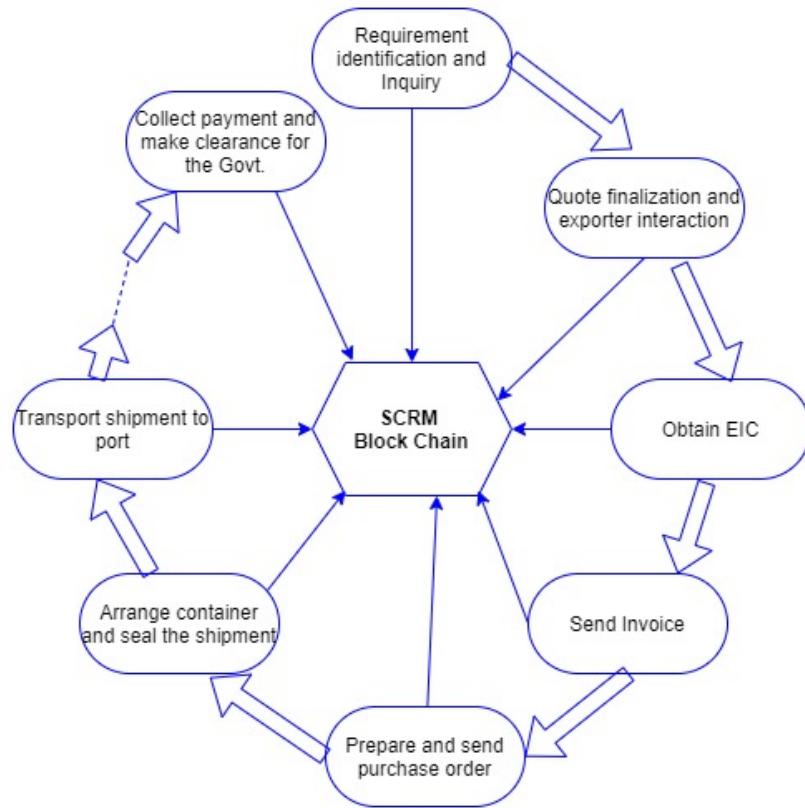


Fig. 3. Process model for SCRM blockchain for raw material shipment

2.2. Feasibility test of blockchain for shipping raw material

The feasibility of the blockchain solution can be decided on the basis of four facts given below-

- Is a shared database required to store the data and state of a transaction? If it is not then the blockchain solution is not feasible.
- Are there multiple users who want to write in the blockchain? If it is not then the blockchain solution is not feasible.
- Is there a high degree of trust among the users? If it is then the blockchain solution is not feasible.
- Is there a third Trusted Party (TTP)? If it is then the blockchain solution is not feasible.

Figure 5 depicts the flowchart for testing the feasibility study of our proposed blockchain i.e. SCRM. In raw material shipping, all these characteristics are satisfied and hence a blockchain solution is feasible.

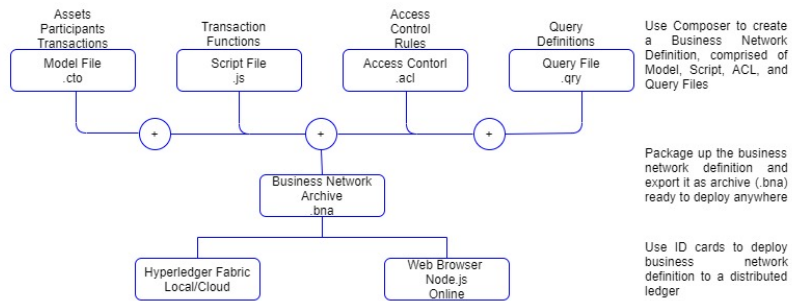


Fig. 4. Framework of Hyperledger Composer

3. Proposed Model

The selection of a blockchain development environment and toolset is an important step before implementing a blockchain application. There are two major options out of many other options that are blockchain implementation using the Ethereum network, and Hyperledger Fabric.

In the case of Ethereum, network anonymity is at the principal edge and in the Ethereum blockchain, anyone can do the transaction but there is no way to identify who has done it. This is not in our favour as we want to track back the history to know who is involved in a transaction. However, Hyperledger Composer is providing all the requirements to develop a blockchain-based application with all robust, transparent and secure records. Therefore, we opted to use Hyperledger Composer for implementing our proposed SCRM system. In the Hyperledger Composer, a functional flow of the SCRM is created as depicted in figure 6. In this, the whole functionality of SCRM is divided into four sections including participants, core modules, blockchain modules, and distributed storage.

Participants: They are the real stakeholders in the raw material shipping process. They can change the state of SCRM by doing transaction-related activities. They are authorized persons in the SCRM for doing transactions and all the details of the transaction by any participant are recorded transparently, securely, and with integration in SCRM.

Core Modules: Core modules are responsible for managing the communication among participants and our proposed blockchain i.e. SCRM. A call of appropriate core module participants can get details of the transaction or submit a transaction.

Blockchain Module: In the case of the blockchain module there are P2P networks and consensus protocols to control the communication through the P2P network.

Distributed Transaction Store: It is the combination of distributed database and a module responsible for the authorization and authentication of participants. All the transaction details are distributed and stored in this module. When any participant wants the details they can get these details after authentication.

Front End: Front end is responsible to provide an interactive interface for participants to our proposed blockchain i.e. SCRM. This interface is created by using composer-REST-server and Yeoman Framework of Hyperledger Composer. Composer-REST-server

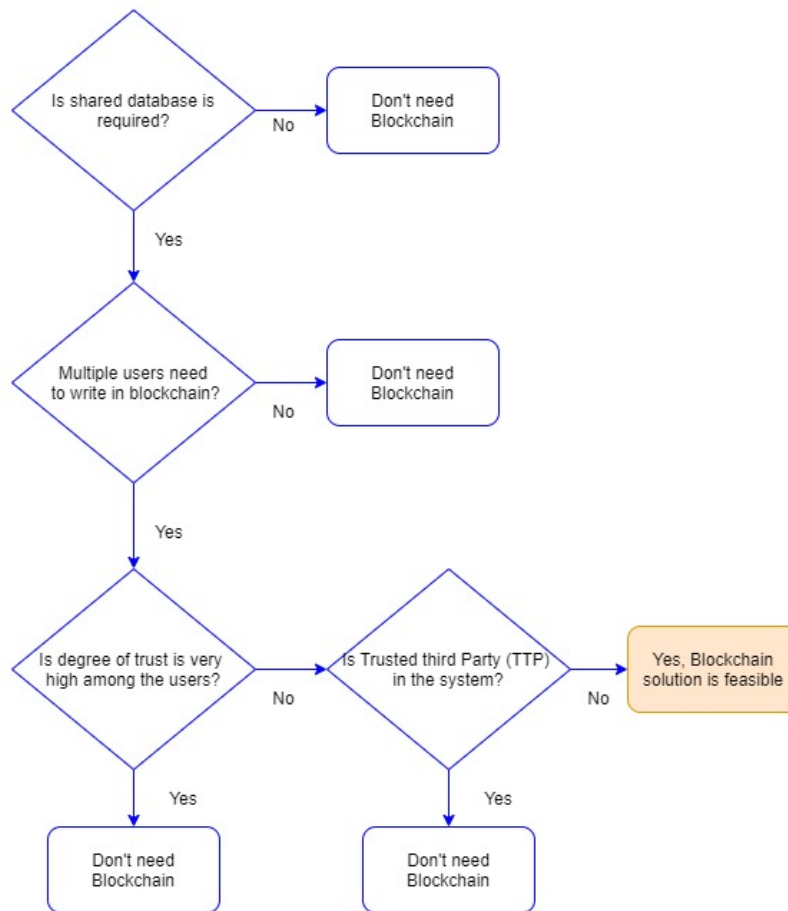


Fig. 5. Framework Flowchart for the feasibility of Blockchain

is responsible to generate the RESTful APIs for all the entities of the SCRM including assets, participants, and transitions. On the other hand, the Yeoman Framework is responsible to generate LoopBack Application. Figure 7 is depicted the Communication of the SCRM blockchain with participants through the front end.

3.1. Proof-of-Concepts

Assets, participants, and transactions are the major components to describe the blockchain applications [20], [4]. Instances of these components are stored in the blockchain. In the case of SCRM their components are described below.

Participants: Suppliers, Buyers, Manufacturers, Distributors, Transporters, Govt. authorities, and Customers are the significant participants of SCRM. Here, the administrator is not a modelled participant as it has only a Composer identity and will not invoke a transaction.

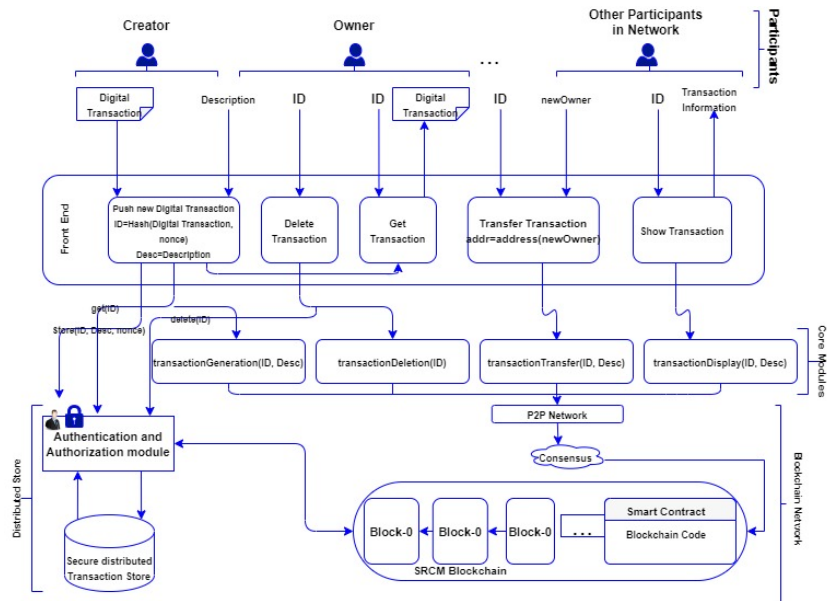


Fig. 6. Functional Flow of SCRM Blockchain

Assets: Any object having value can be an asset. These assets can be shared and transacted in the blockchain. In our case, raw materials, containers, warehouses, etc. are the tangible assets and workforce, securities, IPRs, or any referral documents are intangible assets. Specifically, the data records, used throughout the shipping process, are major assets which are used to track back the history. A model file is used to model the participants. New instances can be created and registered in the registry file. Identities of the participants are also recorded in the identity registry file. Who so ever participates is involved in any of the processes, a record of that task with the performer’s identity has to be recorded by the organization of the blockchain consortium.

Transactions: Action on assets done by participants is known as a transaction. It can be in different forms like- the creation of transactions, updating a transaction, deletion of transactions, and transfer of a saved transaction on the network. A data structure is used to define an SCRM transaction as shown below:

```

struct transaction contains
string/bytes32 transactionID;
address creator;
address owner;
string transactionDescription;
unit caseID;
address transferChain[];
dateTime transferTime[];
//other possible options
    
```

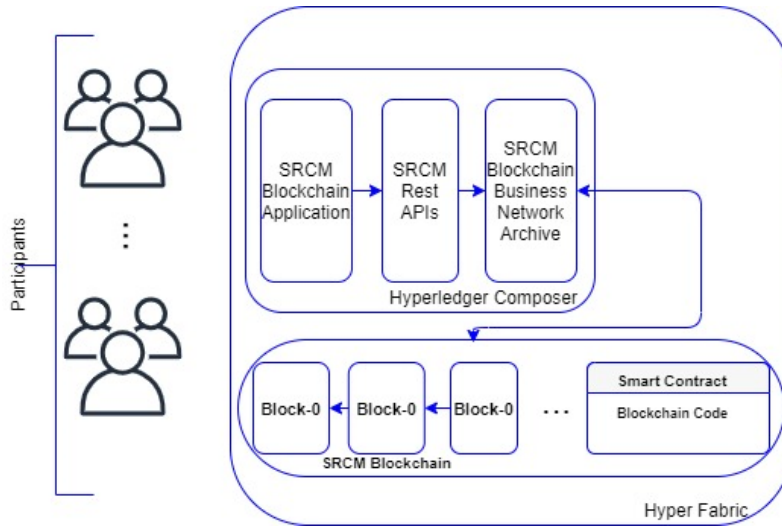


Fig. 7. Communication of SCRM blockchain with participants

3.2. Description of Terminology used

We have used different terminologies in the process and it is mentioned in the following table 1 along with its brief description.

Table 1. Terminology used

Terminology	Brief description
transactionID	It is a unique random number generated by SHA256 for every digital transaction.
creator	It is the participant who inserted the transaction in the first place.
owner	It is the participant who possesses the transaction presently.
transactionDescription	It contains the necessary attributes to define a transaction.
caseID	Unique number for material shipment to initialize the smart contract.
transferChain	It contains the address of the owner of the transaction and it is an array.
transferTime	It contains the date and time of the transaction and it is an array.

The SCRM model has essential functionalities including transaction creation, transaction update, transaction deletion, and transaction display from the blockchain. These functionalities are triggered by the participants of the SCRM. Rules to access a particular functionality by a particular participant are mentioned in permissions.acl file of SCRM blockchain of Hyperledger composer. Different classes are used in the implementation of the SCRM and their script segments are shown below:

```

Script Segment#1: Transaction class
Transaction class {
  "$class": \org.example.basic.Transaction", \ID":
    \0xPQRSTUVWXYZ",
  "creator": \resource : org.example.basic.
    Participants#9876",
  "owner": \resource : org.example.basic.
    Participants#9876",
  "Description": \Necessary information
    describe to transaction",
  "transferChain": [\resource : org.example.
    basic.Participants#9876"],
  "timeChain": [\2020-12-30 T15:03:756z"]
}
Script Segment#2: Participant class
Participant class {
  "$class": \org.example.basic.Participants",
  "ParticipantID": \9876",
  "firstName": \XYX",
  "lastName": \PQR"
}
Script Segment#3: Transaction Transfer class
Transaction Transfer class {
  "$class": \org.example.basic.TransactionTransfer",
  "ID": \resource : org.example.basic.
    Transaction#0xPQRSTUVWXYZ",
  "newOwner": \resource : org.example.basic.
    Participants#6789",
  "transactionID":
    \34c3456-c356-8634-b7d7-c78b56a56d98",
  "timestamp": \2020-12-30 T15:12:755z"
}
Script Segment#4:
Transaction state after several transfers
between participants class {
  "$class": \org.example.basic.Transaction",
  "ID": \0xPQRSTUVWXYZ",
  "creator": \resource : org.example.basic.
    Participants#9876",
  "owner": \resource : org.example.basic.
    Participants#9876",
  "Description": \Necessary information
    describe to transaction",
  "transferChain": [
    "resource : org.example.basic.Participants#9876",
    "resource : org.example.basic.Participants#8769",
    "resource : org.example.basic.Participants#7698",
    "resource : org.example.basic.Participants#6987",
  ],
  "timeChain": [
    "2020-12-30 T15:12:852z"
    "2020-12-30 T17:12:955z"
    "2020-12-30 T18:12:457z"
    "2020-12-30 T15:12:753z"
  ]
}
}

```

This permissioned SCRM blockchain is built on the Hyperledger Composer and runs under the controlled environment governed by the consortium.

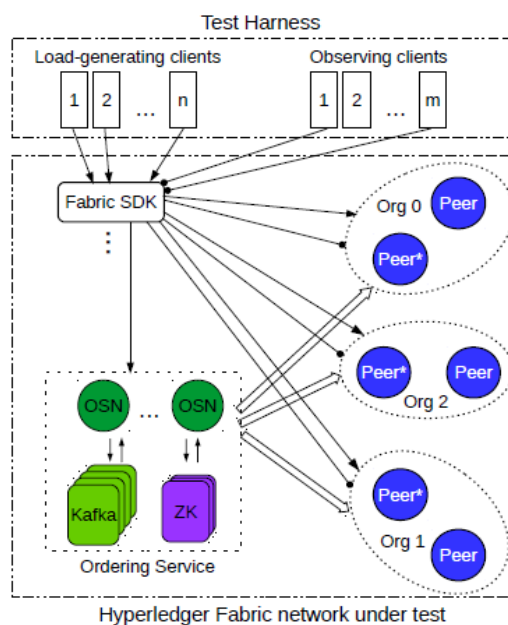


Fig. 8. Framework for Hyperledger Fabric VI performance evaluation [10]

4. Performance Evaluation

Due to the specific characteristics of the blockchain, necessary performance metrics has to be applicable to it. The central authority is not present in the blockchain and hence data or transaction is replicated for different peers of various organizations at different times. Instead of a single transaction, a block of transactions is committed after a consensus process [19], [11] and committed. The consensus process included all the blockchain nodes and until or unless a positive response is not achieved at least by two third of the nodes the transaction will not be committed. This process is dependent on the input traffic i.e. how many transactions are there in the queue for the commit and even on the hardware efficacy of the blockchain miner, therefore, it is uncertain to determine the commit time [13].

4.1. Experimental setup

Hyperledger Fabric VI blockchain setup is used to evaluate the performance of SCRM and it is depicted in figure 8. In this, the test harness is the collection of nodes of two categories including Load-generating clients and Observing clients. Load-generating clients

are responsible to submit the transactional and Observing clients are responsible to do the queries to know the state of the transactions. We have used the Hyperledger Representational State Transfer (REST) interface to provide the connectivity of clients and blockchain. The performance parameter of the blockchain is given below.

4.2. Transaction Latency

It is the time period taken by the process of blockchain to commit the transaction from the time it is submitted [21]. As transactions are committed in blocks after consensus so, it is difficult to find out the transaction latency for an individual transaction. Therefore, average transaction latency has to be calculated for the blockchain network and it is defined as below-

$$\text{AverageTransactionLatency} = \frac{(\sum \text{TransactionLatency})}{\text{TotalCommittedTransactions}} \quad (1)$$

5. Transaction throughput

It is the rate at which valid transactions are committed in blockchain in a definite time interval [22] and it is given as below-

$$\text{AverageTransactionthroughput} = \frac{\text{TotalCommittedTransactions}}{\text{TotalTime@percentageofcommittedtransactions}} \quad (2)$$

5.1. Scalability

Scalability is defined as how much total time is maintained for low transaction latency with an increased number of workloads [16].

There are many organizations involved to transport the raw material and so at a time, many participants from many organizations can be involved to submit the transactions to the SCRM blockchain. These submitted transactions completed the consensus process and finally has a state such as committed or failed which is depicted by figure 9.

We have used four scenarios to test the performance of SCRM, 1-organizations- 2-client-10 test runs, 2-organizations- 6-client-10 test runs, 3-organizations- 10-client-10 test runs, and 4-organizations- 24-client-10 test runs. The test runs are completed on Ubuntu 16.04 system including Intel(R) Core(TM) i5-7400 CPU @ 3.00GHz, 3000 Mhz, 4 Core(s), 4 Logical Processor(s), and 8GB RAM. We have considered multiple fabric networks and measured the performance of SCRM blockchain. Table 2 represents the different values achieved for average latency and throughput for all four scenarios.

Throughput increases when the block size is increased for the same type of network and this pattern remains the same for all the other different types of networks. Therefore, the behaviour of the SCRM is non-fluctuated wrt throughput and this is depicted in figure 10. For block size 5 the minimum average latency is 1.8 ms and maximum average latency is 51.77 ms and the throughput range is from 4 to 9 transactions for all four scenarios. Similarly, almost the same kinds of patterns are achieved for other remaining block sizes

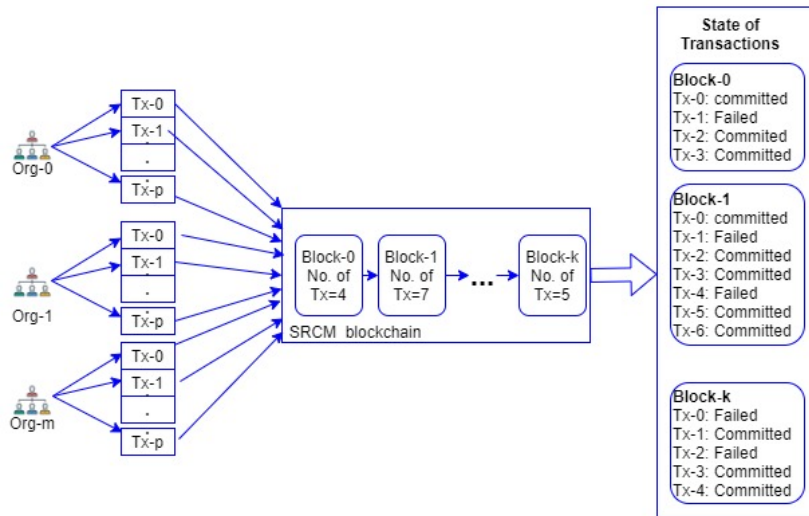


Fig. 9. Pictorial Representation of Performance Metric

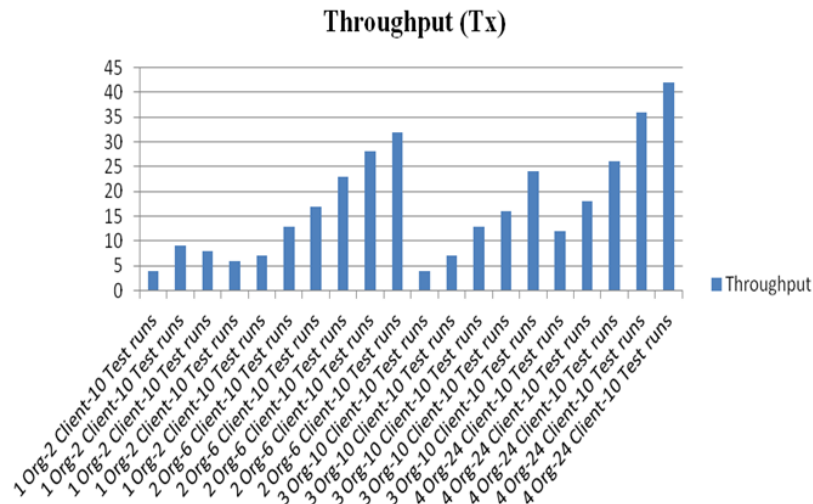


Fig. 10. Throughput for our all four scenarios

including 10, 15, 20, and 25 as depicted in figure 11. Therefore, block size up to 25 the SRCM is supporting the scalability perfectly. Average latency increases when the block size is increased for the same type of network and this pattern remains the same for all the other different types of networks. Therefore, the behaviour of the SRCM is non-fluctuated wrt of average latency and this is depicted in figure 12. We can infer from the results that the proposed SRCM solution is far better than the other traditional solutions like relational databases. In the relational database system, the transactions are not transparent

Table 2. Values achieved for average latency and throughput for all four scenarios

Terminology	Brief description				
Network Type	Block Size	Tx(tps)	Tx time	Avg Latency	Throughput
1 Org-2 Client-10 Test runs	5	8	40	1.8	4
2 Org-6 Client-10 Test runs	5	48	226	9.81	9
3 Org-10 Client-10 Test runs	5	176	365	35.68	8
4 Org-24 Client-10 Test runs	5	257	687	51.77	6
1 Org-2 Client-10 Test runs	10	3	33	0.39	7
2 Org-6 Client-10 Test runs	10	38	228	3.97	13
3 Org-10 Client-10 Test runs	10	187	379	19.19	17
4 Org-24 Client-10 Test runs	10	289	702	29.31	23
1 Org-2 Client-10 Test runs	15	9	58	0.76	28
2 Org-6 Client-10 Test runs	15	43	287	3.02	32
3 Org-10 Client-10 Test runs	15	165	366	11.45	4
4 Org-24 Client-10 Test runs	15	317	764	21.55	7
1 Org-2 Client-10 Test runs	20	7	54	0.48	13
2 Org-6 Client-10 Test runs	20	46	276	2.47	16
3 Org-10 Client-10 Test runs	20	199	403	10.44	24
4 Org-24 Client-10 Test runs	20	297	701	15.27	12
1 Org-2 Client-10 Test runs	25	11	65	0.61	18
2 Org-6 Client-10 Test runs	25	54	305	2.34	26
3 Org-10 Client-10 Test runs	25	178	376	7.59	36
4 Org-24 Client-10 Test runs	25	307	735	12.7	42

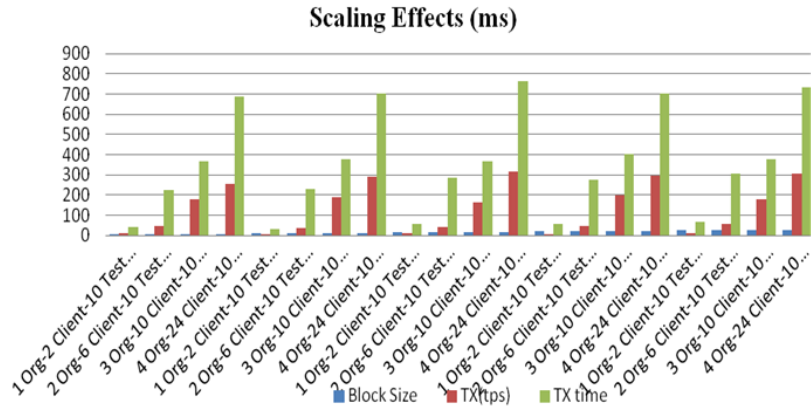


Fig. 11. Scaling effects for different block sizes

as they are under the control of the database administrator but in the case of blockchain the transaction is transparent to all the nodes of the blockchain once it is committed and no one can stop it. In the case of security, the database can be compromised by many means like SQL injection attacks but in the case of blockchain, it is next to impossible to breach the security as the key is almost unpredictable. In the case of scalability, the blockchain can be extended at any limit but the database can't be scaled after a limit as it has lots of consequences regarding efficiency due to exhaustive search space.

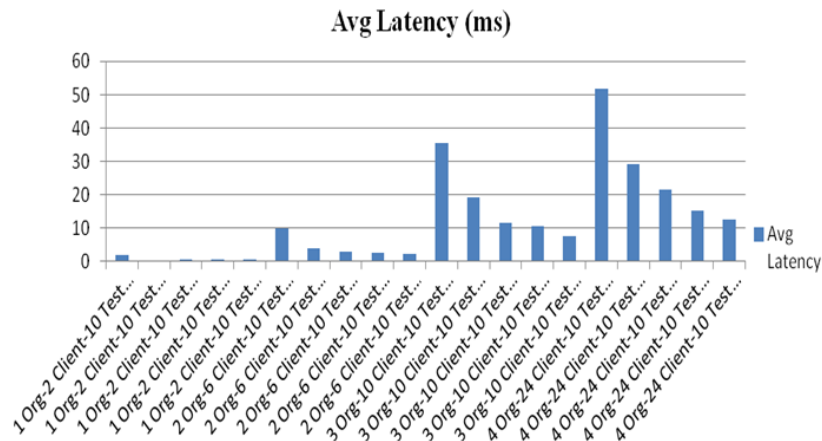


Fig. 12. Average latency for our four scenarios

6. Conclusion

SCRM chain has been proposed with a proof-of-concept for shipping raw materials. SCRM provides a secure, transparent, and integrated raw material delivery. It records all the transactions by various participants of different organizations transparently to all stakeholders. These records can be tracked back if they need an investigation or if there is any fraud that occurred knowingly or unknowingly. A performance evaluation of SCRM is also provided by using the Hyperledger Caliper tool for different four types of modelled networks at different transactions per second. Scalability, Latency, and throughput behaviours for all four modelled networks are provided with that show acceptability of the SCRM. We hope it will be an important tool for automating the raw material shipping industry in future communication as well.

References

1. Abreu, P.W., Aparicio, M., Costa, C.J.: Blockchain technology in the auditing environment. In: 2018 13th Iberian Conference on Information Systems and Technologies (CISTI). pp. 1–6. IEEE (2018)
2. Akram, A., Bross, P.: Trust, privacy and transparency with blockchain technology in logistics. In: MCIS. p. 17 (2018)
3. van Baar, R.B., van Beek, H.M., Van Eijk, E.: Digital forensics as a service: A game changer. *Digital Investigation* 11, S54–S62 (2014)
4. Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M., Walton, M.: From proof of concept to scalable policies: Challenges and solutions, with an application. *Journal of Economic Perspectives* 31(4), 73–102 (2017)
5. Baset, S.A., Desrosiers, L., Gaur, N., Novotny, P., O’Dowd, A., Ramakrishna, V.: Hands-on blockchain with Hyperledger: building decentralized applications with Hyperledger Fabric and composer. Packt Publishing Ltd (2018)
6. Casino, F., Dasaklis, T.K., Patsakis, C.: A systematic literature review of blockchain-based applications: Current status, classification and open issues. *Telematics and informatics* 36, 55–81 (2019)

7. Casino, F., Politou, E., Alepis, E., Patsakis, C.: Immutability and decentralized storage: An analysis of emerging threats. *IEEE Access* 8, 4737–4744 (2019)
8. Choi, W., Hong, J.W.K.: Performance evaluation of ethereum private and testnet networks using hyperledger caliper. In: 2021 22nd Asia-Pacific Network Operations and Management Symposium (APNOMS). pp. 325–329. IEEE (2021)
9. Farnaghi, M., Mansourian, A.: Blockchain, an enabling technology for transparent and accountable decentralized public participatory gis. *Cities* 105, 102850 (2020)
10. Kinory, E., Smith, S.S., Church, K.S.: Exploring the playground: Blockchain prototype use cases with hyperledger composer. *Journal of Emerging Technologies in Accounting* 17(1), 77–88 (2020)
11. Lashkari, B., Musilek, P.: A comprehensive review of blockchain consensus mechanisms. *IEEE Access* 9, 43620–43652 (2021)
12. Liu, M., Wu, K., Xu, J.J.: How will blockchain technology impact auditing and accounting: Permissionless versus permissioned blockchain. *Current Issues in auditing* 13(2), A19–A29 (2019)
13. Messias, J., Alzayat, M., Chandrasekaran, B., Gummadi, K.P.: On blockchain commit times: An analysis of how miners choose bitcoin transactions. In: The Second International Workshop on Smart Data for Blockchain and Distributed Ledger (SDBD2020) (2020)
14. Namasudra, S., Deka, G.C., Johri, P., Hosseinpour, M., Gandomi, A.H.: The revolution of blockchain: State-of-the-art and research challenges. *Archives of Computational Methods in Engineering* 28(3), 1497–1515 (2021)
15. Naz, M., Al-zahrani, F.A., Khalid, R., Javaid, N., Qamar, A.M., Afzal, M.K., Shafiq, M.: A secure data sharing platform using blockchain and interplanetary file system. *Sustainability* 11(24), 7054 (2019)
16. Shahriar Hazari, S., Mahmoud, Q.H.: Improving transaction speed and scalability of blockchain systems via parallel proof of work. *Future internet* 12(8), 125 (2020)
17. Soltani, S., Seno, S.A.H.: A formal model for event reconstruction in digital forensic investigation. *Digital Investigation* 30, 148–160 (2019)
18. Taylor, P.J., Dargahi, T., Dehghantaha, A., Parizi, R.M., Choo, K.K.R.: A systematic literature review of blockchain cyber security. *Digital Communications and Networks* 6(2), 147–156 (2020)
19. Wang, W., Hoang, D.T., Hu, P., Xiong, Z., Niyato, D., Wang, P., Wen, Y., Kim, D.I.: A survey on consensus mechanisms and mining strategy management in blockchain networks. *Ieee Access* 7, 22328–22370 (2019)
20. Wang, Y., Zhang, C., Xiang, X., Zhao, Z., Li, W., Gong, X., Liu, B., Chen, K., Zou, W.: Revery: From proof-of-concept to exploitable. In: Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. pp. 1914–1927 (2018)
21. Yasaweerasinghelage, R., Staples, M., Weber, I.: Using architectural modelling and simulation to predict latency of blockchain-based systems. School of Computer Science and Engineering, UNSW Australia, Tech. Rep 201704 (2017)
22. Yu, G., Wang, X., Yu, K., Ni, W., Zhang, J.A., Liu, R.P.: Survey: Sharding in blockchains. *IEEE Access* 8, 14155–14181 (2020)
23. Zhang, R., Xue, R., Liu, L.: Security and privacy on blockchain. *ACM Computing Surveys (CSUR)* 52(3), 1–34 (2019)

Hemraj Saini is a Professor in the School of Computing, DIT University, Dehradun, INDIA. Prior to that he has worked in Jaypee University of Information Technology, Waknaghat (2012-2021), AIET, Alwar (2011-2012); OEC, Bhubaneswar (2008-2011); HIE, Baniwalid (Libya) (2007-2008); BITS, Pilani (2005-2007); IET, Alwar (2001-2005);

REIL, Jaipur (2000-2001) and Dataman System, Delhi (1999-2000) for almost 23 years in Academics, Administration and Industry. He has obtained PhD (Computer Science) from Utkal University, VaniVihar, Bhubaneswar; M.Tech. (Information Technology) from Punjabi University, Patiala; and B.Tech. (Computer Science & Engineering) from Regional Engineering College, Hamirpur (H.P.), Now NIT. six (06) Ph.D. degrees have been awarded under his valuable guidance. He is an active member of various professional technical and scientific associations such as IEEE, ACM, IAENG, etc. Presently he is providing his services in various modes like, Editor, Member of Editorial Boards, Member of different Subject Research Committees, reviewer for International Journals and Conferences including Springer, Science Direct, IEEE, Wiley, IGI Global, Bentham Science etc. and as a resource person for various workshops and conferences. He has published more than 150 research papers in International/National Journals and Conferences of repute. He has also organized various conferences and workshops including- NCRTDM 2011, OEC, BBSR (AICTE, DST and CSIR sponsored) and BSSCAD 2009, OEC, BBSR (DST and CSIR sponsored, INSPIR CAMP under the DST Internship, funded by DST in August 2012, IEEE PDGC-2012 as member of International TPC, 2013-IEEE ICIP as Technical Program Co-Chair, 2020-IEEE PDGC and 2015-IEEE ICIP as Conference General Chair, PDGC-2016 as the Publicity Committee Chairs and ICIP-2017 as registration Chair.

Satyabrata Dash presently working as Professor in department of Computer Science & Engineering at Ramachandra College of Engineering, Eluru, Andhra Pradesh. He received his B.Tech. in Computer Science Engineering from Biju Pattnaik University of Technology (BPUT) in 2003 and M.Tech. degree in Computer Science Engineering from KIIT university Bhubaneswar in 2006 and PhD in Computer Science & Engineering from Centurion University of Technology and Management, (CUTM) Paralakhumandi, Odisha. He is currently member of different professional society like SMIEEE, ISTE, IE, The Society of Digital Information and many more. He is having more than 17 Years of UG and PG teaching experience and has published two book, 8 Indian and 1 International Patents (Granted), 8 Book chapters, and 41 research papers in international and national journals and conferences. His research areas include Smart Agriculture solutions, Cloud computing, IOT, intrusion detection, and e-Gov applications. He Received Prof. Tribikaram Pati BEST PRESENTATION AWARD Instituted by Prof. Gopal Krishna Panda in Memory of Late Satyananda Panda for the year 2016-17 at 44th Annual Conference of ODISHA MATHEMATICAL SOCIETY, Got the second top performer in eGov campus initiative in 2013-14 from Engineering watch, for study on e-District MMP of Puri District, Received Best Teacher Award at Orissa Technological Conclave 2018, Organized by EGF Odisha.

Subhendu Kumar Pani is Professor in the Department of Computer Science and Engineering at Krupajal Engineering College (KEC) Bhubaneswar, India. He has more than 19 years of teaching and research experience. His research interests include data mining, big data analysis, web data analytics, fuzzy decision-making, and computational intelligence. He is the recipient of five researcher awards. In addition to research, he has guided many PhD and MTech students. He has published over 70 international journal papers (many of which are Scopus indexed). His professional activities include roles as associate editor, editorial board member, and reviewer of various international journals. He is associated

with several conferences and societies. He has more than 150 international publications, several authored books and edited books, and book chapters to his credit. He is a fellow of the Scientific Society of Advanced Research and Social Change and a life member in many other professional organizations. Dr. Pani is also editor of the book series AAP Advances in Artificial Intelligence & Robotics.

Maria José Sousa (PhD in Management) is Pro-Rector for the Development of Distance Learning and a professor and a research fellow at ISCTE/Instituto Universitário de Lisboa. Her research interests currently are public policies of Innovation and Education. She is a best seller author in Research Methods, ICT and People Management and has co-authored over 100 articles and book chapters and is the guest-editor of more than 5 Special Issues from Elsevier and Springer. Is the former President of the ISO/TC 260 – Human Resources Management, representing Portugal in the International Organization for Standardization. She has coordinated several European projects of innovation and is also External Expert of COST Association - European Cooperation in Science and Technology.

Álvaro Rocha holds the title of Honorary Professor, and holds a D.Sc. in Information Science, Ph.D. in Information Systems and Technologies, M.Sc. in Information Management, and BCs in Computer Science. He is a Professor of Information Systems at the University of Lisbon - ISEG, researcher at the ADVANCE (the ISEG Centre for Advanced Research in Management), and a collaborator researcher at both LIACC (Laboratory of Artificial Intelligence and Computer Science) and CINTESIS (Center for Research in Health Technologies and Information Systems). His main research interests are maturity models, information systems quality, online service quality, requirements engineering, intelligent information systems, e-Government, e-Health, and information technology in education. He is also Vice-Chair of the IEEE Portugal Section Systems, Man, and Cybernetics Society Chapter, and Editor-in-Chief of both JISEM (Journal of Information Systems Engineering & Management) and RISTI (Iberian Journal of Information Systems and Technologies). Moreover, he has served as Vice-Chair of Experts for the European Commission's Horizon 2020 Program, and as an Expert at the COST - intergovernmental framework for European Cooperation in Science and Technology, at the European Commission's Horizon Europe Program, at the Government of Italy's Ministry of Universities and Research, at the Government of Latvia's Ministry of Finance, at the Government of Mexico's National Council of Science and Technology, at the Government of Polish's National Science Centre, and at the Government of Cyprus's Research and Innovation Foundation.

Received: September 30, 2021; Accepted: April 01, 2022.

An Innovative Quality Lane Change Evaluation Scheme based on Reliable Crowd-ratings

Konstantinos Psaraftis, Theodoros Anagnostopoulos, and Klimis Ntalianis

Department of Business Administration, Division of Information Systems and Decision
Making
University of West Attica,
250 Thivon & P. Ralli, Egaleo 12241, Greece
kostaspsaraftis@hotmail.com
theodoros.anagnostopoulos@uniwa.gr
kntal@teiath.gr

Abstract. Intelligent Transportation Systems (ITSs) and their applications are attracting significant attention in research and industry. ITSs make use of various sensing and communication technologies to assist transportation authorities and vehicle drivers in making informative decisions and provide leisure and safe driving experience. Data collection and dispersion are of utmost importance for the proper operation of ITSs applications. Numerous standards, architectures and communication protocols have been anticipated for ITSs applications. In recent years, crowdsourcing methods have shown to provide important benefits to ITSs, where ubiquitous citizens, acting as mobile human sensors, help respond to signals and providing real-time information. In this paper, the problem of mitigating crowdsourced data bias and malicious activity is addressed, when no auxiliary information is available at the individual level, as a prerequisite for achieving better quality data output. To achieve this goal, an innovative algorithm is designed and tested on a crowdsourcing database of lane change evaluations. A three-month crowdsourcing campaign is performed with 70 participants, resulting in a large number of lane changes evaluations. The proposed algorithm can negate the noisy ground-truth of crowdsourced data and improve the overall quality.

Keywords: crowdsourcing, intelligent transportation systems, subjective ratings, lane change evaluation, bias reduction, malicious activities, fuzzy logic.

1. Introduction

Road accidents constitute a major social problem in modern societies. Approximately 1.35 million people die every year on the roads worldwide, and another 20 to 50 million sustain non-fatal injuries as a result of road traffic accidents [1]. It is estimated that lane-change crashes account for 4 to 10 % of all crashes. These injuries and fatalities have an immeasurable impact on the families affected, whose lives are often changed irrevocably by these tragedies, and on the communities in which these people lived and worked [2]. In the meantime, careful and lawful drivers are not rewarded for their responsible driving. According to [3] in the Belonitor project in Denmark by rewarding the participants' good driving behaviour, the percentage of kilometres they travelled within

the speed limit increased from 68% to 86%, and the number of kilometres driven a safe distance from the car in front rose from 58% to 77%. However, as soon as the feedback and reward system ended, most drivers returned to their old habits. Most participants acknowledged that the combination of feedback and reward had a strong positive effect. For driving evaluation reports in general, there are three main sources of data that researchers are utilizing. The first source of data collection is by sensors [4] during the act of driving vehicles. The second main category is from the CAN bus data [5], which basically records everything on the running state of the car, such as car speed, steering wheel, gps location, brake and other. The third category includes video data [6], which often derive from a mobile phone or a dashcam. To this end, the first research question is addressed: Is it possible to perform a reliable driver's evaluation out of video data by utilizing crowdsourcing solutions?

Crowdsourcing, in general terms, is the act of taking a job traditionally performed by a designated employee and outsourcing it to an undefined, generally large group of people (a "Crowd") in the form of an open call [7]. Technically speaking, Crowdsourcing is a distributed problem-solving and production model. In such a model, initially, the problems are formulated in a format that can be understood easily by technical and non-technical people. This model is used in many applications. For instance [8], proposed an approach to develop a crowd sensing framework to allow an easier cooperation between the citizens and the authorities by collecting information on crimes and suspects through an e-participatory infrastructure. Specifically, in transportation it has emerged as a novel mechanism for accomplishing temporal and spatial critical tasks with the collective intelligence of individuals and organizations.

In ITSs, Xiao Wang [9] performed a quantitative analysis of related research topics and categorized seven kinds of main crowdsourcing based ITSs services: Crowdsourced geospatial data collection, which is contributed by non-expert end-users for altruistic reasons, which both fully utilize end-users' significant local expertise and provide better data and temporal coverage. Urban traffic planning and management, which focuses on bus arrival time prediction, common trajectory pattern identification, shortest-path computing, optimal route planning, customized deployment of cycle length and signal transition time of traffic light systems, travel information recommendation, etc. Green transportation, which aims to reduce fuel consumption and carbon emission, and to provide a highly efficient trip mode for both public and private transports using crowdsourcing based mobile applications. Social navigation, which leverages public online information with users' social network resources, providing real time exploration in novel and strange environments. Road condition monitoring and assessment, which enables people to effectively take part in solving time-spatial critical traffic tasks without generating the additional financial burden on the transportation agencies with the help of social media sites like Facebook, Twitter, YouTube, and Flickr. Smart parking, which is a long-standing problem in ITSs, because searching for street parking and navigating to it in a crowded urban area impose great societal and environmental challenges. Traffic network construction and communication, where employees' ubiquitous roadside units and vehicular ad hoc networks integrate the capabilities of new generation wireless networks and provide infrastructural support of the inter-vehicle, vehicle-to-roadside and inter-roadside communications in hybrid vehicular ad hoc networks.

In this research, a crowdsourced geospatial data collection solution for peer-to-peer lane change vehicle evaluations is proposed. To this end, the second research question is addressed: How can workers' bias and malicious responses be reliably mitigated? It is proved that data acquired from tasks that comprise a subjective component (e.g. ratings, opinion detection, sentiment analysis) is potentially affected by the inherent bias of crowd workers who contribute to the tasks [10]. In addition, workers may not take tasks seriously. Gadiraju [11] in his research analysed the malicious behaviour in the crowd and defined five categories of untrustworthy workers. Ineligible Workers (IW), who do not comply to the prior stated pre-requisites, e.g. 'Please attempt this task only after you successfully complete 3 tasks. Fast Deceivers (FD), who give random answers in order to finish a task as fast as possible, e.g., entering random numerical values. Rule Breakers (RB), who do not provide the required quality of the answer, e.g., giving 1 keyword, when the task requires at least 3 keywords. Smart Deceivers (SD), who conform to the rules but give semantically wrong answers. Finally, Gold Standard Preys (GSP), who follow instructions and provide valid responses but are caught with providing different answers on repeated test questions during the evaluation. In the context of crowdsourcing, subjective tasks with numerical responses, three of the mentioned categories are applied, namely FD, SD and GSP. Consequently, an algorithm to negate this effect is proposed as a worker might provide biased and/or malicious feedback.

The rest of the paper is structured as follows: Section 2 contains related work. In Section 3, the proposed crowdsourcing framework is described and explained. Section 4 describes the 3-month crowdsourcing campaign conducted and discusses the experimental results. Last, section 5 concludes this paper by pointing out some future research directions.

The three major contributions of this paper are summarized below:

- An innovative algorithm is designed to negate the effect of bias and malicious activities with regards to subjective crowdsourcing environments.
- A crowdsourcing peer-to-peer evaluation framework is proposed, which mainly focuses on lane-change driving acts.
- A large-scale crowdsourcing campaign is carried out with regards to lane-change evaluations and results demonstrate the effectiveness and overall data quality improvement.

2. Literature Review

One of the main and most challenging issues that still exist in crowdsourcing applications, especially in subjective studies where no ground truth exists, is ensuring the reliability of workers' ratings. For that purpose, in case of crowdsourced datasets that record auxiliary information from participants (such as gender, age, income or education level), the work in [12] proposed to apply quasi-randomization techniques in which pseudo-inclusion probabilities are estimated based on covariates available for samples and non-sample units. In other approaches such as in [13, 14], in order to reduce sample bias and adjust the non-probability samples to the target population distributions, pseudo-sampling weights are estimated that are predictive of the outcome of interest and/or the probability of selection. Wang et al. [15] propose a multilevel

regression and post-stratification (MRP) method, which is an extension of the hierarchical regression modelling. The work in [16] focuses on debiasing crowdsourcing answers to estimate the average innate opinion of the social crowd with a small number of samples and depends on the social dependency among workers. Other common techniques used to correct worker bias are Bayesian Additive Regression Trees (BART), Inverse Probability Bootstrapping [17], the Least Absolute Shrinkage and Selection Operator, LASSO [18] and the Propensity Score Adjustment [19]. However, these approaches record large samples of highly relevant variables. In this research, minimal information is available at the individual level because it is common that workers in volunteered geographic information applications do not provide auxiliary individual information apart from the measure of interest and the geographical information.

Researchers have also studied worker bias estimation techniques in crowdsourcing platforms. The works in [20, 21] study the data annotation bias, when data items are presented as batches to be judged by workers simultaneously and propose models to characterize the annotating behavior on data batches. However, they focus on binary answers and their goal is to properly categorize each data item instead of estimating worker bias and malicious activity. In [21], the authors show that crowdsourcing workers have both bias and variance and propose an approach to recover the true quantity values for crowdsourcing tasks with an unsupervised probabilistic model to jointly assess task difficulties. In [22] the proposed scheme aims to solve the above problem by building and using probabilistic graphical models for jointly modeling task features, workers' biases, worker contributions and ground truth answers of tasks, so that task-dependent bias can be corrected. In order to achieve effective models, the aforementioned approaches need a large number of worker responses and additionally, they do not consider workers' malicious activities at all. On the contrary, the proposed algorithm performs well and improves data output even when the number of worker responses is low. In addition, it detects and negates the effect of malicious activities.

Quality control in the data output is also approached with task assignment techniques. For example, [23] investigates the accuracy of workers by evaluating their performance on the completed tasks and predicts which tasks the workers are well acquainted with; [24] propose a framework comprising an inference model and an online task assigner. They prove that inserting a gold standard question helps estimate the worker accuracy and supports blocking of poor workers; [25] estimate the workers' accuracy according to their previous performance and the core quality-sensitive model is able to control the processing latency; [26] developed a quality-sensitive answering model, which guides the crowdsourcing engine to process and monitor the human tasks. The model achieves reliable results by providing an estimated accuracy for each generated result based on the human workers' historical performances. Different from the previous approaches, in this research, only non-auxiliary subjective tasks are considered. Moreover, while these works focus on selecting the reliable workers to perform the tasks, they do not consider workers' bias or malicious activity combined in their responses and they do not focus on estimating quantitative values, such as evaluations. Authors in [27] aim to reliably identify crowdsourced events by selecting a small subset of human sensors to perform tasks. Similarly, to the algorithm proposed in this research, they exploit linear regression to estimate worker bias in each task and attempt to eliminate it, the moment workers provide their ratings. However, in their setting, they assume that, although answers

might be subjective, all are considered as truthful. Hence, they do not take malicious activities into consideration at all.

Researchers in the field have acknowledged the importance and need for techniques to deal with inattentive workers, scammers, incompetent and malicious workers. Authors in [11] analyzed the prevalent malicious activity on crowdsourcing platforms and studied the behavior exhibited by trustworthy and untrustworthy workers. Eickhoff et al. [28] aimed to identify measures that one can take in order to make crowdsourced tasks resilient to fraudulent attempts. The authors concluded that understanding worker behavior better is pivotal for reliability metrics. Difallah et al. [29] reviewed existing techniques used to detect malicious workers and spammers and described the limitations of these techniques. In another relevant work by Gadiraju et al. [11], the authors proposed to design and plan micro-tasks such that they are less attractive for cheaters. In order to do so, the authors evaluated factors such as the type of micro-task, the interface used, the composition of the crowd and the size of the micro-task. All previous research, however, does not take into consideration workers' bias in the datasets.

To the best of the authors' knowledge, no study has reported a similar algorithm to evaluate workers' performance and the use of crowdsourcing techniques for lane-change vehicles' evaluation. A novel algorithm is presented to mitigate bias and malicious activity on workers' ratings and improve overall quality of data output. These instructions and the corresponding MS Word document template are based on the corresponding Springer instructions and MS Word document template for preparing camera ready papers to be published in the Springer series Lecture Notes in Computer Science.

The preparation of manuscripts which are to be reproduced by photo-offset requires special care. Papers submitted in a technically unsuitable form will be returned for retyping, or canceled if the volume cannot otherwise be finished on time.

3. Framework Overview

3.1. System overview

A first high-level overview of the proposed architecture is presented in Figure 1. In general, the requester will post tasks to a crowdsourcing platform as an input. These tasks will be given to a pool of workers for evaluation. The workers' submitted responses will be the raw data for the system. Raw data will be processed through the recommended algorithm to negate crowdsourced data bias and malicious activities. Finally, the proposed algorithm will output the data with improved overall quality.

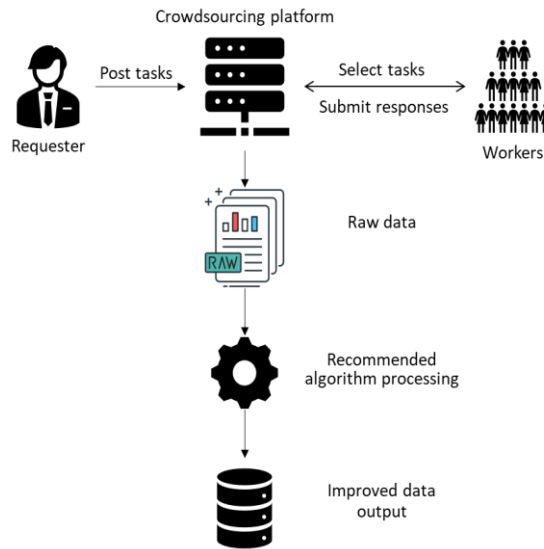


Fig. 1. System Overview

3.2. Problem formulation

Given a set of workers denoted as $w \in W$ that take part in the system and a set of events denoted as $e \in E$, the goal is data acquisition regarding an evaluable event. During this process, it is very important to acquire as many reliable ratings R for as many events as possible. More specifically, the goal is to populate a database of tuples of the form $T = \langle w, e, r \rangle$, where w is a worker, e is the event and r is the rating provided by w for an event e . Given that subjective events that are evaluable may contain bias and/or malicious activity, the framework of data acquisition has the following two secondary goals. First goal is that worker w should not be biased. Therefore, bias, denoted as $bias_w$, which indicates the likelihood of providing ratings above or lower than the average rating must be calculated. Second goal is that worker should not be deceitful. The level of maliciousness, if any, is denoted as $cheat_w$ and must be measured. Both goals, calculating $bias_w$ and $cheat_w$ are important for the realization of the primary goal, acquiring reliable ratings for events. In the next section, the framework that realizes the above goals to achieve the primary goal of acquiring reliable ratings for as many evaluable events as possible from a crowd of workers is described.

3.3. Framework

In a nutshell, the framework works as follows. Initially, a crowdsourcing system with a well of events to be rated into E itemset e_1, \dots, e_n is formulated. Each event e is associated with the following attributes: $\langle id_e, lat_e, long_e \rangle$. Attribute id_e is the event's unique identifier in the framework, and $lat_e, long_e$ corresponds to the current location in terms of latitude, longitude of the event. The total set of possible ratings R is:

$$\left\{ \begin{array}{l} 1 \text{ (Dangerous),} \\ 2 \text{ (Needs improvement),} \\ 3 \text{ (Neither good nor bad),} \\ 4 \text{ (Very good),} \\ 5 \text{ (Excellent)} \end{array} \right\}$$

Ratings are mapped to values between 1 and 5 to stay compatible with the 5-star rating paradigm proposed by [30] and used by most recommender systems. In the system, a worker w has the following attributes: $\langle id_w, bias_w, cheat_w, prev_w[] \rangle$ where id_w is the worker's unique identifier in the system, $bias_w$ represents bias based on expertise and skills when estimating the evaluation of an event, $cheat_w$ correspond to the worker's maliciousness score, and $prev_w[]$ is used to store information about the evaluations completed by the worker w . The calculated value for the event's rating is denoted as val_e^W and computed based on input from all workers in the set W . The crowdsourcing answer of worker w to the event e is denoted as $a_{w,e}$. In the rest of the section, the proposed algorithm for mitigating bias and malicious activity is described.

3.4. Worker Bias

Similar to the intuition of authors in [27], worker ratings are considered having a bias which is defined as a linear function of their answers (x -axis) with respect to the difference of their ratings from the average rating when all workers in W are considered (y -axis). Each worker $w \in W$, who is requested to evaluate an event, is assumed to provide an answer $a_{w,e}$ with a bias $b(a_{w,e})$ and thus the estimated debiased response is:

$$val_{w,e} = \frac{\sum_{w \in W} a_{w,e}}{|W|} + b(a_{w,e}) \quad \#(1)$$

The bias $b(a_{w,e})$ is defined as a linear function of the worker's response. Thus, it is possible to estimate the difference from the average value for each worker's response $a_{w,e}$. Consequently, linear regression is exploited, to adjust the worker's bias estimation whenever the worker provides a response. Linear regression is a useful tool in many applications to model the relationship between a scalar response and one or more explanatory variables. For this framework, the matrix points are defined from the worker's response $a_{w,e}$ and its difference from the average value of the event. Thus, the response $a_{w,e}$ provided for each event's evaluation and the respective difference from

the average rating is recorded and the ratio among these two dimensions is defined. This is computed easily using simple linear regression that produces a linear function:

$$b(a_{w,e}) = \mu * a_{w,e} + v \quad \#(2)$$

where μ is the slope and v is the intercept of the line, which are calculated from the linear regression. This way, an estimation for the difference of each worker's rating compared to the average rating from all workers by computing $b(a_{w,e})$ for each rating $a_{w,e}$ is made.

3.5. Worker maliciousness

In this section, the quality of each worker is described by estimating the worker's weight in a simplified setting. To estimate each worker's maliciousness, a fuzzy logic controller is utilized.

Fuzzy logic has proven itself as a promising mathematical approach for addressing subjectivity, ambiguity, imprecision, and uncertainty of linguistic expressions [31]. A fuzzy logic inference system may contain many inputs and outputs and allows the implementation of the rules described in a natural language. An explanatory diagram is shown in Figure 2. The input consists of numerical signals and are called 'crisp', which are later translated into the fuzzy sets through the fuzzification process. A fuzzy set is a pair consisting of linguistic variables. Different membership functions are used to perform the fuzzification process. Often, triangular, or trapezoid functions are used to keep the computational cost low. After the transformation into linguistic variables, the inference rules could be applied. After that process, a so-called defuzzification is needed to generate a sharp output value.

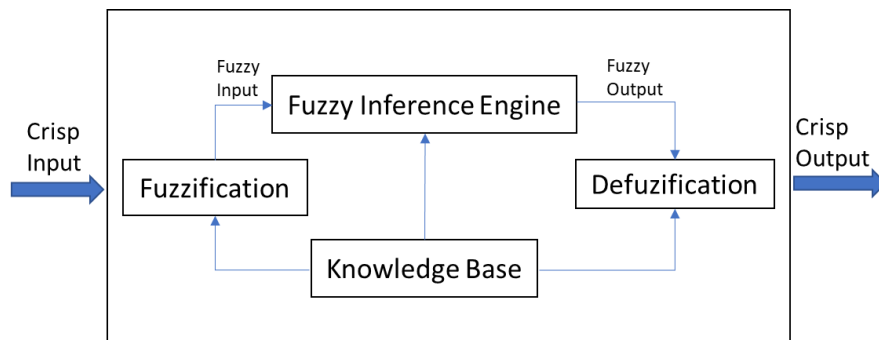


Fig. 2. Fuzzy logic inference diagram

Fuzzy logic controllers are used in many applications. For instance [32], developed a system with the internet of things (IoT) concept for making right decision according to the situation for monitoring and determining fire confidence and reduce the number of rules by doing so sensor activities also reduced and extend battery lifetime as well as

improve efficiency whereas [33] developed an efficient and intelligent IoT communication system to ensure data security with network consistency.

The proposed fuzzy logic maliciousness evaluator has a ‘2 input – 1 output’ structure. First input is based on a Euclidean distance similarity measure between the vector distance of the estimated debiased worker’s evaluation to the estimated average debiased evaluation. Second input is the relative normalized number of times the worker’s estimated vote is above the estimated debiased average evaluation of the event. More formally, the fuzzy logic controller will solve the following problem:

Given two sets of numbers between 0 and 1, where 1 is excellent, that respectively represents the Distance Score and the Over-Under Score of a crowdsourcing worker, what maliciousness weight should be assigned?

In the following paragraphs, the calculation process of the controller’s input is described.

For the first input, namely *Distance Score*: let w_i be a random worker, W the whole set of workers and E events. Also, let $\langle val_{w_i,e_1}, val_{w_i,e_2}, \dots, val_{w_i,e_E} \rangle$ be the vector of debiased ratings the random worker provided. Similarly, let $\langle \frac{\sum_{w \in W} val_{w,e_1}}{|W|}, \frac{\sum_{w \in W} val_{w,e_2}}{|W|}, \dots, \frac{\sum_{w \in W} val_{w,e_E}}{|W|} \rangle$ be the vector of average debiased evaluations. Based on the Euclidean-distance similarity vector similarity, the *Distance Score* is computed as follows:

$$Distance\ score(w_i) = \frac{1}{1 + \sqrt{\sum_{j=1}^E \left(val_{w_i,j} - \frac{\sum_{w \in W} val_{w,e_j}}{|W|} \right)^2}} \#(3)$$

This result in a value in the range of [0,1]. The higher the score, the better the results for the worker, indicating that no malicious activity is detected. Of course, any other suitable similarity measure can be seamlessly used instead, depending on the context for which the framework is used.

For the second input, namely *Over-Under score*: Once again, w_i is assumed to be a random worker, W the whole set of workers and E is the total number of events. For each crowdsourcing event, the number of times the worker’s estimated vote is above the estimated debiased average evaluation of the event, $N_{w_i}(voted\ over)$ and the number of times less $N_{w_i}(voted\ under)$ is tracked. Consequently, the *Over-Under score* is calculated as follows:

$$Over - Under\ score(w_i) = \frac{|N_{w_i}(voted\ over) - N_{w_i}(voted\ under)|}{E} \#(4)$$

In a similar manner as with the *Distance Score*, this result in a value in the range of [0,1]. With this score alone, it is not safe to make assumptions about the worker. Combining *Distance* with the *Over-Under score*, fuzzy sets can be created. The set of rules for the fuzzy logic system are shown in Table 1. To train the fuzzy inference system, two datasets of evaluations were formulated. For the first dataset, a local police department of Athens is contacted where three traffic enforcement officers (experts) assisted by performing lane-change event evaluations. Specifically, officers performed

132 lane-change evaluations and for each event, the averaged of their response was calculated to minimize any possible bias. Their evaluations are all considered truthful and, for the framework, the ground truth of these events. For the second dataset, for the same set of events, the average of the evaluations made by the 70 workers from the crowdsourcing campaign was calculated. Finally, based on the two sets of evaluations, the fuzzy logic controller’s fuzzy rules and membership functions were manually adjusted by trial and error until the root mean square of the framework was minimal.

Table 1. Fuzzy logic evaluation rule-base

Worker weight	Over-Under score			
	Low	Average	Excellent	
Low	Low	Below Average	Average	
Distance score	Average	Below Average	Average	Above Average
	Excellent	Average	Above Average	Excellent

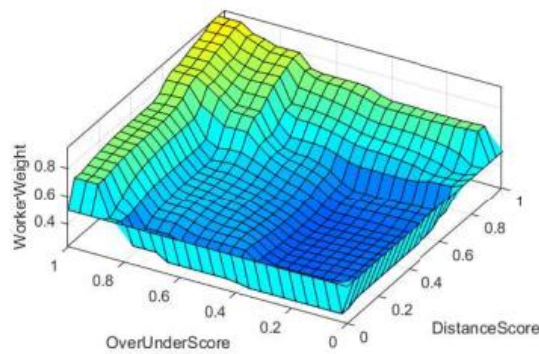


Fig. 3. Worker weight fuzzy evaluator 3D surface

The output variable represents the maliciousness of a worker and is denoted as: $cheat_w \in [0,1]$. The three-dimensional surface of the designed fuzzy logic inference mechanism is displayed in Figure 3 and the corresponding fuzzy logic rules in Figure 4. Variable $cheat_w$ provides an estimation on how malicious a worker may be in the ratings provided. The closest to the upper bound, 1, the better the score for the worker. Hence, to better improve the data quality, the top-K workers are selected with the highest score achieved. The next step after acquiring the fuzzy logic’s output and selecting the top-K workers is to assign each worker a weight which is calculated as shown in equation (5) where K is the selected set of workers with the best score.

$$weight_w = \frac{cheat_w}{\sum_{w \in K} cheat_w} \#(5)$$

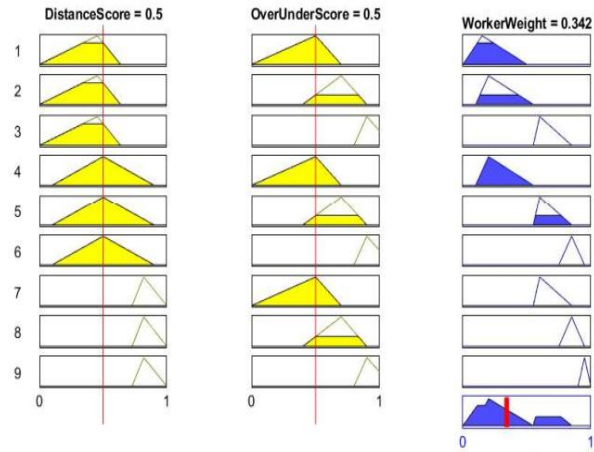


Fig. 4. Worker weight fuzzy evaluator rules

3.6. Event output computation

Finally, the last step is to determine the output of the event. The output of the event is computed using the following equation:

$$val_e^K = \sum_{w \in K} \left((a_{w,e} - b(a_{w,e})) (weight_w) \right) \#(6)$$

Thus, the average value of the retrieved ratings is computed after selecting the top-K workers with the best $cheat_w$ score and eliminating the estimated bias and maliciousness for each of those individual workers.

4. Experimental results

With regards to subjective crowdsourcing environments, the proposed algorithm is a promising solution. It combines linear regression techniques to negate the effect of human bias and a fuzzy logic controller with a trained fuzzy inference system to detect the low-skilled workers. In the paragraphs below, the 3-month crowdsourcing campaign conducted is described and the experimental results are discussed.

The proposed algorithm was evaluated by analyzing the results from a 3-month crowdsourcing campaign of 70 workers¹. Specifically, the data were collected from 7th December 2020 to 19th March 2021. The campaign was performed locally, in the department's research lab in University of West Attica. Participants were all adults with

¹ More information on datasets and applied software can be found here: [GitHub](#)

a valid driving license registration. Of them, 29 (41%) were female and 41 (58%) were male, with an age ranging from 19 to 54 years. Most of the participants (72.8%) were University graduates. For the purposes of the experiments, no other personal information is published since in this study, the problem of mitigating crowdsourced data bias and malicious activity when no auxiliary information is available at the individual level is addressed.



Fig. 5. Assigned lane change event

The first 132 lane change videos were extracted from the UAH-DriveSet [34] which is a public collection of data captured by their driving monitoring application, DriveSafe by various testers in different environments. Thus, workers were given lane changing acts to rate in the form of videos through a 5-point Likert scale. These lane change videos cover national highways, state highways and district roads but no rural or village roads. Additionally, videos cover all traffic volume scenarios, from light to high traffic roads.

Figure 5 shows frames of a lane-change event's video clip with the accompanying task request. A valid worker's participation required all 132 events to be completed, so the total experiment's evaluation dataset contains 9240 lane change evaluations.

In Figure 6, in a clustered bar chart, the total number of answers that were retrieved for each of the ratings for all events is presented. There are 2 noticeable observations made on the chart. Workers' most preferred response was the neutral number 3 (Neither good nor bad) which is a small indication that they were unable to decisively decide whether a lane change is more or less than safe. In addition, the total number of '1' as a feedback, which is the worst evaluation possible, exceeds the total number of '5' which indicates the best response ($1776 > 1714$); these are the video footages, city authorities need to reevaluate.

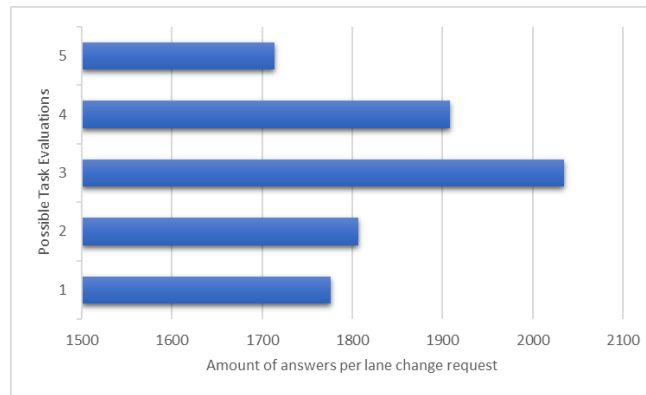


Fig. 6. Number of evaluations for each lane change event rating

In Figure 7, the average evaluations provided by the workers for each individual lane change event were presented. The first 40 lane change events (30.3%) out of 132 in total are classified as ‘1’ (Dangerous) and ‘2’ (Need improvement). Further actions should be planned for these drivers who performed so poorly.

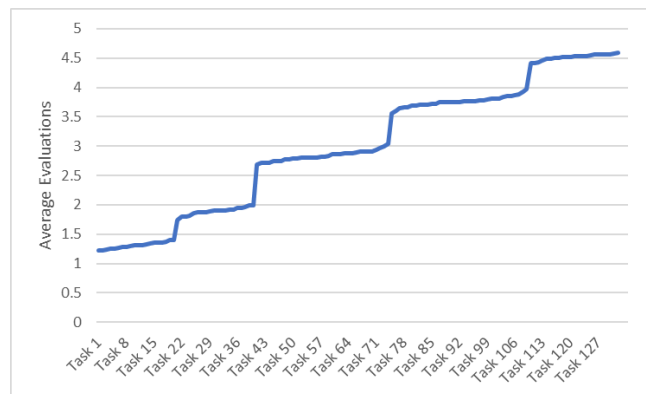


Fig. 7. Average evaluations made per lane change event

In Figure 8, the average evaluation is presented that each individual worker provides for the lane changes in total, whereas in Figure 9 the assigned weight for each worker is illustrated. It appears that in both cases, workers can be classified into two large categories. The first 32 workers (45.7%) appear to provide low responses on average while the rest (54.3%) appear to provide higher responses. In the meantime, in figure 9, the graph results show that the worker weights are clustered into the following mentioned weight classes:

- 34 workers (48.5%) have a weight in the range [0.2205 – 0.2755],
- 5 workers (7.1%) have a weight between [0.4704 – 0.4719],
- 145 workers (20%) between [0.6996-0.7603]

- 17 and the rest of the workers (24.2%) a range [0.8121 - 0.95].

Clustered results arise, due to the fuzzy logic input variables, over-under and distance scores as these are depended on the average debiased score of each task. Therefore, the variables act like a ranking scale and the better the worker’s data quality is, the larger the assigned weight will be.

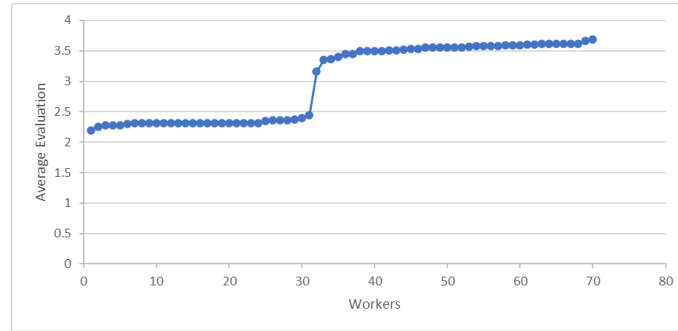


Fig. 8. Average evaluation made per worker (original dataset)

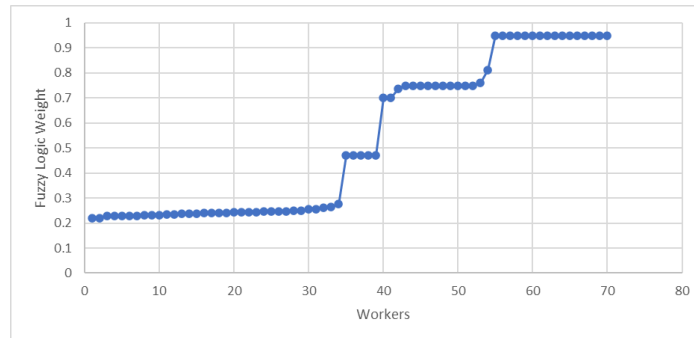


Fig. 9. Average evaluation made per worker

The root mean square error (RMSE) was used to evaluate the accuracy of the recommended algorithm in terms of reliability and effectiveness. Specifically, RMSE is a goodness-of-fit measure of how close the suggested values from different models are, to the initial values. Higher RMSE values indicate poor results, while a smaller RMSE indicates better performance. The relevant formula of RMSE is denoted in equation (7) where n is the sample size and e the difference between each evaluation to the average.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2} \#(7)$$

RMSE is used under three different baseline approaches to better test the proposed algorithm. Specifically:

Random sampling, where the average evaluations from a random set of workers are computed and is denoted by the equation (8) where $\alpha_e^R = \frac{\sum_{w \in R} a_{w,e}}{|R|}$.

$$RMSE (Random Sampling) = \sqrt{\frac{1}{E} \sum_e \left(\alpha_e^R - \frac{\sum_{w \in W} a_{w,e}}{|W|} \right)^2} \quad \#(8)$$

Average sampling from the workers who had acquired the best individual $cheat_w$ score, where their average initial evaluation is computed. Specifically, it can be obtained as shown in equation (9), where $\alpha_e^K = \frac{\sum_{w \in K} a_{w,e}}{|K|}$.

$$RMSE (Average from workers with best score) = \sqrt{\frac{1}{E} \sum_e \left(\alpha_e^K - \frac{\sum_{w \in W} a_{w,e}}{|W|} \right)^2} \quad \#(9)$$

Finally, RMSE calculation based on the recommended algorithm is specified by the following formula:

$$RMSE(Recommended) = \sqrt{\frac{1}{E} \sum_e (val_e^K - \frac{\sum_{w \in W} a_{w,e}}{|W|})^2} \quad \#(10)$$

For the above equations (8, 9, 10), W represents the complete set of workers in the datasets that have evaluated lane change events e , E represents the total number of events and K the varying sample size.

Primarily, for the first set of experiments, the proposed algorithm was examined in terms of effectiveness and accuracy when the number of malicious workers in a dataset varies. For that purpose, the initial 70-worker dataset is injected with a varying number of malicious workers, denoted as $Malicious(m)$, where m is the total number of malicious workers. Truly random evaluations were submitted for each of their lane change task assignments, to simulate real life malicious workers. Figure 10 presents the RMSE(Recommended) score for the lane change events under various numbers of sample size (5-35) and m malicious workers. In all cases where the sample size varies and the injected malicious workers were in the range $m \in [0 - 50]$, the algorithm manages to perform well and keep RMSE minimal (0.112–0.2108). As variable m increases, so does RMSE(Recommended) with performance ranging in (0.5669-0.6066). Therefore, the proposed model has a limitation when the number of malicious users' ratio is known in advance and exceeds 41.6% of the total worker population.

For the rest of the experiments, two versions of datasets were compared in terms of RMSE. The first dataset is the initial, which contains the evaluations of 70 workers for 132 lane change events and the respective algorithm's data processing results. The second dataset is injected with 15 malicious workers ($\approx 17.6\%$ of the total worker population), with the same approach as in the first set of experiments. The injected malicious workers do not exceed the proposed model's ratio limitation as described before.

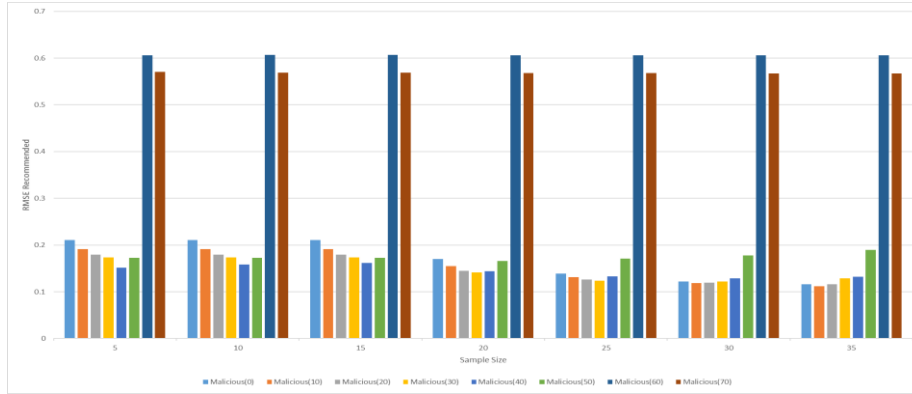


Fig. 10. Varying malicious users sample size evaluation

For both the initial dataset and the dataset with 15 added malicious workers, Figure 11 and Figure 12 respectively, present the RMSE score for the lane change events under various numbers of sample size (5-30). It makes sense that, in all cases, RMSE decreases as the sample size increases in all approaches. In addition, it is also reasonable that RMSE(Average) has high initial scores (0.71 on the initial dataset and 0.74 on the malicious) because biased workers retain better fuzzy logic scores. Furthermore, results from the charts show that the recommended algorithm outperforms all other cases, especially when the sample size is very small. So, by taking 5 workers as a sample, the original dataset results were $RMSE(Random) = 0.278$, $RMSE(Average) = 0.717$, $RMSE(Recommended) = 0.21$, whereas with the malicious dataset, results were even better: $RMSE(Random) = 0.394$, $RMSE(Average) = 0.742$, $RMSE(Recommended) = 0.179$.

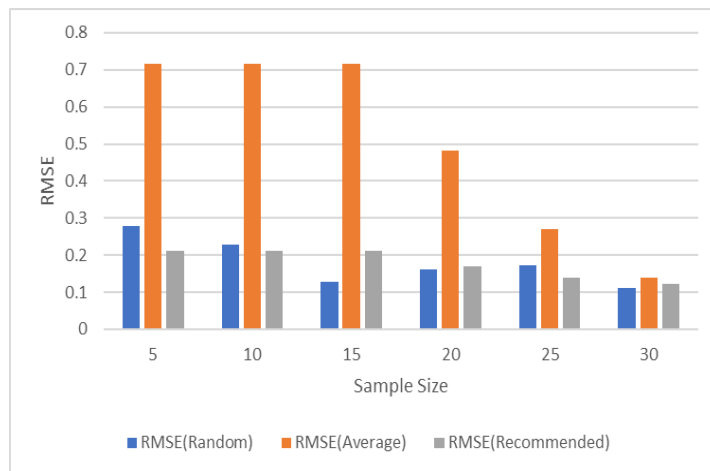


Fig. 11. Varying sample size evaluations (Original dataset)

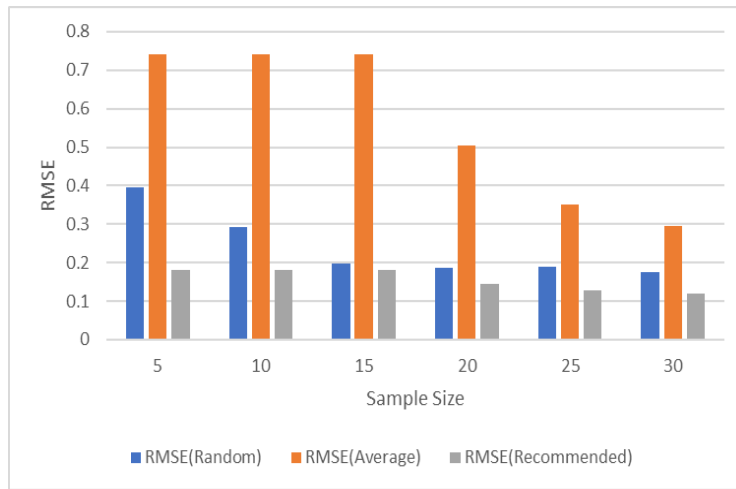


Fig. 12. Varying sample size evaluations (Malicious dataset)

Figures 13 and 14 illustrate how RMSE behaves when the sample size is kept to 15 workers, but the number of total workers varies from 30 to 50 and lastly at 70 workers. For the malicious dataset, for each worker group population, 5 valid workers are replaced with 5 malicious. Thus, the first 30 worker population had 5 malicious workers, the second 10 and the third 15. Both charts show that the recommended algorithm outperforms the two baseline approaches. In addition, although RMSE(Random) and RMSE(Average) were increased in the malicious dataset in comparison to the initial, the proposed algorithm performed similar results. Finally, the proposed algorithm is resilient to malicious workers, since RMSE values range between 0.15 and 0.192.

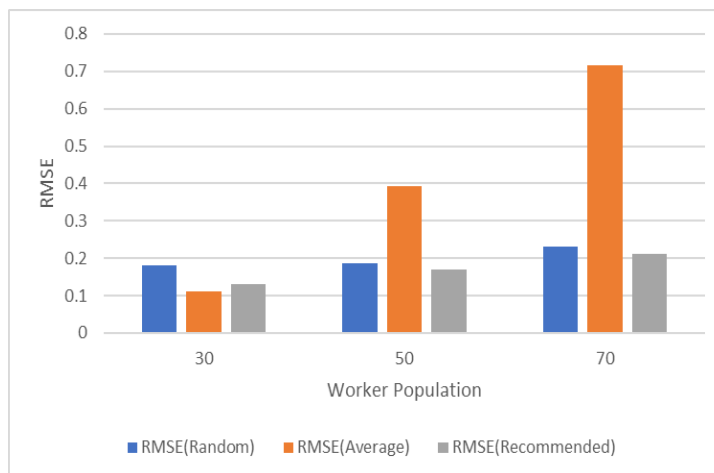


Fig. 13. Varying total workers evaluations (Original dataset)

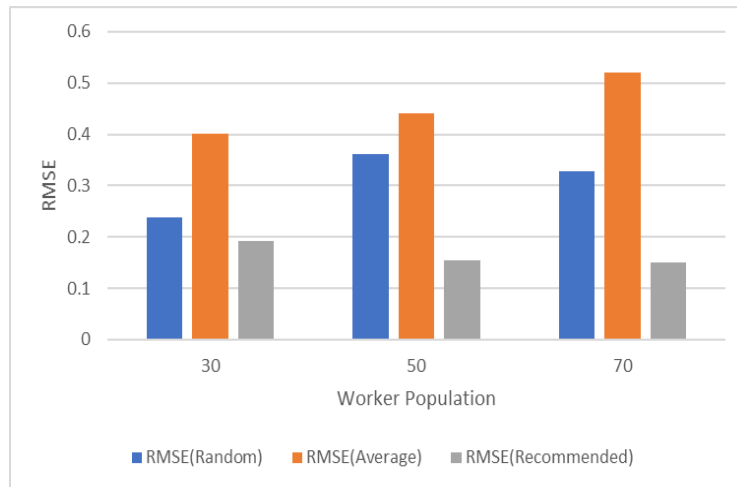


Fig. 14. Varying total workers evaluations (Malicious dataset)

Results prove that the proposed method has many benefits. The inherit human bias is estimated by exploiting linear regression for each task and negates it before the fuzzy logic controller accepts it as an input. Moreover, as a worker may not always be trustworthy, the construction of the fuzzy logic controller proved to be a suitable solution for estimating her quality. The worker's quality estimation is then used to select the best (*top - K*) for an improved data output quality.

5. Conclusion

Crowdsourcing applications have been proposed in intelligent transportation systems for multiple case studies and undeniably it has been an effective tool in bringing people together to solve a problem that affects their community. However, crowdsourcing, like all systems, has its own set of limitations that must be resolved through proper planning and understanding of the system. Specifically, there are concerns about data quality and data management.

To address these concerns, in this paper, a novel algorithm is proposed to address the problem of mitigating crowdsourced data bias and malicious activity to evaluable subjective events when no auxiliary information is available at the individual level as a prerequisite for achieving better quality data output. Experiments involving a crowdsourcing campaign of 70 workers for three months are conducted to evaluate lane changes. Results reveal that the proposed algorithm outperforms in terms of RMSE all other baseline approaches. It should be noted that the settings for the proposed algorithm are tailored for the lane change event. Different traffic events may require a modified version of the proposed algorithm. In pursuit of a more universal approach, additional experiments with other traffic events should be performed and investigate what modifications may be required.

There are many potential enhancements regarding the current framework. Initially, since data anonymization is of crucial importance nowadays, the information recorded for each crowd worker can be protected using relevant techniques [35]. Additionally, a mobile application for real-time vehicle evaluations could automate the data gathering process. Lastly, the fuzzy logic controller has room for improvements to further improve its accuracy and assign even better worker weights.

Acknowledgment. This paper presents a part of the first author's PhD research. This research is partially supported by the Special Research Account of the University of West Attica.

References

1. Agrawal, R., Srikant, R.: Fast Algorithms for Mining Association Rules. In Proceedings of the 20th International Conference on Very Large Databases. Morgan Kaufmann, Santiago, Chile, 487-499. (1994)
2. 1. World Health Organization Global Status Report on Road Safety 2018, <https://www.who.int/publications/i/item/9789241565684>, (2018)
2. Barr L, Najm W (2001): Crash problem characteristics for the intelligent vehicle initiative. Presented at the
3. Hattem J, Mazurek U (2005): Good Driving! The Power of Rewarding. Presented at the November 10
4. Cao W, Lin X, Zhang K, Dong Y, Huang S, Zhang L (2017): Analysis and evaluation of driving behavior recognition based on a 3-axis accelerometer using a random forest approach. In: Proceedings - 2017 16th ACM/IEEE International Conference on Information Processing in Sensor Networks, IPSN 2017, Association for Computing Machinery, Inc, pp. 303-304
5. You CW, Lane ND, Chen F, Wang R, Chen Z, Bao TJ, Montes-de-Oca M, Cheng Y, Lint M, Torresani L, Campbell AT (2013): CarSafe App: Alerting drowsy and distracted drivers using dual cameras on smartphones. In: MobiSys 2013 - Proceedings of the 11th Annual International Conference on Mobile Systems, Applications, and Services, ACM Press, New York, New York, USA, pp. 13-26
6. Ma X, Chau LP, Yap KH (2018): Depth video-based two-stream convolutional neural networks for driver fatigue detection. In: Proceedings of the 2017 International Conference on Orange Technologies, ICOT 2017, Institute of Electrical and Electronics Engineers Inc., pp. 155-158
7. Vander Schee BA: Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business 2009 Jeff Howe. Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business . New York, NY: Crown Business 2008. 320 pp. \$26.95 . J Consum Mark 26, 305-306 (2009)
8. Hasna EE, Abdelaziz E, Zohra EF, Mohamed S: A mobile crowd sensing framework for suspect investigation: An objectivity analysis and de-identification approach. Comput Sci Inf Syst 17, 253-269 (2020)
9. Wang X, Zheng X, Zhang Q, Wang T, Shen D: Crowdsourcing in ITS: The State of the Work and the Networking. IEEE Trans Intell Transp Syst 17, 1596-1605 (2016)
10. Hube C, Fetahu B, Gadiraju U (2019): Understanding and mitigating worker biases in the crowdsourced collection of subjective judgments. In: Conference on Human Factors in Computing Systems - Proceedings, Association for Computing Machinery, New York, NY, USA, pp. 1-12

11. Gadiraju U, Kawase R, Dietze S, Demartini G (2015): Understanding malicious behavior in crowdsourcing platforms: The case of online surveys. In: Conference on Human Factors in Computing Systems - Proceedings, Association for Computing Machinery, New York, NY, USA, pp. 1631–1640
12. Elliott Michael R., Richard Valliant: Inference for Nonprobability Samples. *Stat Sci JSTOR* 32, 249–264 (2017)
13. Baker R, Brick JM, Bates NA, Battaglia M, Couper MP, Dever JA, Gile KJ, Tourangeau R: Summary report of the aapor task force on non-probability sampling. *J Surv Stat Methodol* 1, 90–105 (2013)
14. Elliott MR: Combining Data from Probability and Non- Probability Samples Using Pseudo-Weights. *Surv Pract* 2, 1–7 (2009)
15. Wang W, Rothschild D, Goel S, Gelman A: Forecasting elections with non-representative polls. *Int J Forecast* 31, 980–991 (2015)
16. Das A, Gollapudi S, Panigrahy R, Salek M (2013): Debiasing social wisdom. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, pp. 500–508
17. Nahorniak M, Larsen DP, Volk C, Jordan CE: Using Inverse Probability Bootstrap Sampling to Eliminate Sample Induced Bias in Model Based Analysis of Unequal Probability Samples. *PLoS One* 10, e0131765 (2015)
18. Chen JKT Using LASSO to Calibrate Non-probability Samples using Probability Samples, (2016)
19. Lee S: Propensity score adjustment as a weighting scheme for volunteer panel web surveys. *J Off Stat* (2006)
20. Zhuang H, Parameswaran A, Roth D, Han J (2015): Debiasing crowdsourced batches. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, pp. 1593–1602
21. Ouyang RW, Kaplan L, Martin P, Toniolo A, Srivastava M, Norman TJ (2015): Debiasing crowdsourced quantitative characteristics in local businesses and services. In: IPSN 2015 - Proceedings of the 14th International Symposium on Information Processing in Sensor Networks (Part of CPS Week), Association for Computing Machinery, Inc, pp. 190–201
22. Kamar E, Kapoor A, Horvitz E: Identifying and Accounting for Task-Dependent Bias in Crowdsourcing. *HCOMP* (2015)
23. Fan J, Li G, Ooi BC, Tan KL, Feng J (2015): ICrowd: An adaptive crowdsourcing framework. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, Association for Computing Machinery, pp. 1015–1030
24. Hu H, Zheng Y, Bao Z, Li G, Feng J, Cheng R (2016): Crowdsourced POI labelling: Location-aware result inference and Task Assignment. In: 2016 IEEE 32nd International Conference on Data Engineering, ICDE 2016, Institute of Electrical and Electronics Engineers Inc., pp. 61–72
25. Khan AR, Garcia-Molina H (2017): CrowdDQS: Dynamic question selection in crowdsourcing systems. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, Association for Computing Machinery, pp. 1447–1462
26. Liu X, Lu M, Ooi BC, Shen Y, Wu S, Zhang M: CDAS: A crowdsourcing data analytics system. *Proc VLDB Endow* 5, 1040–1051 (2012)
27. Boutsis I, Kalogeraki V, Guno D (2016): Reliable crowdsourced event detection in smartcities. In: 2016 1st International Workshop on Science of Smart City Operations and Platforms Engineering (SCOPE) in Partnership with Global City Teams Challenge (GCTC), SCOPE - GCTC 2016, Institute of Electrical and Electronics Engineers Inc.
28. Eickhoff C, deVries A: How Crowdsourcable is Your Task. undefined (2011)
29. Difallah DE, Demartini G, Cudré-Mauroux P: Mechanical cheat: Spamming schemes and adversarial techniques on crowdsourcing platforms. *CEUR Workshop Proc* 842, 20–25 (2012)

30. CHIVERS TC, ROGERS WJ, WILLIAMS ME (1974): a Technique for the Measurement of Gas-Leakage.
- 31.: Uncertain Rule-Based Fuzzy Systems - Introduction and New Directions, 2nd Edition | Jerry M. Mendel | Springer, <https://www.springer.com/gp/book/9783319513690>
32. Maksimović M, Vujović V, Perišić B, Milošević V: Developing a fuzzy logic based system for monitoring and early detection of residential fire based on thermistor sensors. *Comput Sci Inf Syst* 12, 63–89 (2015)
33. Khattak HA, Ameer Z, Din IU, Khan MK: Cross-layer design and optimization techniques in wireless multimedia sensor networks for smart cities. *Comput Sci Inf Syst* 16, 1–17 (2019)
34. Romera E, Bergasa LM, Arroyo R (2016): Need data for driver behaviour analysis? Presenting the public UAH-DriveSet. In: *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, Institute of Electrical and Electronics Engineers Inc.*, pp. 387–392
35. Psaraftis K, Anagnostopoulos T, Ntalianis K, Mastorakis N: Customized Recommendation System for Optimum Privacy Model Adoption. *Int J Econ Manag Syst* 03

Konstantinos Psaraftis received the MSc degree from the Electrical and Computer Engineering Department, National Technical University of Athens (NTUA), in 2016. He is currently a PhD candidate in the Department of Business Administration at the University of West Attica. His research interests include location-based services and crowdsourcing, software engineering, and ridesharing applications. He is a member of the IEEE Technical Committee on Hyper-Intelligence since 2021.

Theodoros Anagnostopoulos received the BEng degree in informatics from the Department of Informatics and Engineering, Technical Educational Institution (TEI) of Athens, Greece, in 1997. He also received the BSc degree from the Athens University of Economics and Business (AUEB), Greece, in 2001, and the MSc degree in information systems from the Athens University of Economics and Business (AUEB), Greece, in 2002. He was a visiting PhD student at the University of Geneva (Uni Dufour), Switzerland, in 2007. He received the PhD degree in advanced location prediction techniques in mobile computing from the National and Kapodistrian University of Athens (NKUA), Greece, in 2012. In 2013, he was a postdoctoral researcher in awareness systems at the TEI of Athens, Greece. In 2014, he was a senior postdoctoral researcher in internet of things at the Information Technologies Mechanics and Optics (ITMO) University, Russia. In 2015, he was also a senior postdoctoral researcher in machine learning at the Oulu University, Finland. Currently, he is a principal research scientist in smart cities at the Research & Education at the Ordnance Survey: Britain's Mapping Agency, United Kingdom. He is also a lecturer of internet of things at ITMO University. His current research interests are in the areas of smart cities, internet of things, connected and autonomous vehicles, cyber-security and privacy, artificial intelligence, and awareness systems for geographic information science and systems. He is a member of the ACM, the IEEE Computer and Communications Societies, and the IEEE.

Klimis S. Ntalianis received the Diploma and Ph.D. degrees from the Electrical and Computer Engineering Department, National Technical University of Athens (NTUA), in 1998 and 2003, respectively. Since 1998, he has participated in more than 25 research

and development projects in different frameworks. From 2004 to 2009, he was a Senior Researcher and Projects Coordinator at the Image, Video and Multimedia Laboratory, NTUA. In 2020, he became a Professor at the University of West Attica. He has published more than 160 scientific articles (IEEE, ACM, Springer, and Elsevier) and has received more than 900 citations. He also worked as a Research Evaluator for several international journals and conferences, such as the European Union, the Romanian Executive Agency for Higher Education Research Development and Innovation Funding, the Greek Secretariat of Research and Technology, the Cyprus Promotion Foundation, the Polish National Science Center, the Natural Sciences and Engineering Research Council of Canada, the University of Jeddah (Saudi Arabia), the University of Magdeburg (Germany), the Sant Longowal Institute of Engineering & Technology (India), the Cyprus University of Technology, and other organizations. His main research interests include multimedia analysis, social computing, and new technologies for disruptive business and innovation. He has served as the General Executive Chair for the 3rd IEEE Cyber Science and Technology Congress, the 16th IEEE International Conference on Dependable, Autonomic Secure Computing, the 16th IEEE International Conference on Pervasive Intelligence and Computing, and the 4th IEEE International Conference on Big Data Intelligence and Computing.

Received: August 30, 2021; Accepted: April 03, 2022.

COVID-19 Datasets: A Brief Overview

Ke Sun¹, Wuyang Li¹, Vidya Saikrishna², Mehmood Chadhar³, and Feng Xia^{3,*}

¹ School of Software, Dalian University of Technology,
Dalian 116620, China
{kern.sun,wuyang.li}@outlook.com

² Global Professional School, Federation University,
Ballarat 3353, Australia
v.saikrishna@federation.edu.au

³ Institute of Innovation, Science and Sustainability, Federation University Australia,
Ballarat 3353, Australia
m.chadhar@federation.edu.au
f.xia@ieee.org

Abstract. The outbreak of the COVID-19 pandemic affects lives and social-economic development around the world. The affecting of the pandemic has motivated researchers from different domains to find effective solutions to diagnose, prevent, and estimate the pandemic and relieve its adverse effects. Numerous COVID-19 datasets are built from these studies and are available to the public. These datasets can be used for disease diagnosis and case prediction, speeding up solving problems caused by the pandemic. To meet the needs of researchers to understand various COVID-19 datasets, we examine and provide an overview of them. We organise the majority of these datasets into three categories based on the category of applications, i.e., time-series, knowledge base, and media-based datasets. Organising COVID-19 datasets into appropriate categories can help researchers hold their focus on methodology rather than the datasets. In addition, applications and COVID-19 datasets suffer from a series of problems, such as privacy and quality. We discuss these issues as well as potentials of COVID-19 datasets.

Keywords: COVID-19, Data science, Datasets, Artificial intelligence.

1. Introduction

In late 2019, a novel virus, named COVID-19 emerged all over the world. This virus was declared as a global pandemic by the World Health Organization on March 11, 2020. The COVID-19 has incalculable influences on the world's health, social and economic conditions [54]. With the increase in the number of people infected with COVID-19 every day, it is essential to find a fast and effective way to manage the problems caused by the COVID-19 for biological, medical, and public health issues. Recently, effective utilisation of Artificial Intelligence (AI) technologies [33, 50] to perform analysis, prediction and diagnosis are ongoing researches to fight against coronavirus [2, 47, 59]. We know that AI-based models rely on the available datasets. Thus, datasets play a key role in fighting against the COVID-19 pandemic [34, 37, 79].

* Corresponding author

With the advancement of the Internet and digital media technologies, many open datasets are now available on the websites of research institutions [39]. These datasets are public and can be downloaded for free. For conveniently use of research, this paper provides a summary of the datasets collected from official websites and academic publications. The purpose of our work is to provide researchers, professionals and scholars a quick reference of datasets in application. Based on usages, datasets are organised into several meaningful taxonomies such as, dataset fields, organization structure, and purpose. The categories identified for COVID-19 datasets include: time-series, knowledge and media datasets. We briefly introduce each category of datasets in the following.

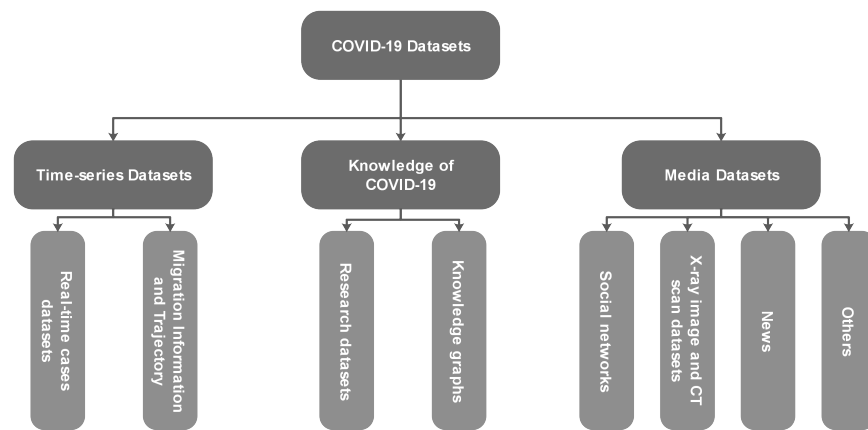


Fig. 1. Taxonomy of COVID-19 open-source datasets

We review the COVID-19 datasets according to the data format and applications. The first category of COVID-19 datasets is the time-series datasets [78]. This kind of datasets is relatively simple in structure and usually contains time information, death statistics, population movement data, etc. Thus, they can be used without complicated preprocessing, such as directly showing the number of cases every day. In addition, this kind of datasets can be used to predict the impact of the disease on human society [6] and trend of virus growth in the future [38].

The second category of datasets is knowledge base of COVID-19, for instance, scholar knowledge graph [36], medical knowledge graph [73]. A part of COVID-19 knowledge bases [46] contain complex contents in datasets, such as text data and image data. Therefore, it is necessary to extract and preprocess the contents to structured data before using the inner information, such as converting the text data into knowledge graphs. In addition, the COVID-19 knowledge base [46] has location information with timestamps and can be used to estimate the risk of infection on people.

The third category of datasets is the media datasets. They have many types of information such as text, image and video. There are lots of media datasets of COVID-19 published on the Internet such as, social networks [8], news information [52]. These information can be used to analyse the current emotional state of people and can be used to

monitor current concerned topics for the public. The taxonomy of COVID-19 datasets are summarized in Fig 1.

The rest of this paper is organized as follows. The time-series datasets are introduced in Section 2. The knowledge base are described in Section 3 and the media datasets are presented in Section 4. In Section 5, several issues of the datasets are discussed followed by conclusions in Section 6.

2. Time-series Datasets of COVID-19

This section is organized to cover information on the source of real-time case datasets and migration datasets around the COVID-19 time. The comparison of several representative datasets is presented in Table 1.

Table 1. Comparison of COVID-19 Time-series Datasets

Dataset	Data type	Level	Main Contents
nCoV2019 [72]	Text, values	Country, province	Symptoms, key dates, and travel history
COVID-19 [13]	Text, values	Country, state, city	Daily case reports, regional coordinates
covid19india-cluster	Text, values	India	Outbreak and transmission COVID19 in India
CoronaWatchNL	Text, values	World	COVID-19 case number, age and sex information
qianxi [†]	Text, map	China	Population migration information of China
2019-nCov-data	Text, values	China	Migration, news, and rumor data
OAG	Text, values	World	Flight schedules data
IATA ^{**}	Text, values	World	Flight schedules data

2.1. Real-time cases datasets

Real-time cases datasets are import for fighting against the COVID-19 disease. For example, these real-time cases datasets could provide people intuitive trend information of COVID-19 and can support public health decision making. To analyse and track the COVID-19 pandemic from real-time cases, Xu et al. [72] collected the real-time cases data from national health reports, and provincial health reports. They also collected information from online reports to supplement their dataset. The dataset contains richer information compared with other real-time case datasets. For example, it has geo-coded information, such as travel history, symptoms, and key dates information (dates of onset, admission, confirmation). In addition, this dataset is being updated in real-time and can be downloaded for free.

Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE) published the Johns Hopkins epidemiological dataset, including daily case reports and time-series summary tables [13]. The dataset has data fields, including country names, state names, place names, the time of last update, and regional longitude. It is available for the public and can be downloaded from the Github repository. Several studies were conducted based on this dataset. For example, Domenico et al. [5] utilized a variant regressive model to predict the epidemiological trend of COVID-19. Punn et al. [48] developed

a deep learning-based model to monitor people's behaviour every day. The model could utilize the real-time dataset of the Johns Hopkins dashboard to predict the trend of the COVID-19 spreading across nations in the future. Tátrai et al. [58] used the same dataset to investigate how well the logistic equation can predict the influence of the COVID-19 pandemic on the place where the outbreak occurred. The proposed model was used to estimate the risky point, the date of reaching a certain percentage of infections and the number of infected persons in the future.

Several official institutions recorded real-time data and published them to the public. For example, the National Health Commission of the People's Republic of China daily publishes the latest cases information COVID-19 on the official website[†]. Roda et al. [51] used the data to predict the cases of COVID-19 in Wuhan city after lockdown. Singapore collected and published the real-time cases' information of COVID-19 on the Ministry of Health official website[‡]. In addition, the Singapore reports cover detailed analysis of the COVID-19 data. The Tianjin Health Commission daily published the local COVID-19 cases in the form of online press release on their official website[§]. The detailed report along with analysis can be obtained from the same official website.

2.2. Migration information and trajectory

Population migration influences on the spread of virus. To track and discover their relations, it is necessary to record the population migration for studying the trend of virus transmission. There are many open datasets to track the COVID-19 transmission shared on the Internet. A well known migration dataset is available on the Baidu Migration site[¶]. This data can be used to study the pattern of population migration during the Spring Festival of China [27,71]. In addition, researchers can utilize this data to visualize population migration around China. In this dataset, "qianxi" index is used to reflect the size of population moving in or out, and the cities can be compared horizontally. The intensity of city travel is calculated as the ratio of the number of people travelling to the city to the resident population in the same city. In addition, Baidu built a data federation platform (Baidu FedCube), which provides usage instructions and data download services.

Migration datasets of confirmed COVID-19 patients usually contain the travel information including start time, end time, travel type, number of trips, travel description, departure station, arrival station, and other information of the confirmed patients. Several studies were conducted based on migration datasets. For instance, to estimate the geographical scope of the spread of COVID-19 and its potential risks, Lai et al. [26] proposed a deep learning-based model, which could learn from population migration dataset and give future prediction results. Huang et al. [20] attempted to utilize the nationwide mobility data to study the economic impact caused by the COVID-19.

Several trajectory data of proprietary airline are commercially available, such as Official Airlines Guide (OAG) database^{||}, International Air Travel Authority (IATA) database^{**}. The IATA database contains about 90% of passenger information of commercial flights,

[†] <https://www.nhc.gov.cn/>

[‡] <https://www.moh.gov.sg/>

[§] <https://wsjk.tj.gov.cn/>

[¶] <https://qianxi.baidu.com/>

^{||} <https://www.oag.com/>

^{**} <https://www.iata.org>

including the direct starting point from Wuhan to the destination and the indirect starting point from Wuhan with a connecting flight to the final destination [26]. Chinazzi et al. [9] utilized a global aggregate population disease transmission model and proprietary airline data to predict the impact of travel restrictions on the spread of the COVID-19. Proprietary airline data can be used to evaluate the capacity to detect the COVID-19 of different locations. For instance, Rene et al. [44] utilized a Bayesian-based model and the proprietary airline data to estimate the capacity of 194 locations. In addition, the authors designed a mathematical model to calculate the rate of local residents being infected by foreign tourists.

Descartes Labs collected and released DL-COVID-19 dataset, which is mobility dataset at state and county level of US [67]. The dataset was published on the GitHub repository under the Creative Commons Attribution license. Based on the DL-COVID-19 dataset, Michael et al. [67] have found significant changes in the flow of people caused by COVID-19 in US and around the world through mobility data in US.

Transportation Security Administration (TSA) published a confirmed COVID-19 cases dataset^{††}. They notified the public about the airport locations where the employees belonging to TSA were found positive for the COVID-19 virus. TSA listed airports with confirmed COVID-19 cases and also the corresponding employees inflicted by the virus.

3. Knowledge Bases of COVID-19

This section is divided into two subsections. The first subsection introduces research datasets of COVID-19 from knowledge point of view. The second subsection deals with the knowledge graphs and their importance. Table 2 gives the summary of these datasets.

Table 2. Summary of COVID-19 knowledge bases

Dataset	Type	Size	Main Contents
CORD-19 [63]	Article	128,000 articles	Coronaviruses related
CORD-NER [66]	Text	75 entity types	Entity types related to the COVID-19
COVID-19-epidemiology ^{‡‡}	Knowledge graph	374 instances	Epidemiological knowledge
covid19kg [12]	Knowledge graph	4016 nodes, 10 entity types	Virus protein, potential drug target, etc
covid-19-medical ^{‡‡}	Knowledge graph	383 instances	Clinical knowledge

3.1. Research dataset for COVID-19

Since the outbreak of COVID-19, many research papers for the study and analysis of the new coronavirus have surged, especially in the fields of medicine and biology. CORD-19 [63] is one of a extensive machine-readable and large new coronavirus paper collection for data mining to date, which contains historical and the latest scientific research papers of the coronavirus. The CORD-19 was provided by the leading research groups of Semantic Scholars at Allen AI. The dataset has a collection of more than 128,000 academic

^{††} <https://www.tsa.gov/>

articles and 59,000 full texts on new coronaviruses, containing articles related to such as, COVID-19, SARS (Severe Acute Respiratory Syndrome), and MERS (Middle East Respiratory Syndrome). The dataset contains more than 50,000 metadata files of coronavirus research articles, including but not limited to COVID-19.

CORD-19 is an open knowledge research dataset, and it is free for use by the global research community. For instance, the worldwide AI research community utilized this dataset and data mining methods to fight against the COVID-19. Since its release, CORD-19 has been downloaded more than 75,000 times. It becomes the basis of many COVID-19 text mining and discovery systems and could promote generating new insights into the fight against the ongoing COVID-19. In addition, the dataset has links to other publication databases such as PubMed, Microsoft Academic Graph, Semantic Scholar, and WHO through unique keywords. Thus, CORD-19 has richer information than other knowledge bases.

At present, CORD-19 has been used in information extraction, information retrieval and knowledge graphs by natural language processing and deep learning techniques. Besides, it can also be used in multiple directions, such as question answering [43], pre-trained language models [31], summarization [55], and recommendations [53]. To inspire developers to find new insights in the large-scale COVID-19 epidemic, Kaggle utilized the CORD-19 dataset to host an open research dataset challenge. The research challenge includes the topic of tracing the history of the virus [35], the study of the transmission characteristics of the virus, the diagnosis of the virus, etc.

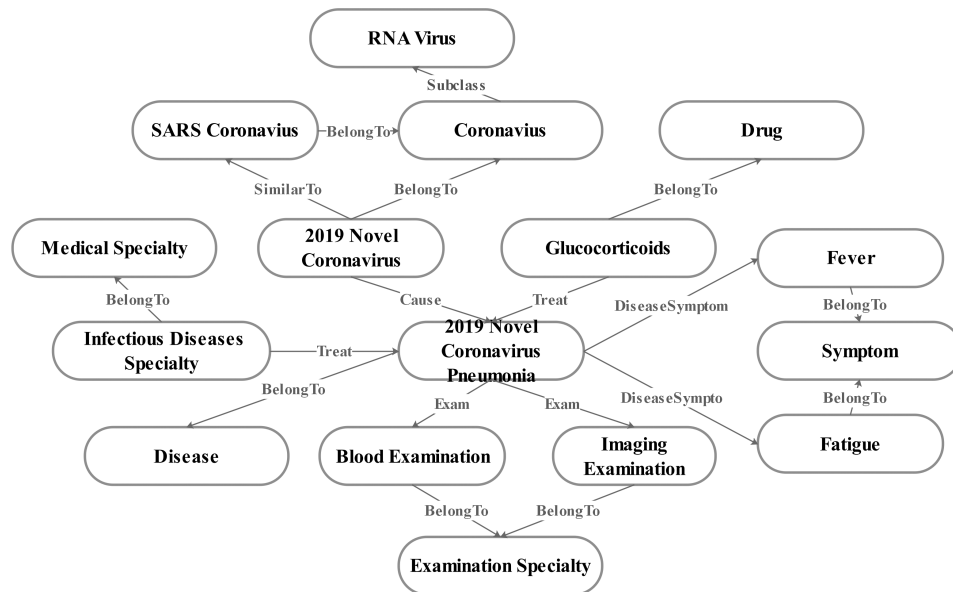


Fig. 2. A medical knowledge graph of COVID-19. This knowledge graph contains entities such as viruses, bacteria, epidemics, and infectious diseases. Entities connected to each other with relations, such as “Subclass”, “SimilarTo”, and “Cause”

3.2. COVID-19 knowledge graphs

Knowledge graphs associated with COVID-19 attracts lots of attention of researchers. Many kinds of knowledge graphs are published such as, pathophysiology knowledge graph [12] and biomedical knowledge graphs [1, 81]. Knowledge graphs are useful in a wide range of applications, for instance, knowledge answer system of COVID-19 [28], auxiliary diagnostic system of COVID-19 [30].

Directly getting desired knowledge from lots of research articles is relative time-consuming. An efficient approach is to obtain knowledge from pre-constructed knowledge graphs. Knowledge graphs utilize topology to integrate data and cover a wide range of knowledge. For instance, the COVID-19 knowledge graphs contains biological processes, drug-target interactions, genes and proteins of the new coronavirus [16]. Thus, the COVID-19 knowledge graph can even help researchers discover hidden interactions of protein.

Metrological analysis and visualization of knowledge graph methods can help extract and formalize structured knowledge. For example, network embedding theories [19, 56] and graphic visualization technologies can be used to visualise knowledge graphs. A visualization of a medical knowledge of COVID-19^{‡‡} is presented in Fig 2. The visualization of knowledge graphs can provide intuitive connections between entities and can promote researchers to better understand knowledge in less amount of time. Moreover, in the visual web application of knowledge graphs, users can browse and query the network, filter nodes or edges according to their own needs, or calculate the path between nodes of interest.

There are several knowledge graphs built from COVID-19 [63]. The literature of medicine and biology are the main contents of this dataset. Several natural language processing methods were applied to this dataset to construct knowledge graphs. The constructed knowledge graph contains information on medicine and biology, which are important for researchers [81]. In addition, the rich relations in knowledge graphs could help researchers to discover hidden information and thus contribute to fight against COVID-19. To apply COVID-19 to the field of knowledge graphs, the first work is to create a named entity recognition (NER) dataset. Wang et al. [66] built an entity recognition dataset for COVID-19. The NER dataset contains 4 different sources. The entities in the NER dataset are mapped into 75 fine-grained entity types. The results obtained by identifying named entities on the COVID-19 dataset helps in constructing knowledge graphs. For example, it can be applied to build medical knowledge graphs.

Several other researchers devoted themselves to constructing knowledge graphs of COVID-19. Domingo-Fernández et al. [12] constructed an extensive a knowledge graph based on the COVID-19 paper collection. The knowledge graph contains information of the COVID-19 virus protein, potential drug target and the biological transmission path of the virus. The knowledge graph could provide a new research perspective for exploring the physiology of COVID-19 cases. First, the authors filtered the unimportant information from the available source. Second, they collected a part of free and open scientific articles related to COVID-19. Then, the collected articles were scored and ranked by using modelling language tools based on importance. Finally, the knowledge graph was constructed from the selected articles.

^{‡‡} <https://openkg.cn/>

The integration of large-scale knowledge graphs and information mining functions are urgently needed for filtering plenty of new coronaviruses, especially in the field of medicine. The drug knowledge graphs can help medical researchers quickly find potential drug candidates. For instance, Ge et al. [16] designed a knowledge graph building method, which is a data-driven drug framework. The built knowledge graph is virus related, including knowledge of drug-target, protein-protein interactions. Three different types of nodes exist in the knowledge graph, namely drugs, human targets and viral targets. Entities in knowledge graph are connected with edges, which describe the relationship, similarities and interaction between entities. A total of seven networks are considered to construct the knowledge graph, including human protein-protein interaction network, human target-drug interaction network, and so on.

Network embedding algorithms [70] are well known for network analysis. It can be applied to knowledge graphs to predict the drug candidate list, saving the time and cost of discovering effective drugs for disease. For instance, Hong et al. [18] proposed a relation extraction method based on deep learning technology, namely BERE. The method can be applied to mining large-scale literature. Relying on this method, only a small number of candidate drugs on the list need to be manually checked, thus, the list of candidate drugs is further narrowed.

4. COVID-19 Related Media Datasets

This section covers datasets collected from social networks, news and other media sources in three different sub-sections. These datasets are summarized in Table 3.

Table 3. Summary of COVID-19 related media Datasets

Dataset	Format	Contents
COVID19socialscience [75]	Text	Tweet of 69 institutional/media Twitter accounts
covid19twitterevent [83]	Text, JSON	COVID-19 Events from Twitter
covid19twitter [4]	Text, JSON	Twitter chatter of COVID-19
CoronaVis [22]	Text, JSON	Personal opinions, facts, news, status
COVID-19-TweetIDs [8]	Text	50 million tweets
COVID-19-InstaPostIDs [77]	Text	Public posts from Instagram
covid19_dataset [15]	TSV	Tweet, user ID and Weibo ID
COVID-CT [82]	Text, xls,image	CT scans from medRxiv, bioRxiv, etc
covid-chestxray-dataset [10]	Text, csv, image	Chest images of COVID-19 or other pneumonias
COVID-19 [45]	Image	Normal, pneumonia chest images

4.1. Social networks

Many datasets for social networks [69] are published such as, Twitter and Facebook datasets. These datasets can be used to support urgent research to address the outbreaks caused by COVID-19. Considering that there is no specialized collection of tweets posted

by the government or news media, Yu [75] published a COVID-19 Twitter dataset for social science research, which is built on the keywords of coronavirus and COVID-19.

Department of Social Psychology, Universitat Autònoma de Barcelona published an Institutional and News Media Tweet dataset of COVID-19 for social science research [75]. The dataset was obtained from Twitter accounts of 69 institutions/news media, including 17 government and international organizations and 52 news media in North America, Europe, etc. There are 8 categories in the collection: “Government Tweets” (government, international agencies, etc.), “US News Tweets”, “British News Tweets”, “Spanish News Tweets”, “German News Tweets”, “France News Tweets”, “China News Tweets”, and “Additional News Tweets”. Each category contains different collection targets. This microblogging data can support sociologists to analyse the impact of the pandemic on public interest, health information, and social response to policy-makers [14].

The Department of Computer Science at the University of Missouri published the coronaVirus Twitter (focused on the United States) dataset [22]. They collected and processed more than 100 million tweets related to the novel coronavirus using Twitter Streaming API and Tweepy since March 5, 2020. The collected raw data around 700GB up until April 24, 2020, and saved these collected data in the format of JSON. To improve the usability of data, the dataset has been dynamically processed in real-time, which is stored and being updated in the Github repository. Every single file in dataset contains intra-day data. Date is set as the single file name. The file in the dataset contains 6 different attributes (tweet_id, created_at, loc, text, user_id, verified). The tweet_id represents the unique id of a tweet. The created_at represents the creation time of a tweet. The loc represents the state user location. The text represents the text of the tweet being processed, with all the text in lowercase, non-English characters, and some stop words removed. The user_id means that the exact user name of the pseudo-user ID is converted to an anonymous ID to protect the user’s privacy. The verified field indicates whether the tweet is verified (1 or 0), 0 means unverified, 1 means verified. During the pandemic, people were isolated at home, but social media allowed people around the world to stay connected. Collecting information of people shared on social media, such as personal opinions, status and location, can help researchers understand public behaviour during a pandemic. This dataset can be used for such as, sentiment analysis [41, 42], behavioural decision-making [76].

Another public Twitter dataset [8] related to coronavirus was collected by the Information Sciences Institute, University of Southern California. The dataset has more than 50 million tweets from the inception until March 16, 2020, about 450 GB of raw data. The dataset could be used to track scientific coronavirus misinformation and unverified rumours, and help researchers to understand the fear and panic of the public [80]. There is another Twitter dataset of COVID-19 for scientific research [?]. This open dataset enables researchers to carry out research projects on emotional and psychological responses to social distance measurements, identification of false sources, and stratified measurements of pandemic emotions.

The first Instagram dataset for COVID-19 was collected by researchers at Queen Mary University of London, England [77]. The dataset is published on a Github repository. The dataset contains four main parts: (1) publisher information content, (2) post content, (3) like features, and (4) comment metrics content. The posts content part has key attributes, such as captions, hashtag lists, images/videos, likes, comments, locations, dates, and tagged lists. Posters can be public accounts (or public Instagram pages) and datasets

contains information about individuals, fan pages, news agencies, influencers, bloggers, and so on. Each post receives a response, such as a comment issued by a viewer/follower. The dataset helps researchers study the analysis of fake news, false alarms, rumours, the robot population and robot-generated content, and behavioural changes during the spreading of the COVID-19 pandemic.

Georgetown University built a Tweets dataset of COVID-19 using Twitter's Streaming API (Twitter Streaming API) [15]. Tweets related to COVID-19 are defined as tweets with 16 tags, such as 2019nCoV Corona SARI. The dataset has 2,792,513 tweets, 456,878 quotes, and 18,168,161 retweets. Most tweets in dataset were in English (57.1%), followed by tweets in Spanish (11.6%). The dataset was divided into two parts: one was grouped according to the location information in the tweet's content and the other was grouped according to the location information based on the time of the tweet. More than 351,000 tweets in the data have links to news organizations, which accounted for about a tenth of the original tweets of the samples. Researchers found that more than 63,000 tweets were linked to high-quality sources and more than 1,000 were linked to low-quality sources. Currently, the dataset has been used to find a correlation between the virus outbreak and the activity level of local social media. Although rumours and low-quality information still exist, they have little impact on general trends such as the direction of public opinion. The dataset can also be used by natural language processing models for more sophisticated spatiotemporal analysis of information flows and the spread of COVID-19, aiming at identifying rumours and topics [61].

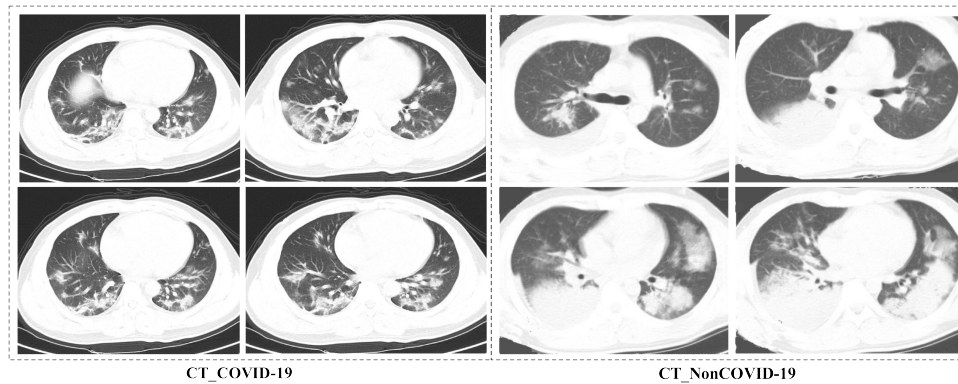


Fig. 3. Examples of CT scans for COVID-19. The left images are the CT scans of a patient's chest infected with COVID-19. The right images are the CT scans of the chest of a normal person

4.2. X-ray image and CT scan datasets

Medical images are important sources to diagnose COVID-19, such as chest CT scans and X-ray images. To improve the efficiency of medical diagnosis, many institutions attempt to automatically diagnosis this disease by exploiting deep learning algorithms and

COVID-19 medical image datasets. A part of institutions has shared the collected COVID-19 medical image datasets on the Internet to promote the research of artificial intelligence in pneumonia diagnosis. In the following, we present these datasets in this part.

Many medical images of COVID-19 can be obtained from GitHub repository or the official website. For example, the COVID_CT dataset [82] was published on GitHub repository. Examples of CT scans for COVID-19 of this dataset are shown in Fig 3. This dataset was collected from COVID-19 related papers published in several open databases such as, medRxiv, bioRxiv. The authors collected images from papers by matching the pneumonia name of image titles. This dataset contains normal and infected CT scans. The subset of COVID_CT contains 349 images from the clinical findings of 216 patients. The two categories of images are stored separately in two different files. Cohen et al. [10] collected open sources and diagnostic data from hospitals and created a public image dataset of COVID-19. This dataset contains chest images of COVID-19 or other types of pneumonia (MERS, SARS and ARDS) of positive or suspected patients. The dataset contains COVID-19 chest X-rays images of 412 people from 26 countries/regions, including 679 images. In addition, this dataset contains clinical records of patients such as blood tests, ICU stay. Ozturk et al. [45] created an integrated COVID-19 X-ray dataset from a part of two open-source datasets, including the dataset published by Cohen JP [10] and the ChestX-ray8 dataset provided by Wang et al. [65]. The integrated dataset covers normal, pneumonia chest images and COVID-19 images of people, containing diagnostic images of 43 female and 82 male patients, and a total of 127 COVID-19 chest images. In addition, the dataset contains the age information of 26 patients. Wang et al. [62] created the largest publicly accessible COVID-19 chest X-ray image dataset, namely COVIDx. This dataset was collected from five open-access data repositories and contains a total of 13,725 patient cases and 13,800 chest images. Several studies [21, 60, 68] have conducted AI-based diagnosis with the help of COVIDx. The authors [64] collected a total of 259 chest CT scans from several hospitals in China and 15 recruited patients. The dataset contains 180 typical viral pneumonia and 79 confirmed SARS-COV-2 cases. This dataset can be obtained from the supplementary data of the study.

4.3. News

News media data serves as a convenient and direct data for the public [74]. People can obtain the exact situation of the pandemic in the current region, and take corresponding measures to protect themselves. We can use the existing data to predict the follow-up development of the epidemic in each region. For instance, researchers can use this kind of dataset to apply statistical analysis to compare pandemic features among different countries, and try to find features that can bring new insights to fight the pandemic [17]. The dataset can also be used with other kinds of datasets, such as regional and subregional social demographic [24].

4.4. Others

A question-and-answer system was developed to obtain the CovidQA dataset [40] to help the research community find answers and gain insight into coronavirus infectious diseases [57]. The CovidQA dataset is for questions and answers about COVID-19 and is

built from the source from Kaggle's COVID-19 open research dataset challenge. The CovidQA dataset is the first publicly available COVID-19 Q&A dataset, which contains 124 question-article pairs according to the CovidQA dataset. The current version is 0.1, and the database will be further expanded as resources get evaluated. The dataset includes fields such as subcategory, title, answer. Title is the title of the scientific research article, of which the answer is derived or the title of the announcement issued by an authoritative institution to verify the reliability of the answer. The datasets were reviewed by epidemiologists, MDs (Medical Practitioners), and medical students. The dataset can be used in the natural language processing model (NLP) field to test the validity of the model. In addition, the dataset can be used to build the deep learning-based question and answer model, which consists of two main parts: the question context component and the answer component [29].

University College London recently published a dataset called RWWD (Real World Worry Dataset) [25]. The dataset utilizes a direct questionnaire approach to obtain written descriptions of how people feel about COVID-19 and their current emotions. Instead of relying on a third party to annotate, the dataset relies on the writer's ratings of their own mood after writing, which makes the dataset more reliable. RWWD has two versions, and each version has 2,500 English texts. The first version is of variable length, with a minimum of 500 words. The second version uses the Twitter format (i.e., no more than 240 words long), and the short text is mainly used for comparison with the Twitter data. All subjects were asked to use a 9-point scales to indicate their internal emotions including worry, anger, etc. The study results on this dataset showed that Britons were more worried about their families and finances. Short texts (in the form of tweets) tend to be inspirational and chants, while long texts prefer to express their inner emotions, for example, people's concerns about the epidemic. This dataset has been used to measure changes in the mood of citizens during the COVID-19 outbreak.

5. Discussions

More and more datasets related to COVID-19 pandemic are emerging gradually over time. However, only a part of datasets are helpful for researchers as most the published datasets for COVID-19 analysis or treatments tend to be incomplete, possibly biased, and limited to national samples. Especially, problems existing in COVID-19 datasets such as incompleteness [23] and small scale [29] are urgent problem for research. Thus, how to obtain valuable datasets is still a challenging task for researchers. For the sake of effective COVID-19 research, we still need more valuable datasets by adopting appropriate processing and collection methods.

Nevertheless, the presented datasets have significant implications to fight against COVID-19. For instance, real-time datasets can be used for contact tracing and finding out the influence scope. These datasets can help publish reasonable policy of lockdown. Similar, these datasets can be used to identify potential risk points such as, places for common public interests. The listed datasets can have practical implications when integrated with the other datasets. For instance, real-time datasets can provide information regarding the impact of diseases on human life. These datasets can be used with other datasets including temperature and humidity. These factors appear to influence the COVID-19 effects on

human lives [49]. Future studies can use these factors to reveal their moderating effect between COVID-19 and human society regarding deaths and population movements.

Companies can use knowledge bases to develop strategies to fight against diseases. The listed knowledge bases are a valuable tool to extract practical knowledge, experience and facts to formulate policies for the businesses. For instance, companies can use these datasets to publish economic policy recommendations as small and medium businesses are heavily impacted by diseases. The business survival depends on the new business models and how quickly these models are adopted. Therefore, new strategies are inevitable for governments and local businesses to endure this unpredicted scenario [7].

Researchers can use the media datasets to publish policies to counter fake news regarding the COVID-19. Online platforms are flooded with unauthentic misinformation such as the negative impacts of the COVID-19 vaccine and the dangerous nature of the virus [32]. This fabricated news can make negative impact on the efforts to counter the disease. Firstly, people might not take the required precautions and consider the COVID-19 a conspiracy [3], as witnessed in several anti-lockdown protests. Second, people might resist receiving vaccinations based on social media fake news. This is also evident with the slow progress of vaccination in several countries.

Social media plays a critical role in addressing the issue of misinformation. It can also be a valuable tool to provide relevant and authentic information for patients, doctors and clients. Therefore, better social media approaches and strategies are required for social media to play an influential role in policy and decision making for government, organizations and individuals [11]. The listed media datasets can provide researchers with a platform to generate recommendations of such policies and strategies.

6. Conclusion

This paper presents several key sources of COVID-19 datasets of different categories including time-series datasets (real-time cases datasets, migration information datasets), knowledge base (knowledge graphs, research dataset), and media datasets (social networks, X-ray image and CT scan datasets, news, and others). Then, we discuss how various organisations gather the data. Most of the datasets examined in this paper are publicly available. We provide relevant links to the datasets wherever possible. We also discussed several drawbacks of the current COVID-19 datasets. An efficient COVID-19 dataset evaluation procedure is also missing. We suggest building a more effective mechanism for collecting more valuable data for research.

Acknowledgments. The authors would like to thank Xiangtai Chen, Huazhu Cao, Mengyuan Wang, and Xu Feng from Dalian University of Technology for their help with the first draft of this paper.

References

1. Al-Saleem, J., Granet, R., Ramakrishnan, S., Ciancetta, N.A., Saveson, C., Gessner, C., Zhou, Q.: Knowledge graph-based approaches to drug repurposing for covid-19. *Journal of Chemical Information and Modeling* 61(8), 4058–4067 (2021)

2. Albahri, A., Hamid, R.A., Alwan, J.K., Al-Qays, Z., Zaidan, A., Zaidan, B., Albahri, A., AlAmoodi, A., Khlaf, J.M., Almahdi, E., et al.: Role of biological data mining and machine learning techniques in detecting and diagnosing the novel coronavirus (covid-19): a systematic review. *Journal of Medical Systems* 44, 1–11 (2020)
3. Apuke, O.D., Omar, B.: Fake news and covid-19: modelling the predictors of fake news sharing among social media users. *Telematics and Informatics* 56, 101475 (2021)
4. Banda, J.M., Tekumalla, R., Wang, G., Yu, J., Liu, T., Ding, Y., Artemova, E., Tutubalina, E., Chowell, G.: A large-scale covid-19 twitter chatter dataset for open scientific research—an international collaboration. *Epidemiologia* 2(3), 315–324 (2021), <https://www.mdpi.com/2673-3986/2/3/24>
5. Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., Ciccozzi, M.: Application of the arima model on the covid-2019 epidemic dataset. *Data in Brief* p. 105340 (2020)
6. Cao, W., Fang, Z., Hou, G., Han, M., Xu, X., Dong, J., Zheng, J.: The psychological impact of the covid-19 epidemic on college students in china. *Psychiatry Research* p. 112934 (2020)
7. Carracedo, P., Puertas, R., Marti, L.: Research lines on the impact of the covid-19 pandemic on business. a text mining analysis. *Journal of Business Research* 132, 586–593 (2021)
8. Chen, E., Lerman, K., Ferrara, E., et al.: Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set. *JMIR Public Health and Surveillance* 6(2), e19273 (2020)
9. Chinazzi, M., Davis, J.T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., y Piontti, A.P., Mu, K., Rossi, L., Sun, K., et al.: The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. *Science* 368(6489), 395–400 (2020)
10. Cohen, J.P., Morrison, P., Dao, L., Roth, K., Duong, T.Q., Ghassemi, M.: Covid-19 image data collection: Prospective predictions are the future. *arXiv* 2006.11988 (2020), <https://github.com/ieee8023/covid-chestxray-dataset>
11. Cuello-Garcia, C., Pérez-Gaxiola, G., van Amelsvoort, L.: Social media can have an impact on how we manage and investigate the covid-19 pandemic. *Journal of Clinical Epidemiology* 127, 198–201 (2020)
12. Domingo-Fernández, D., Baksi, S., Schultz, B., Gadiya, Y., Karki, R., Raschka, T., Ebeling, C., Hofmann-Apitius, M., Kodamullil, A.T.: Covid-19 knowledge graph: a computable, multi-modal, cause-and-effect knowledge model of covid-19 pathophysiology. *Bioinformatics* 37(9), 1332–1334 (2021)
13. Dong, E., Du, H., Gardner, L.: An interactive web-based dashboard to track covid-19 in real time. *The Lancet Infectious Diseases* 20(5), 533–534 (2020)
14. Ferreira, C.M., Sá, M.J., Martins, J.G., Serpa, S.: The covid-19 contagion–pandemic dyad: A view from social sciences. *Societies* 10(4), 77 (2020)
15. Gao, Z., Yada, S., Wakamiya, S., Aramaki, E.: Naist covid: Multilingual covid-19 twitter and weibo dataset. *arXiv preprint arXiv:2004.08145* (2020)
16. Ge, Y., Tian, T., Huang, S., Wan, F., Li, J., Li, S., Yang, H., Hong, L., Wu, N., Yuan, E., et al.: A data-driven drug repositioning framework discovered a potential therapeutic agent targeting covid-19. *BioRxiv* (2020)
17. Hamzah, F.B., Lau, C., Nazri, H., Ligot, D.V., Lee, G., Tan, C.L., Shaib, M., Zaidon, U.H.B., Abdullah, A.B., Chung, M.H., et al.: Coronatracker: worldwide covid-19 outbreak data analysis and prediction. *Bull World Health Organ* 1(32), 1–32 (2020)
18. Hong, L., Lin, J., Tao, J., Zeng, J.: Bere: An accurate distantly supervised biomedical entity relation extraction network. *arXiv preprint arXiv:1906.06916* (2019)
19. Hou, M., Ren, J., Zhang, D., Kong, X., Zhang, D., Xia, F.: Network embedding: Taxonomies, frameworks and applications. *Computer Science Review* 38, 100296 (2020)
20. Huang, J., Wang, H., Xiong, H., Fan, M., Zhuo, A., Li, Y., Dou, D.: Quantifying the economic impact of covid-19 in mainland china using human mobility data. *arXiv preprint arXiv:2005.03010* (2020)

21. Jaiswal, A., Gianchandani, N., Singh, D., Kumar, V., Kaur, M.: Classification of the covid-19 infected patients using densenet201 based deep transfer learning. *Journal of Biomolecular Structure and Dynamics* pp. 1–8 (2020)
22. Kabir, M., Madria, S., et al.: Coronavis: A real-time covid-19 tweets analyzer. *arXiv preprint arXiv:2004.13932* (2020)
23. Karlinsky, A., Kobak, D.: Tracking excess mortality across countries during the covid-19 pandemic with the world mortality dataset. *Elife* 10, e69336 (2021)
24. Karmakar, M., Lantz, P.M., Tipirneni, R.: Association of social and demographic factors with covid-19 incidence and death rates in the us. *JAMA Network Open* 4(1), e2036462–e2036462 (2021)
25. Kleinberg, B., van der Vegt, I., Mozes, M.: Measuring emotions in the covid-19 real world worry dataset. In: *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020* (2020)
26. Lai, S., Bogoch, I.I., Ruktanonchai, N.W., Watts, A., Lu, X., Yang, W., Yu, H., Khan, K., Tatem, A.J.: Assessing spread risk of wuhan novel coronavirus within and beyond china, january-april 2020: a travel network-based modelling study. *MedRxiv* (2020)
27. Lai, H., et al.: Changingepidemiologyofhug man brucellosis, china, 1955g2014. *Emerg Infect Dis* 23(2), 184 (2017)
28. Lee, J., Sean, S.Y., Jeong, M., Sung, M., Yoon, W., Choi, Y., Ko, M., Kang, J.: Answering questions on covid-19 in real-time. In: *Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020* (2020)
29. Levy, S., Mo, K., Xiong, W., Wang, W.Y.: Open-domain question-answering for covid-19 and other emergent domains. *arXiv preprint arXiv:2110.06962* (2021)
30. Li, X., Geng, M., Peng, Y., Meng, L., Lu, S.: Molecular immune pathogenesis and diagnosis of covid-19. *Journal of Pharmaceutical Analysis* (2020)
31. Lin, J., Nogueira, R., Yates, A.: Pretrained transformers for text ranking: Bert and beyond. *Synthesis Lectures on Human Language Technologies* 14(4), 1–325 (2021)
32. van der Linden, S., Roozenbeek, J., Compton, J.: Inoculating against fake news about covid-19. *Frontiers in Psychology* 11, 2928 (2020)
33. Liu, J., Kong, X., Xia, F., Bai, X., Wang, L., Qing, Q., Lee, I.: Artificial intelligence in the 21st century. *IEEE Access* 6, 34403–34421 (2018)
34. Liu, J., Kong, X., Zhou, X., Wang, L., Zhang, D., Lee, I., Xu, B., Xia, F.: Data mining and information retrieval in the 21st century: A bibliographic review. *Computer Science Review* 34, 100193 (2019)
35. Liu, J., Nie, H., Li, S., Chen, X., Cao, H., Ren, J., Lee, I., Xia, F.: Tracing the pace of covid-19 research: Topic modeling and evolution. *Big Data Research* 25, 100236 (2021)
36. Liu, J., Ren, J., Zheng, W., Chi, L., Lee, I., Xia, F.: Web of scholars: A scholar knowledge graph. In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 2153–2156 (2020)
37. Liu, J., Tian, J., Kong, X., Lee, I., Xia, F.: Two decades of information systems: a bibliometric review. *Scientometrics* 118(2), 617–643 (2019)
38. Mandal, M., Jana, S., Nandi, S.K., Khatua, A., Adak, S., Kar, T.: A model based study on the dynamics of covid-19: Prediction and control. *Chaos, Solitons & Fractals* p. 109889 (2020)
39. Mohamadou, Y., Halidou, A., Kapen, P.T.: A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of covid-19. *Applied Intelligence* pp. 1–13 (2020)
40. Möller, T., Reina, A., Jayakumar, R., Pietsch, M.: Covid-qa: A question answering dataset for covid-19. In: *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020* (2020)
41. Naseem, U., Razzak, I., Khushi, M., Eklund, P.W., Kim, J.: Covidsent: A large-scale benchmark twitter data set for covid-19 sentiment analysis. *IEEE Transactions on Computational Social Systems* (2021)
42. Nemes, L., Kiss, A.: Social media sentiment analysis based on covid-19. *Journal of Information and Telecommunication* pp. 1–15 (2020)

43. Ngai, H., Park, Y., Chen, J., Parsapoor, M.: Transformer-based models for question answering on covid19. *arXiv preprint arXiv:2101.11432* (2021)
44. Niehus, R., De Salazar, P.M., Taylor, A.R., Lipsitch, M.: Using observational data to quantify bias of traveller-derived covid-19 prevalence estimates in wuhan, china. *The Lancet Infectious Diseases* (2020)
45. Ozturk, T., Talo, M., Yildirim, E.A., Baloglu, U.B., Yildirim, O., Acharya, U.R.: Automated detection of covid-19 cases using deep neural networks with x-ray images. *Computers in Biology and Medicine* 121, 103792 (2020)
46. Pepe, E., Bajardi, P., Gauvin, L., Privitera, F., Lake, B., Cattuto, C., Tizzoni, M.: Covid-19 outbreak response, a dataset to assess mobility changes in italy following national lockdown. *Scientific Data* 7(1), 1–7 (2020)
47. Prakash, K.B., Imambi, S.S., Ismail, M., Kumar, T.P., Pawan, Y.: Analysis, prediction and evaluation of covid-19 datasets using machine learning algorithms. *International Journal* 8(5) (2020)
48. Punna, N.S., Sonbhadra, S.K., Agarwal, S.: Covid-19 epidemic analysis using machine learning and deep learning algorithms. *MedRxiv* (2020)
49. Qi, H., Xiao, S., Shi, R., Ward, M.P., Chen, Y., Tu, W., Su, Q., Wang, W., Wang, X., Zhang, Z.: Covid-19 transmission in mainland china is associated with temperature and humidity: A time-series analysis. *Science of The Total Environment* 728, 138778 (2020)
50. Ren, J., Xia, F., Chen, X., Liu, J., Hou, M., Shehzad, A., Sultanova, N., Kong, X.: Matching algorithms: Fundamentals, applications and challenges. *IEEE Transactions on Emerging Topics in Computational Intelligence* 5(3), 332–350 (2021)
51. Roda, W.C., Varughese, M.B., Han, D., Li, M.Y.: Why is it difficult to accurately predict the covid-19 epidemic? *Infectious Disease Modelling* (2020)
52. Shahi, G.K., Nandini, D.: Fakecovid—a multilingual cross-domain fact check news dataset for covid-19. *arXiv preprint arXiv:2006.11343* (2020)
53. Shen, I., Zhang, L., Lian, J., Wu, C.H., Fierro, M.G., Argyriou, A., Wu, T.: In search for a cure: recommendation with knowledge graph on covid-19. In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. pp. 3519–3520 (2020)
54. Sohrabi, C., Alsafi, Z., O’Neill, N., Khan, M., Kerwan, A., Al-Jabir, A., Iosifidis, C., Agha, R.: World health organization declares global emergency: A review of the 2019 novel coronavirus (covid-19). *International Journal of Surgery* (2020)
55. Su, D., Xu, Y., Yu, T., Siddique, F.B., Barezi, E.J., Fung, P.: Caire-covid: a question answering and query-focused multi-document summarization system for covid-19 scholarly information management. *arXiv preprint arXiv:2005.03975* (2020)
56. Sun, K., Wang, L., Xu, B., Zhao, W., Teng, S.W., Xia, F.: Network representation learning: From traditional feature learning to deep learning. *IEEE Access* 8, 205600–205617 (2020)
57. Tang, R., Nogueira, R., Zhang, E., Gupta, N., Cam, P., Cho, K., Lin, J.: Rapidly bootstrapping a question answering dataset for covid-19. *arXiv preprint arXiv:2004.11339* (2020)
58. Tátrai, D., Várallyay, Z.: Covid-19 epidemic outcome predictions based on logistic fitting and estimation of its reliability. *arXiv preprint arXiv:2003.14160* (2020)
59. Tuli, S., Tuli, S., Tuli, R., Gill, S.S.: Predicting the growth and trend of covid-19 pandemic using machine learning and cloud computing. *Internet of Things* 11, 100222 (2020)
60. Ucar, F., Korkmaz, D.: Covidiagnosis-net: Deep bayes-squeezenet based diagnosis of the coronavirus disease 2019 (covid-19) from x-ray images. *Medical Hypotheses* 140, 109761 (2020)
61. Ullah, A., Das, A., Das, A., Kabir, M.A., Shu, K.: A survey of covid-19 misinformation: Datasets, detection techniques and open issues. *arXiv preprint arXiv:2110.00737* (2021)
62. Wang, L., Lin, Z.Q., Wong, A.: Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Scientific Reports* 10(1), 1–12 (2020)

63. Wang, L.L., Lo, K., Chandrasekhar, Y., Reas, R., Yang, J., Burdick, D., Eide, D., Funk, K., Katsis, Y., Kinney, R.M., et al.: Cord-19: The covid-19 open research dataset. In: Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020 (2020)
64. Wang, S., Kang, B., Ma, J., Zeng, X., Xiao, M., Guo, J., Cai, M., Yang, J., Li, Y., Meng, X., et al.: A deep learning algorithm using ct images to screen for corona virus disease (covid-19). *European Radiology* pp. 1–9 (2021), <https://doi.org/10.1016/j.mehy.2020.109761>
65. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M.: Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2097–2106 (2017)
66. Wang, X., Song, X., Li, B., Guan, Y., Han, J.: Comprehensive named entity recognition on cord-19 with distant or weak supervision. *arXiv preprint arXiv:2003.12218* (2020)
67. Warren, M.S., Skillman, S.W.: Mobility changes in response to covid-19. *arXiv preprint arXiv:2003.14228* (2020)
68. Wynants, L., Van Calster, B., Collins, G.S., Riley, R.D., Heinze, G., Schuit, E., Bonten, M.M., Dahly, D.L., Damen, J.A., Debray, T.P., et al.: Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *British Medical Journal* 369 (2020)
69. Xia, F., Jedari, B., Yang, L.T., Ma, J., Huang, R.: A signaling game for uncertain data delivery in selfish mobile social networks. *IEEE Transactions on Computational Social Systems* 3(2), 100–112 (2016)
70. Xia, F., Sun, K., Yu, S., Aziz, A., Wan, L., Pan, S., Liu, H.: Graph learning: A survey. *IEEE Transactions on Artificial Intelligence* 2(2), 109–127 (2021)
71. Xia, F., Wang, J., Kong, X., Wang, Z., Li, J., Liu, C.: Exploring human mobility patterns in urban scenarios: A trajectory data perspective. *IEEE Communications Magazine* 56(3), 142–149 (2018)
72. Xu, B., Gutierrez, B., Mekaru, S., Sewalk, K., Goodwin, L., Loskill, A., Cohn, E.L., Hswen, Y., Hill, S.C., Cobo, M.M., et al.: Epidemiological data from the covid-19 outbreak, real-time case information. *Scientific Data* 7(1), 1–6 (2020)
73. Xu, J., Kim, S., Song, M., Jeong, M., Kim, D., Kang, J., Rousseau, J.F., Li, X., Xu, W., Torvik, V.I., et al.: Building a pubmed knowledge graph. *Scientific Data* 7(1), 1–15 (2020)
74. Yang, C., Zhou, X., Zafarani, R.: Checked: Chinese covid-19 fake news dataset. *Social Network Analysis and Mining* 11(1), 1–8 (2021)
75. Yu, J.: Open access institutional and news media tweet dataset for covid-19 social science research. *arXiv preprint arXiv:2004.01791* (2020)
76. Yu, S., Qing, Q., Zhang, C., Shehzad, A., Oatley, G., Xia, F.: Data-driven decision-making in covid-19 response: A survey. *IEEE Transactions on Computational Social Systems* 8(4), 1016–1029 (2021)
77. Zarei, K., Farahbakhsh, R., Crespi, N., Tyson, G.: A first instagram dataset on covid-19. *arXiv preprint arXiv:2004.12226* (2020)
78. Zeroual, A., Harrou, F., Dairi, A., Sun, Y.: Deep learning methods for forecasting covid-19 time-series data: A comparative study. *Chaos, Solitons & Fractals* 140, 110121 (2020)
79. Zhang, D., Zhang, M., Peng, C., Jung, J.J., Xia, F.: Metaphor research in the 21st century: A bibliographic analysis. *Computer Science and Information Systems* 18, 303–322 (2021)
80. Zhang, J., Wang, W., Xia, F., Lin, Y.R., Tong, H.: Data-driven computational social science: A survey. *Big Data Research* p. 100145 (2020)
81. Zhang, R., Hristovski, D., Schutte, D., Kastrin, A., Fiszman, M., Kilicoglu, H.: Drug repurposing for covid-19 via knowledge graph completion. *Journal of Biomedical Informatics* 115, 103696 (2021)
82. Zhao, J., Zhang, Y., He, X., Xie, P.: Covid-ct-dataset: a ct scan dataset about covid-19. *arXiv preprint arXiv:2003.13865* (2020)

83. Zong, S., Baheti, A., Xu, W., Ritter, A.: Extracting covid-19 events from twitter. arXiv preprint arXiv:2006.02567 (2020)

Ke Sun received the B.Sc. and M.Sc. degrees from Shandong Normal University, Jinan, China. He is currently Ph.D. Candidate in Software Engineering at Dalian University of Technology, Dalian, China. His research interests include deep learning, network representation learning, and knowledge graphs.

Wuyang Li is currently working toward the Bachelor's degree in the School of Software, Dalian University of Technology, China. His research interests include data science, natural language processing, and social network analysis.

Vidya Saikrishna received her PhD from Monash University, Australia in 2017. She completed her Master's in Technology (M. Tech) in 2012 and Bachelor's in Engineering in 2002 from Maulana Azad National Institute of Technology, India and Barkatullah University, India respectively. She is currently a Scholarly Teaching Fellow in Global Professional School, Federation University Australia. Her current research interests include Machine Learning, Artificial Intelligence, String Matching, and Data/Text Mining.

Mehmood Chadhar received his PhD in Information Systems from the University of New South Wales, Sydney, Australia. He is currently a Lecturer teaching business analytics, supply chain management, and real-time analytics at Federation University Australia. His areas of interest include enterprise systems implementation, organizational learning, IT business value and social media benefits.

Feng Xia received the BSc and PhD degrees from Zhejiang University, Hangzhou, China. He was Full Professor and Associate Dean (Research) in School of Software, Dalian University of Technology, China. He is Associate Professor and former Discipline Leader (IT) in Institute of Innovation, Science and Sustainability, Federation University Australia. Dr. Xia has published 2 books and over 300 scientific papers in international journals and conferences (such as IEEE TAI, TKDE, TNNLS, TBD, TCSS, TNSE, TETCI, TC, TMC, TPDS, TETC, THMS, TVT, TITS, TASE, ACM TKDD, TIST, TWEB, TOMM, WWW, AAAI, SIGIR, CIKM, JCDL, EMNLP, and INFOCOM). His research interests include data science, artificial intelligence, graph learning, anomaly detection, and systems engineering. He is a Senior Member of IEEE and ACM.

Received: August 22, 2021; Accepted: April 22, 2022.

Development of Recommendation Systems Using Game Theoretic Techniques

Evangelos Sofikitis¹ and Christos Makris²

¹ Department of Computer Engineering and Informatics, University of Patras,
Rio 26500, Patras, Greece
sofikitis@ceid.upatras.gr

² Department of Computer Engineering and Informatics, University of Patras,
Rio 26500, Patras, Greece
makri@ceid.upatras.gr

Abstract. In the present work, we inquire the use of game theoretic techniques for the development of recommender systems. Initially, the interaction of the two aspects of the systems, query reformulation and relevance estimation, is modelled as a cooperative game where the two players have a common utility, to supply optimal recommendations, which they try to maximize. Based on this modelling, three basic recommendation methods are developed, namely collaborative filtering, content based filtering and demographic filtering. The different methods are then combined to create hybrid systems. In the weighted combination, the use of game theoretic techniques is extended, as it is modelled as a *cooperative game*. Finally, the methods are combined with the use of a genetic algorithm where game theory is used for the parent selection process. Our work offers a baseline for the efficient combination of recommendation methods through game theory and in addition the novelty method, *Choice by Game*, for the parent selection process in genetic algorithms which offers consistent performance improvements.

Keywords: recommendation system, game theory, genetic algorithm.

1. Introduction

Information retrieval systems, consist of two aspects, namely the formulation of an optimal query to best represent the target user's need of information and the estimation of the documents' relevance to this query [29,16]. According to Rocchio's fundamental theory [26], the optimal query reformulation, in text retrieval, is achieved through relevance feedback [11]. In more detail, the query is reformulated through an iterative process, where the system returns results and based on the user's feedback on the results, expands the terms of the query. Depending on the individual situations, the feedback signal might be implicit, such as clicks or play-lists, or blinded, assuming the top-k returned results as relevant ones [31]. The same ground rules could be applied to recommendation systems after making some modifications. In recommender systems where there is no initial query provided, the user's need is represented by a profile based on its historical interactions and may be inferred from other similar user profiles [30]. As far as the second aspect is concerned, the main goal is to assign relevance scores to the available documents given the query. The classic retrieval model [24] and its variations, BM25 [25], and language models [34], utilize term weighing to calculate the relevance scores.

In related work [36], based on game-theoretical analysis, a new equilibrium theory of information retrieval is proposed, according to which, the two basic aspects of Information Retrieval are correlated and participate in a cooperative game. More specifically, the query reformulation player would refine the query that is the best response to the results returned from the given retrieval model player, i.e., formulate an optimal query that would maximize its utility. Simultaneously, the retrieval model player would also need to produce the document relevant estimation that is the best response toward the formulated query. The game play provides the retrieval solutions that is in a Nash equilibrium or multiple Nash equilibria [19] when each becomes the best response to the other. In the recommendation systems setting, this novelty equilibrium theory is evaluated through a practical implementation in collaborative filtering tasks and the experiments show that, it outperforms the query reformulation and retrieval model reformulation, when applied separately.

In this paper, in order to confirm that, the modelling of the recommendation process as a cooperative game offers performance improvements, we apply this equilibrium theory [36] in more recommendation methods, namely collaborative filtering, demographic filtering, content based filtering and hybrid combinations of the three [7], while scaling the experiments in a larger dataset. The experiments show that the initial claim [36] is confirmed for the variations of the recommendation methods, however the same does not apply for the scaled experiments. Thereinafter, we extend the integration of game theory in recommendation systems by developing novelty algorithms. An algorithm for the weighted combination of two recommendation methods is developed, where the weights of every method derived from a cooperative game [6]. The methods are the players of the game and they are, initially, assigned with a weight. During the game, which is an iterative process, every player chooses a strategy, i.e. increase, or decrease its weight to maximize the common utility. An equilibrium is reached when the two players can not increase their utility by altering their weights. This algorithm is evaluated with experiments in two datasets where, although different combinations of methods are attempted, the weighed combination through game excels in performance compared to the individual methods combined. Finally, game theoretic techniques are also applied in genetic algorithms, where the parent selection process is modelled as a game. Two user-based recommendation methods are executed, and the resulting top-k users are combined to create an initial population for the genetic algorithm. In the method *Choice by Game*, the probability of two individuals being selected as parents, is calculated through a game and is proportional to the utility of the combination. The results of the experiments in two datasets show that, the *Choice by Game* method offers consistent performance improvements when compared to the existing *Choice by Roulette* method and to the individual methods from whom derived the initial population for the genetic algorithm.

Although a simple linear retrieval model is used for the basic recommendation methods, our work offers a baseline for the efficient development of hybrid recommendation systems through game theory, which could be applied to more refined systems. Utilizing our algorithms, two recommendation methods can be combined linearly or through a genetic algorithm resulting to a hybrid that provides more qualitative recommendations than the individual methods. Furthermore, the *Choice by Game* method can be used not only in recommendation systems, but in the various applications [27] of the genetic algorithms. Briefly, the paper is organized as follows. In the following section (Section 2), we con-

duct a brief review of and how has game theory utilized to produce more efficient and improved recommender systems. Then, in Section 3, we present the baseline of previous research on the modelling of the recommendation process as a cooperative game. In Section 4, we elaborate on the basic recommendation methods that were developed using this modelling. The integration of game theory and recommendation systems is extended on Section 5, where we introduce two algorithms for the efficient combination of two recommendation methods, namely linearly and through a genetic algorithm. Both algorithms utilize game theoretic techniques. In Section 6 the practical implementation used for the experimental evaluation of all the above algorithms is explained, while in Section 7 the results of the experiments in two datasets are presented. In the epilogue of this paper (Section 8), the conclusions of our inquiry is discussed along with suggestions for future research.

2. Related Work

2.1. Recommender Systems

On the Internet, there is an overabundance of options, choices and products, rendering it impossible for users to browse and evaluate all of them, in order to choose the most desirable ones. In our daily lives we rely on the recommendations of others, such as friends or critics, for the selection of movies, restaurants, products, etc. On the Internet, this role is carried out by the recommendation systems [5]. These systems are a subset of information filtering systems and aim to present the most relevant products from a set, making predictions about users' preferences. In the initial stages of recommender systems, the collaborative filtering method was developed, where recommendations are based on the preferences of users who are relevant to the target user. Other methods such as demographic filtering and content based filtering were later developed [23]. The demographic filtering, like the collaborative filtering is based on users' similarity while content based filtering is based on item similarity, meaning that the systems suggest to the user items similar to those they have already evaluated and rated highly [17]. The ever-increasing need for efficient recommendations has led to more sophisticated systems that often utilize two or more methods, these are hybrid recommender systems [3], or draw on methods and techniques from other scientific fields [35]. The evolution of recommendation systems has led to extending their usage in other fields other than improving users' online experience, even in the education [20]. State of the art recommendation systems utilize neural networks and matrix factorization to develop better performing systems [22]. Hierarchical Recurrent Neural Networks are currently one of the most efficient methods to achieve felicitous recommendations [21]. In our research, however, we do not utilize such sophisticated methods but rather, try to find ways to efficiently combine recommendation methods and techniques.

2.2. Recommender Systems and Game Theory

Game Theory is the scientific field that studies games, which are interdependence situations of players [18]. Techniques from this field are widely applied in various scientific fields with machine learning being one of the latest. The combination of recommender

systems and game theory is an innovative field of research with a relatively limited literature.

However, researchers have already utilized game theoretic techniques to various aspects of recommender systems. Such techniques have been used to balance accuracy and coverage of recommendations through rough sets [2], to counterpoise profit of strategic content providers to application usability [4] and to find an equilibrium between qualitative recommendations and data privacy [10]. Moreover, game theory has been used to locate trustworthy users in a set more efficiently and consequently provide more accurate recommendations [1] and even to provide a novel formulation of the recommendation process i.e., as a cooperative game between the user and the systems enhancing the process as a whole [33]. In our work, we follow a more straight-forward approach, as we attempt to provide more efficient and qualitative recommendations through game theoretic techniques.

3. Baseline of Previous Approach

As mentioned in the previous section, any retrieval process consists of two main aspects. The first aspect is to formulate the query to best represents the user's need for information. The second aspect is to calculate the relevance of the available documents or items to the query and select the most relevant ones. In recommendation systems instead of a query, a profile is used to represent the preferences of the target user. Since there is no initial query, the profile is based entirely on the user's historical interactions with the products, for example their ratings and comments. The goal of the retrieval models remains the same, that is, to estimate the relevance of available items to the profile and select the items that are more likely to satisfy the target user. When it comes to text retrieval, the classic models assign weights to query terms so that each word has a different weight when calculating relevance. Term weighting can also be adapted, in recommendation systems, utilizing past interactions as terms. Although recommendation systems are a subset of information filtering systems, they can be formulated using information retrieval techniques. In this case in particular, instead of filtering the non-relevant users, in every iteration of the algorithm the users are classified as relevant and non-relevant and both sets contribute to the query and retrieval parameters reformulation. Namely, both offer information to the system in order to conclude to the recommendations.

3.1. Modelling

According to related research [36], for the modelling of the recommendation process as a game, the profile of the target user q and a set of objects D are defined. Each object d_i of the set D can be a product, a service, or another user, since the recommendations for the target user can be derived from the relevant users. The profile and each object of the set are represented by a vector of attributes. The attributes vary depending on the recommendation method used. The recommendation process can be modelled as a cooperative game with three key elements: players, their strategies, and profit functions. The game is collaborative, as the common goal of the two players and the means of maximizing their utilities is to provide qualitative recommendations.

Definition 1. (Game of Recommendation Process)

The game is defined as follows:

- *Players:* the query reformulation Q and the retrieval model reformulation M ,
- *Strategies:* S_Q and S_M are two finite sets of strategies, which are available to players Q and M , respectively. A strategy $s_Q \in S_Q$ can for example be the weight gain of an item in the user profile which is more relevant than the rest of the items to better represent their user preferences,
- *Utility Functions:* u_Q and u_M define the utilities of players Q and M respectively and depend on the players' strategies. Equilibrium occurs when both players have no motivation to change strategy. That is, when a unilateral change of strategy will reduce the profit of the player who changed.

The utility of the retrieval model M depends on the successful distinguishing the relevant objects from the non-relevant ones.

$$u_M(s_Q, s_M) = \frac{1}{|D_r|} \sum_{d_i \in D_r} \log p(r = 1|d_i, q; w) - \frac{1}{|D_n|} \sum_{d_i \in D_n} \log p(r = 0|d_i, q; w) \tag{1}$$

where:

- D_r, D_n are the sets of relevant and non-relevant items respectively,
- w is the weight vector of the retrieval model for each attribute,
- p is the probability that the object is relative ($r = 1$) or non-relevant ($r = 0$) given the profile q , the object d_i and the weights w

The probability is calculated using the sigmoid function

$$s(x) = \frac{1}{1 + e^{-x}} \tag{2}$$

where:

$$\sum_{j=1}^N w_j q_j d_{ij} \tag{3}$$

j being an element of the characteristic vectors of length N , x is the inner product of the vectors w, q, d_i . The gain of M increases if an object belongs to the relevant, i.e. $d_i \in D_r$ and the model has a high probability of distinguishing it as so and vice versa, decreases if the probability is low. The same applies to non-relevant items. Therefore, in the iterative process of reshaping the weights w , the strategy that appropriately shapes the probabilities and maximizes its profit is chosen. On the other hand, the utility of player Q is based on the feedback from the retrieval model

$$u_Q(s_Q, s_M) = \frac{1}{|D_k|} \sum_{d_i \in D_k} \log p(r = 1|d_i, q; w) - \frac{1}{N - |D_k|} \sum_{d_i \notin D_k} \log p(r = 0|d_i, q; w) \tag{4}$$

where:

- D_k are the top-k objects as classified, based on relativity, by the recovery model
- N is the set of all available objects

The method of calculating the probability remains the same. The player Q also tries to configure the probabilities appropriately in order to maximize its utility with the difference that its available strategies aim at reformulating the profile q instead of w . When there is no apriori knowledge of the relevant and non-relevant objects, we assume the top-k as relevant and the others as non-relevant. In that case, the players Q and M share a common utility function.

3.2. Cases: ConvQ - ConvM – Game

For the application of the above modelling of the recommendation process as a game, three cases are distinguished [36], the reformulation of the user's profile, the reformulation of the weights of the retrieval model and the simultaneous reformulation of both. Let q be the target user's profile, d_i an object of the set d , w the weights of the retrieval model and θ_i the relevance score of the object d_i to the profile q . In all three cases, players try to maximize their utility using gradient ascent. Gradient ascent is an iterative algorithm for finding the maximum of a differential function. In the case of *Retrieval Model Reformulation*, θ_i is calculated as follows:

$$\theta_i = \text{sigmoid}(q^\top w d_i) \quad (5)$$

while in the case of *Profile Reformulation*, the calculation method of θ_i is defined as:

$$\theta_i = \text{sigmoid}(q^\top d_i) \quad (6)$$

where q^\top is the inverse vector of q , $q^\top w d_i$ is the inner product of q , w , d_i and respectively, $q^\top d_i$ is the inner product of q and d_i .

Case 1: Profile Reformulation In this case, the player Q tries to maximize its utility u_Q by reformulating the profile q while the weights w remain constant.

$$\frac{\partial u_Q(s_Q, s_M)}{\partial q} = \frac{1}{|D_k|} \sum_{d_i \in D_k} (1 - \theta_i) d_i - \frac{1}{N - |D_k|} \sum_{d_i \notin D_k} \theta_i d_i \quad (7)$$

$$q \leftarrow q + \eta \frac{\partial u_Q(s_Q, s_M)}{\partial q} \quad (8)$$

q is updated with gradient ascent as shown in the above equations Eq. (8) and Eq. (10), where η is the learning rate. We call this case *ConvQ*.

Case 2: Retrieval Model reformulation Similarly, in this case player M tries to maximize its utility u_M by reformulating the weights w while the profile q remains constant. w is updated using the gradient ascent as shown in the following equations, where η is the learning rate. We call this case ConvM.

$$\frac{\partial u_M(s_Q, s_M)}{\partial w} = \frac{1}{|D_r|} \sum_{d_i \in D_r} (1 - \theta_i) q d_i^\top - \frac{1}{|D_n|} \sum_{d_i \in D_n} \theta_i q d_i^\top \tag{9}$$

$$w \leftarrow w + \eta \frac{\partial u_M(s_Q, s_M)}{\partial w} \tag{10}$$

Case 3: Game In the third case, which we call Game, both q and w change. Player Q changes their strategy and reshapes the profile q in response to the strategy chosen by M, who does the same by updating w based on the reformulation of q . Their interaction is a collaborative game.

4. Equilibrium of the Recommendation Process

The modelling of the recommendation process as a cooperative game is product of related work [36], in which, it was applied and evaluated on the collaborative filtering method. We expand the usage of this methodology to other two fundamental recommendation methods and four hybrid combinations of them.

4.1. Collaborative Filtering

In the Collaborative Filtering algorithm, the set of objects D consists of user profiles and each element of the vectors d_i and q represents a product from the set of available products to be recommended. The value of every elements is the rating of each user for the corresponding product. The recommendation system identifies the users $d_i \in d$ who are most relevant to user q , i.e. users who have similar ratings to the target user. The collaborative filtering method is based on the assumption that two users who have rated a set of products similarly, will have the same satisfaction or dissatisfaction from products that have not yet rated [8]. Therefore, the user for whom the recommendations are intended is likely to be satisfied with products that the relevant users have rated highly. The aim of the system is to predict the user’s ratings for the products that he has not evaluated and to present to him the products for which he has provided a high rating. To do this, the system first identifies the k most relevant users. K is a number smaller than the total number of users and varies depending on the application. The predicted rating for each product is the average ratings of the top-k users for that product. The products with the highest predicted rating are then recommended to the target user.

4.2. Demographic Filtering

The demographic filtering method is based on the assumption that users with similar demographics will have the same preferences on products. In this algorithm, the profile q , and the objects d_i of set D are still user profiles, for the representation of which vectors are used. The elements of each vector represent a demographic characteristic, such as age, gender, occupation, place of residence and more. The same steps as described above for the collaborative filtering algorithm are followed, to generate recommendations.

4.3. Content Based Filtering

For the content based filtering algorithm the d_i objects are products and the elements of the vectors representing each $d_i \in D$ are characteristic of the products. The user q profile is also a feature vector where each attribute has a value according to the preferences of the user. The profile is created based on previous ratings of the user. More specifically, we consider a set of N products $P \subset D$, each product $p_i \in P$ for $i \in [1, N]$ has been rated by the user with r_i . Each of these products is represented by a vector of attributes c and $p_{i,j}$ is the value of the product for the attribute $c_j \in c$. The element q_j in the profile q represents the preferred value of the attribute c_j by the target user, and it is calculated by the formula:

$$q_j = \frac{\sum_{i=1}^N p_{i,j} * r_i}{\sum_{i=1}^N r_i} \quad (11)$$

So instead of the bipartite user to item graph an item to item graph is utilized to result in the recommendations for the target user.

4.4. Hybrid Methods

Hybrid1 and Hybrid2 algorithms are hybrid recommendation algorithms that use the above methods. They are serial filtering algorithms, in the sense that the algorithms are executed sequentially, and the results of the first are used to reinforce the recommendations of the second. The cases developed and presented below are combinations of collaborative filtering and demographic filtering or content based filtering.

Hybrid1.1: Demographic Filtering after Collaborative Filtering The Hybrid1.1 algorithm is the combination of collaborative filtering and demographic filtering. It is a simple algorithm that identifies users who have similar demographic characteristics to the target user, among users who have already been judged to be relevant based on their ratings. Running collaborative filtering results in a set of users who are relevant to the target user. This set is then used in demographic filtering to generate the k final users from whom the recommendations derived.

Hybrid1.2: Collaborative Filtering after Demographic Filtering Hybrid1.2 combines the two algorithms like Hybrid1.1 but in a different order. More specifically, running demographic filtering creates a set of relevant users from which, after collaborative filtering, the top- k most relevant are chosen. Essentially, the Hybrid1 algorithms apply double filtering on the set of users D . Once with demographic criteria and once with their product ratings.

Hybrid2.1: Content Based Filtering after Collaborative Filtering In the combination of content based filtering and collaborative filtering where collaborative filtering is performed first, the procedure is as follows, the system performs collaborative filtering from which the top-k users derived. Then, a set of products is created, containing the products that are highly rated by the top-k users. This set is used for the content based filtering algorithm. In other words, the system selects the products that the relevant users liked and filters them for the final recommendations.

Hybrid2.2: Collaborative Filtering after Content Based Filtering In the case of Hybrid2.2, first content based filtering is performed resulting in the top-k products, which are more relevant to the target user. These products are given more weight to the calculation of the most relevant users when performing collaborative filtering. More specifically, in q and d_i , which are vectors and each of their elements represents a product, a second elements is created for the representation of each of the top-k products. Thus, these products have a double weight and affect the relevance score of users that collaborative filtering will identify.

5. Game Theory in Hybrid Systems

In this section, we present two methods of integrating game theory to recommendation systems. Initially, the linear combination of two recommendation methods is modelled as a game. In the second hybrid system, two methods are combined with a genetic algorithm and the game takes place in the selection stage where the novelty method *Choice by Game* is used.

5.1. Equilibrium in Linear Combination

The Hybrid3 algorithm is a hybrid algorithm of linear combination of two recommendation methods. The algorithms of each combination run simultaneously and independently and each produces a score vector according to its predictions. Each element of the vectors corresponds to a product and the value of the element is the rating that the algorithm predicted for it. These vectors are combined linearly to obtain the final ratings predictions. If R_A , R_B are the predicted ratings and α , β are the weights of the algorithms A, B respectively, then the final prediction of the ratings of a product, R results from the combination as follows:

$$R = R_A * \alpha + R_B * \beta \quad (12)$$

Initially the weights have values $\alpha = 1$ and $\beta = 0$ which change while their sum remains constant ($\alpha + \beta = 1$). The final weights occur after a cooperative game.

Definition 2. (*Game of Recommendation Methods Combination*)

The game is defined as follows:

- *Players:* A, B are the algorithms that are combined to produce recommendations and have initial weights α , β respectively
- *Strategies:* s_A and s_B for players A, B respectively are to reduce or increase their weights

- *Utility function: u is common for both players and they try to maximize it. A state of equilibrium occurs when neither of the players benefits from changing their strategies or when one of the two weights reaches a certain upper limit.*

More specifically, let r_A, r_B be the vectors of estimated scores of the algorithms A, B respectively and q the target user's profile. The scores r_A, r_B are combined according to equation Eq. (12) resulting to the vector r . The players' common utility is defined as the correlation of the vectors r and q . The method for calculating correlation may differ in applications. In each round of the game the weights are renewed, α decreases and β increases by a constant c . Then the score vector r is recalculated with the updated weights. At the end of each iteration the new utility u is calculated. The game ends when the utility decreases or when $\beta = 1$ and $\alpha = 0$.

5.2. Equilibrium in Genetic Algorithm

Collaborative filtering and demographic filtering identify the top-k most relevant users to the target user. These users can be represented as vectors every element of which corresponds to a product and the value of the element is the user rating for this product. These vectors can be used as the initial population for a genetic algorithm. Each user is an individual of the initial population and each element of the vector that describes them is a chromosome. The genetic algorithm is performed iteratively and in each iteration, the fittest individuals, i.e. those who are more relevant to the target user in terms of their ratings, pass on to the next generation or reproduce and their offspring pass on to the next generation. The fitness function may differ depending on the characteristics of each application. The stage of selection is modelled as a cooperative game.

Definition 3. (Game of Individual Selection)

The game is defined as follows:

- *Players: A, B are the parents, who will participate in the reproduction*
- *Strategies: s_A and s_B are the individuals whose chromosomes will be combined to produce an offspring*
- *Utility function: u is common to both players and they try to maximize it. Each pair of strategies has a chance to be chosen, which is proportional to the utility it offers.*

More specifically, let D be the population of the genetic algorithm and d_i, d_j two individuals of the population. If the fitness of the individuals is based on their correlation to the target user's profile q , calculated by a function f , namely $fitness(d_i) = f(d_i, q)$ and $fitness(d_j) = f(d_j, q)$ then the shared utility of the players for choosing this pair of individuals is calculated as follows:

$$u(d_i, d_j) = (fitness(d_i) + fitness(d_j) * f(d_i, d_j)) \quad (13)$$

In each generation the utility for every combination of individuals is calculated and the probability of a combination being selected for reproduction is the utility of this combination to the sum of the utilities of all combinations. The goal of the algorithm is to create a population with a higher fitness than the initial, so the algorithm terminates when the sum of the fitness of the new generation is lower than the previous one or when a

predetermined number of generations is exceeded. The recommendations derived from the final population, as described above for the collaborative filtering algorithm.

The following simple example is presented to illustrate the algorithm. Let a target user u_T and three users u_1, u_2 and u_3 who represent the choices of the two players A and B i.e. parents in the crossover stage of the genetic algorithm. The parents aim to select the most dominant users, those who, when reproduced will result in an offspring that is closest to the target user. Every user is represented by six chromosomes c_i for $i = 1, \dots, 6$ which are their rating to a corresponding movie. In this example we will use euclidean distance to calculate the fitness of every user, so this is now a simple minimization problem.

Table 1. Choice by Game Example.

	c_1	c_2	c_3	c_4	c_5	c_6
u_T	0	3	2	4	2	1
u_1	1	3	2	2	4	2
u_2	2	2	3	4	3	1
u_3	1	5	1	5	2	1

Chromosomes of Individuals

		Player B		
		u_1	u_2	u_3
Player A	u_1	-	17.42	25.31
	u_2	17.42	-	21.16
	u_3	25.31	21.16	-

Costs of Strategies

The chromosomes of every individual are presented in the table on the left. Respectively, the cost of the players' choices, calculated using equation 13, are presented in the table on the right. The players aim to minimize their cost i.e. the distance to the target user. The lower the cost the higher the probability of a user combination to be chosen for the crossover stage. Of course the element of randomness exist in every stage of a genetic algorithm, however for the sake of this example we suppose that the most dominant pairs are chosen for the crossover stage. In this example the most dominant pair of users are (u_1, u_2) and (u_2, u_3) with approximate costs of 17.42 and 21.16 respectively, so these two pair of choices will have a higher probability to be chosen and their offspring will proceed to the next generation. The crossover and mutation methods as well as the probabilities thereof depend on every problem's characteristics.

6. Implementation

The evaluation of the algorithms was based on two sets of experiments, for which the MovieLens 100k and MovieLens 25M datasets were used [12].

6.1. MovieLens 100k

The MovieLens 100k dataset contains:

- 100,000 ratings from 943 users for 1,682 movies,
- demographic data for each user which includes age, gender, occupation, and zip code

Pre-processing The collaborative filtering and demographic algorithms were implemented and evaluated for this dataset using user ratings and demographic data, respectively. Pre-processing is performed so that the data takes the appropriate form to run the algorithms.

For the collaborative filtering, user profiles are created, which are represented by vectors of ratings. Users have not rated all the movies in the set, so the missing ratings are initialized with 0. Next step in the preprocessing is the test-train Split, which is done in a ratio of 1 to 3, i.e. 75% is the set of users from which the most relevant users derived and the remaining 25% of users are used to evaluate the algorithms. From each user of the test set, we keep the 75% of the ratings as its history and the remaining 25% for evaluating the recommendations of each algorithm. For the demographic filtering, a different procedure is followed. The data is already in the form of user profiles, so the first step is omitted, however another pre-processing is required. Some of the demographics are categorical variables, so they are converted to numeric. In addition, unlike ratings, each feature in demographic profiles has a different scale, so we normalize the values of each of the four features on a scale of 0 to 1. Then apply train-test split with a ratio of 1 to 3 as described above.

Experiments For every one of the collaborative filtering, demographic filtering, Hybrid1.1 and Hybrid1.2 algorithms we distinguish the three cases ConvQ, ConvM and Game, which have been described above. For each user of the test set we execute the algorithms and their cases to compare the results. For the Hybrid3 and Hybrid4 algorithms the cases differ. More specifically, for the linear combination algorithm, Hybrid3, we compare the case in which the weights of the combined algorithms result from a cooperative game with the case where the weights, α and β , are constant. Respectively for the Hybrid4 genetic algorithm we consider two cases, in which the method of selecting parents for reproduction differs. We compare the selection method Choice by Game with the existing method Choice by Roulette.

Collaborative Filtering: after selecting the target user, the three cases ConvQ, ConvM and Game are executed sequentially. For the ConvQ case, we first calculate the relevance of the users in the train set to the target user. We consider top-k users as relevant and the rest as non-relevant. The parameter k has been given the value 100, since after experiments it showed the best results. The user profile is then reformulated according to equation Eq. (8). The learning rate η , is set to 0.1. This process is repeated until the profit converges to a maximum, using the threshold 10^{-4} . In most cases, the utility function converges before 50 iterations, so the profile is updated 50 times. The learning rate, the ratio of convergence and the iterations are based on precedents experiments [36]. In the case of ConvM the same procedure is followed except that the retrieval model is updated according to equation by Eq. (10). Finally, for the Game, the profile and the retrieval model are reformulated simultaneously in order to maximize their common utility. In each iteration, player Q chooses the best strategy in response to the strategy of M, who does the same. When neither player has an incentive to change their strategy, i.e. to reformulate the profile or retrieval model respectively, equilibrium occurs, and the profit function converges to a maximum.

Demographic Filtering: The same procedure as collaborative filtering is followed, except that instead of rating profiles, the top-100 users are calculated based on demographic profiles. The iterations for the convergence of the utility function with the gradient ascent method are reduced to 25 from 50. The demographic profiles consist of 4 characteristics versus the 1,682 characteristics of the rating profiles, therefore faster convergence

can be achieved. At the end of each case the demographic profiles are replaced with the corresponding rating profiles to calculate the predicted ratings.

Hybrid1.1 & Hybrid1.2: The Hybrid1.1 algorithm combines collaborative filtering and demographic filtering methods to perform double filtering. More specifically, collaborative filtering is initially executed to identify the top-200 most relevant users, who are then filtered again using the demographic filtering method resulting to the top-100 final users. The Hybrid1.2 algorithm is similar to Hybrid1.1 except that the demographic filtering precedes the collaborative filtering.

Hybrid3: The Hybrid3 algorithm is a linear combination of collaborative filtering and demographic filtering. The individual algorithms are executed, and the predicted ratings are calculated. Then, for each movie, the weighted average of the two ratings is calculated using as weights the α , β , which have the initial values 1 and 0, respectively. The utility function is calculated with the inverse hyperbolic sine of the inner product of the predicted ratings and the target user's profile q

$$u = \sinh^{-1}(r^\top q) \quad (14)$$

The next step is to renew the weights, subtracting 0.01 from α and adding it to β , i.e. $\alpha = 0.99$ and $\beta = 0.01$. This process is repeated until the utility is reduced or β has a value of 1 and α 0. To evaluate the above algorithm, we also implement a simple linear combination, where the weights have a constant value of 0.5, i.e. the ratings are derived from the average of the individual scores.

Hybrid4: The Hybrid4 algorithm combines the results of collaborative filtering and demographic filtering. Once the two algorithms are executed, their top-100 users are used as the initial population of the genetic algorithm. For each individual, there is a possibility of mutation in which a chromosome randomly takes on a value. The probability of reproduction and mutation is 0.7 and 0.05 respectively. During reproduction, the two-point crossover method is used. The algorithm terminates when the generation limit is exceeded, which is set to 100, or when the sum of the fitness of the individuals of the current generation is lower than that of the previous one. The fitness of an individual d_i is the inverse hyperbolic sine of the inner product of d_i and the target user's profile q multiplied by the correlation of q and d_i calculated by the Spearman's ρ correlation

$$fitness(d_i) = \sinh^{-1}(q^\top d_i) * \rho(d_i, q) \quad (15)$$

To evaluate the efficiency of the Choice by Game method, the same algorithm is implemented with the Choice by Roulette method and their results are compared.

6.2. MovieLens 25M

The implementation of the algorithms for the MovieLens 25M data set is similar to that for the MovieLens 100k dataset. The main difference is due to the size of the second dataset, which makes it more difficult to handle, while the methods used for the first dataset are prohibitive, due to time complexity. In addition, the MovieLens 25M dataset does not contain demographic characteristics but movie-tag correlations, which describe

the content of the movies. Therefore, content based filtering is implemented instead of demographic filtering. The MovieLens 25M set contains:

- 25,000,095 ratings from 162,541 users for 62,423 movies
- 1,093,360 movie correlations with 1,128 tags

Pre-processing As mentioned above, this data set is significantly larger than the previous one, so the same preprocessing methods cannot be used. To address this, instead of representing profiles as vectors, an inverted index is used. The same procedure is followed for the train-test split as well as for the user ratings in the test set. The preprocessing required for content based filtering is to create profiles for the movies based on the movie-tag correlations. In addition, a corresponding profile should be created that presents the preferences of the target user.

Experiments For the collaborative filtering, content based filtering, Hybrid2.1 and Hybrid2.2 algorithms the three cases ConvQ, ConvM and Game are examined while for the Hybrid3 and Hybrid4 the same cases described for the MovieLens 100k are used.

Collaborative Filtering: The collaborative filtering algorithm is the same as in the MovieLens 100k dataset. For every target user, the set of available users from which the top-100 are retrieved, is limited to the 1,000 most relevant users, instead of using the whole train set, for time efficiency.

Content Based Filtering: this algorithm follows the same procedure as collaborative filtering and top-750 most relevant movies are retrieved. The predicted ratings derived from their degree of relevance to the target user.

Hybrid2.1: Firstly, the collaborative filtering algorithm is executed, and the top-100 most relevant users are retrieved. From each of these users the 10 movies with the highest rating are selected to create a set of 1,000 movies. Content based filtering is applied to this set to retrieve the top-750 movies for the user.

Hybrid2.2: The top-750 most relevant movies for the target user are retrieved from the content based filtering. Then, the user's profiles are modified adding a duplicate element for each of the top-750 movies. Doing so, these movies affect the top-100 user retrieved from the collaborative filtering.

Hybrid3: For the MovieLens 25M, collaborative filtering and content based filtering are combined linearly. The utility function based on which the players of the game renew their weights is the inverse hyperbolic sine of the inner product of the predicted ratings r and the target user's profile q , multiplied by the Spearman's ρ of r and q .

Hybrid4: For the Hybrid4 algorithm, in this dataset we use as initial population the top-100 users from the collaborative filtering and Hybrid2.2 algorithms. The same comparison is made between the parent selection methods, i.e. selection by game and selection by roulette. As a fitness function, which determines the individuals of each generation, the inverse hyperbolic sine of the inner product of the vectors describing the target user and the individual of the population, is used. The probability of reproduction and mutation and the maximum number of generations remain the same.

7. Results

After the algorithms are executed, we evaluate the results [13]. The metrics used for the evaluation are Normalized Discounted Cumulative Gain and Precision for the first 10 and the first 30 recommendations, Mean Average Precision and Mean Reciprocal Rank i.e. NDCG@10, NDCG@30, P@10, P@30, MAP and MRR [28]. For the dataset MovieLens 100k experiments are run for the entire test set, while in the MovieLens 25M dataset the experiments for the "Game of Recommendation Process" algorithms were very time consuming so experiments could not be performed for the entire test set. However, the results are reliable, because we focus on the comparison of the algorithms and not their actual scores on the metrics. The comparisons are confirmed with t-test for each algorithm [15]. For the collaborative filtering, demographic filtering, content based filtering, Hybrid1 and Hybrid2 algorithms we compare the cases ConvQ, ConvM and Game. For Hybrid3 we compare the cases Game and Simple, i.e. the determination of weights for the linear combination with game and the simple linear combination respectively as well as the results of the Game with the individual algorithms that are combined. Respectively, for the Hybrid4 we compare selecting parents for reproduction, we also compare the Game with the individual algorithms as for Hybrid3. The best value of each algorithm for the respective metric of evaluation is written in bold.

It should be noted that, the Content Based Filtering algorithm tested on the MovieLens 25M dataset returns 750 movies out of 62,423 movies as recommendations for the user. Having predicted ratings for only 750 movies, it is difficult to evaluate the algorithm as it is very likely that the target user has rated only a few of them or even none. This is a common problem in recommendation systems address by many elaborate solving methods [14,32]. However, we resolve in a simplified method, during the evaluation we delete the movies that the user has not rated to easier evaluate the algorithms. This affects the evaluation metrics greatly, however the absolute values of the metrics used in the evaluation are not relevant. Rather we focus on the comparison of the algorithms to determine which is more efficient and since this method is used in every algorithm their relative efficiency is not altered.

7.1. MovieLens 100k

Table 2. Collaborative filtering results for the MovieLens 100k dataset.

	NDCG@10	NDCG@30	P@10	P@30	MAP	MRR
ConvQ	0.80551	0.86006	0.54746	0.44802	0.43742	0.25880
ConvM	0.80566	0.86012	0.54788	0.44689	0.43296	0.26080
Game	0.80815	0.86205	0.55000	0.44845	0.43737	0.26085

The Game case of collaborative filtering shows better results than ConvQ and ConvM in five out of six metrics, while the ConvQ case shows better performance only based on the metric MAP. Overall, the Game case is the best in the collaborative filtering algorithm. These results are in accordance to previous experiments [36].

Table 3. Demographic filtering results for the MovieLens 100k dataset.

	NDCG@10	NDCG@30	P@10	P@10	MAP	MRR
ConvQ	0.77903	0.83975	0.53771	0.44562	0.41965	0.26056
ConvM	0.76436	0.83055	0.53432	0.44251	0.41149	0.26079
Game	0.77918	0.84131	0.53602	0.44477	0.42429	0.26353

In the demographic filtering algorithm, the Game case performs better on all metrics except Precision, for the first 10 and 30 results. So, overall, the Game case outperforms the other cases.

Table 4. Hybrid1.1 filtering results for the MovieLens 100k dataset.

	NDCG@10	NDCG@30	P@10	P@10	MAP	MRR
ConvQ	0.77903	0.83975	0.53771	0.44562	0.41965	0.26056
ConvM	0.77795	0.84034	0.53898	0.44605	0.42010	0.26070
Game	0.77880	0.84036	0.53898	0.44675	0.41920	0.25979

In the Hybrid1.1 algorithm, the Game case performs better for the NDCG@30, P@10 and P@30 metrics, the ConvQ case for the NDCG@10 and MAP metrics, while the ConvM case outperforms only the metric MRR. So, we can assume that the Game case is the best of the three.

Table 5. Hybrid1.2 filtering results for the MovieLens 100k dataset.

	NDCG@10	NDCG@30	P@10	P@10	MAP	MRR
ConvQ	0.78641	0.84782	0.54110	0.44718	0.42750	0.26964
ConvM	0.78395	0.84637	0.53898	0.44647	0.42691	0.26478
Game	0.78676	0.84794	0.54068	0.44703	0.42839	0.26854

In Hybrid1.2 algorithm, Game shows better results on the metric NDCG@10, NDCG@30 and MAP, while ConvQ on P@10, P@30 and MRR. So, these are the two best cases, but we cannot say that one outperforms the other.

Below, in Table 6, we compare the results of the Game case of the Hybrid3 algorithm, with the results of the individual algorithms that are combined, i.e. the Game cases of Collaborative Filtering (CF) and Demographic Filtering (DF). We also compare the two cases, Game and Simple, of Hybrid3. The aim is to determine whether the linear combination yields better results than the two algorithms and whether the combination using a game is more efficient than a simple combination. The results show that the Collaborative Filtering method performs better according to the metrics NDCG@10 and NDCG@30, the Demographic Filtering in the metric MRR, while the linear combination of the two shows better results in the metrics P@10, P@30 and MAP, therefore we can conclude that the Hybrid3 algorithm excels the two individual algorithms combined. In addition,

Table 6. Hybrid3 results for the MovieLens 100k dataset.

	NDCG@10	NDCG@30	P@10	P@30	MAP	MRR
CF	0.80815	0.86205	0.55000	0.44845	0.43737	0.26085
DF	0.77918	0.84131	0.53602	0.44477	0.42429	0.26353
H3 (Game)	0.80795	0.86159	0.55042	0.44859	0.43741	0.26183

	NDCG@10	NDCG@30	P@10	P@30	MAP	MRR
H3 (Simple)	0.79729	0.85371	0.54576	0.44788	0.43226	0.27150
H3 (Game)	0.80795	0.86159	0.55042	0.44859	0.43741	0.26183

the linear combination using game is more efficient than the simple linear combination, as it shows better results in all metrics except MRR.

Table 7. Hybrid4 results for the MovieLens 100k dataset.

	NDCG@10	NDCG@30	P@10	P@30	MAP	MRR
CF	0.80815	0.86205	0.55000	0.44845	0.43737	0.26085
DF	0.77918	0.84131	0.53602	0.44477	0.42429	0.26353
H4 (Game)	0.81118	0.86233	0.55127	0.44972	0.43739	0.26190

	NDCG@10	NDCG@30	P@10	P@30	MAP	MRR
H4 (Roulette)	0.77347	0.83472	0.54195	0.44421	0.42848	0.25773
H4 (Game)	0.81118	0.86233	0.55127	0.44972	0.43739	0.26190

To evaluate Hybrid4 we make the same comparison as for Hybrid3. As shown above, on Table 7, when comparing the genetic algorithm, Hybrid4, with the individual collaborative and demographic filtering algorithms, which are combined, the genetic algorithm excels in all metrics, except MRR. Similarly, in the comparison of the different parent selection methods, the Choice by Game method show a consistent performance improvement over Choice by Roulette.

7.2. MovieLens 25M

Table 8. Collaborative filtering results for the MovieLens 25M dataset.

	NDCG@10	NDCG@30	P@10	P@30	MAP	MRR
ConvQ	0.79442	0.85507	0.58550	0.51083	0.43954	0.23508
ConvM	0.79285	0.85353	0.58700	0.51017	0.43932	0.23011
Game	0.79292	0.85436	0.58400	0.51067	0.43782	0.23203

As shown above (Table 8), for the collaborative filtering, the Game case does not show better results in any metric, unlike ConvQ, which is the most efficient on all metrics except P@10. Therefore, the ConvQ case excels the other two cases.

Table 9. Content based filtering results for the MovieLens 25M dataset.

	NDCG@10	NDCG@30	P@10	P@10	MAP	MRR
ConvQ	0.75590	0.77239	0.36500	0.16383	0.42211	0.39969
ConvM	0.76139	0.78382	0.37450	0.18650	0.44950	0.40544
Game	0.76612	0.78302	0.35600	0.16033	0.42157	0.41650

Based on the results of Content Based Filtering as presented in Table 9, the ConvM case excels, since it shows better results in four out of six metrics while the Game case in only two.

Table 10. Hybrid2.1 filtering results for the MovieLens 25M dataset.

	NDCG@10	NDCG@30	P@10	P@10	MAP	MRR
ConvQ	0.73204	0.76125	0.34800	0.19433	0.41748	0.37929
ConvM	0.72591	0.75460	0.36950	0.19950	0.42829	0.39349
Game	0.73524	0.76375	0.33900	0.18933	0.40660	0.39910

Table 11. Hybrid2.2 filtering results for the MovieLens 25M dataset.

	NDCG@10	NDCG@30	P@10	P@10	MAP	MRR
ConvQ	0.74198	0.81873	0.54400	0.48967	0.40090	0.22376
ConvM	0.74414	0.81919	0.54450	0.49117	0.40493	0.21798
Game	0.74195	0.81995	0.54350	0.49217	0.40316	0.22553

As shown in Tables 10 and 11, where the results of Hybrid2.1 and Hybrid2.2 are presented respectively, the Game case yields better results in three out of six metrics and the same applies for the ConvM case. Therefore, we cannot consider that one of the two cases prevails.

Table 12. Hybrid3 results for the MovieLens 25M dataset.

	NDCG@10	NDCG@30	P@10	P@30	MAP	MRR
CF	0.79292	0.85436	0.58400	0.51067	0.43782	0.23203
CB	0.76612	0.78302	0.35600	0.16033	0.42157	0.41650
H3 (Game)	0.79365	0.85527	0.58350	0.51100	0.43846	0.23117

	NDCG@10	NDCG@30	P@10	P@30	MAP	MRR
H3 (Simple)	0.77583	0.84160	0.56850	0.50350	0.42228	0.23116
H3 (Game)	0.79365	0.85527	0.58350	0.51100	0.43846	0.23117

For the Hybrid3 algorithm, we compare the results as for the experiments in the MovieLens-100k dataset. As shown in Table 12, the Game case of the Hybrid3 outperforms the two individual algorithms combined, as it shows better results in all metrics

except MRR and P@10. Also, the Game case is a more efficient than the simple linear combination, based on all metrics.

Table 13. Hybrid4 results for the MovieLens 25M dataset.

	NDCG@10	NDCG@30	P@10	P@30	MAP	MRR
CF	0.79292	0.85436	0.58400	0.51067	0.43782	0.23203
H2.2	0.74195	0.81995	0.54350	0.49217	0.40316	0.22553
H4 (Game)	0.80882	0.86029	0.59350	0.51000	0.45030	0.23711

	NDCG@10	NDCG@30	P@10	P@30	MAP	MRR
H4 (Roulette)	0.79782	0.84916	0.58850	0.50600	0.44753	0.23860
H4 (Game)	0.80882	0.86029	0.59350	0.51000	0.45030	0.23711

As shown in Table 13, the Hybrid4 algorithm outperforms the algorithms collaborative filtering and Hybrid2.2 filtering from which the initial population is derived, as it achieves better performance based on all metrics except P@30. In addition, the method Choice by Game leads to better overall results compared to Choice by Roulette, as it is a more efficient based on all metrics except MRR.

Summing up, to draw conclusions, we categorize the algorithms based on the way they were modelled as games, thus distinguishing three categories. The first category includes the algorithms in which the "Game of Recommendation Process" is used, i.e. the algorithms collaborative filtering, demographic filtering, content based filtering and their combinations, Hybrid1 and Hybrid2. The second category concerns the Hybrid3 algorithm, which uses the "Game of Recommendation Methods Combination" and finally, the third category, which includes the Hybrid4 genetic algorithm where the "Game of Individual Selection" is used, where each parent is considered player of a cooperative game. For the first category, the initial experiments in the MovieLens-100k dataset lead to the conclusion that the recommendation process can be improved if modelled as a game, since in all algorithms except Hybrid1.2, the Game case provides better results than other cases. This conclusion is overturned by the experiments in the MovieLens-25M dataset since the Game case does not outperform in any algorithm. For the experiments in the MovieLens 25M dataset, the same parameters as for MovieLens 100k were used instead of being experimentally optimized, because the experiments were very time consuming. Presumably, the fact that the same parameters were used is responsible for the reduced performance of the Game case. For the second category, i.e. for the Hybrid3 algorithm, although different combinations were performed for MovieLens 100k and MovieLens 25M, in both cases the weighted combination using game outperformed the individual algorithms combined and the simple weighted combination as well. Finally, for the third category, the Hybrid4 algorithm, experiments in the MovieLens 100k dataset show that the use of the top-k users of the collaborative filtering and demographic filtering algorithms as the initial population of the genetic algorithm results in better recommendations than the two algorithms, if the Choice by Game method is used, while the same does not apply for the Choice by Roulette method. This conclusion is confirmed and reinforced by the experiments in the MovieLens-25M dataset since the results are the same although different algorithms were combined.

8. Conclusions & Future Work

As mentioned above, the purpose of the present study is to develop recommendation systems using game theoretic techniques to further inquire and confirm the claim that the recommendation process can be improved by utilizing game theory[36]. After a series of experiments on different datasets for recommendation algorithms developed using game theoretic techniques, we conclude that a recommendation system can become more efficient when game theory is integrated to it. The algorithms developed do not only concern the recommendation systems, since they can be applied to information retrieval systems as well as in machine learning in general. An important finding is the method Choice by Game as it brings about a significant performance improvement over the existing method Choice by Roulette.

For the future work, better integration of Nash equilibrium theory into algorithms may lead to improved results. In particular, the execution of the Hybrid3 and Hybrid4 algorithms is interrupted when an equilibrium state is found, but this state is probably a local maximum of the utility function and not a total one. So, algorithms need to be further developed to find the total maximum. Regarding the Hybrid3 algorithm, it would be interesting to adapt to more players, i.e. in a linear combination of more than two methods of recommendations [9]. Finally, it is important to further research and apply techniques of game theory in more methods and scenarios of recommendations and Machine Learning in general, since it is possible that new optimization techniques will emerge for this field.

References

1. Azadjalal, M.M., Moradi, P., Abdollahpouri, A.: Application of game theory techniques for improving trust based recommender systems in social networks. In: 2014 4th International Conference on Computer and Knowledge Engineering (ICCKE). pp. 261–266 (2014)
2. Azam, N., Yao, J.: Game-theoretic rough sets for recommender systems. *Knowledge-Based Systems* 72 (12 2014)
3. Bashiri, P.: Recommender systems: Survey and possible extensions (08 2018)
4. Ben-Porat, O., Tennenholtz, M.: A game-theoretic approach to recommendation systems with strategic content providers. *arXiv (Nips)* (2018)
5. Bobadilla, J., Ortega, F., Hernando, A., Gutiérrez, A.: Recommender systems survey. *Knowledge-Based Systems* 46, 109–132 (2013), <https://www.sciencedirect.com/science/article/pii/S0950705113001044>
6. Branzei, R., Dimitrov, D., Tijs, S., Ebooks Corporation.: *Models in cooperative game theory : Crisp, fuzzy, and multi-choice games* p. 137 (2005)
7. Burke, R.: Hybrid web recommender systems. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 4321 LNCS, 377–408 (2007)
8. Çano, E., Morisio, M.: Hybrid recommender systems: A systematic literature review. *Intelligent Data Analysis* 21(6), 1487–1524 (2017)
9. Davis, M., Maschler, M.: *The kernel of a cooperative game* (1965)
10. Halkidi, M., Koutsopoulos, I.: A game theoretic framework for data privacy preservation in recommender systems. pp. 629–644 (09 2011)
11. Harman, D.: Relevance feedback revisited. In: *Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. p. 1–10.

- SIGIR '92, Association for Computing Machinery, New York, NY, USA (1992), <https://doi.org/10.1145/133160.133167>
12. Harper, F.M., Konstan, J.A.: The movielens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems* 5(4) (2015)
 13. Hernández del Olmo, F., Gaudioso, E.: Evaluation of recommender systems: A new approach. *Expert Systems with Applications* 35(3), 790–804 (2008)
 14. Hwang, W.S., Li, S., Kim, S.W., Lee, K.: Data imputation using a trust network for recommendation via matrix factorization. *Computer Science and Information Systems* 15(2), 347–368 (2018)
 15. Kim, T.K.: T test as a parametric statistic. *Korean J Anesthesiol* 68(6), 540–546 (2015), <http://ekja.org/journal/view.php?number=8123>
 16. Lavrenko, V., Croft, W.B.: Relevance models in information retrieval. *Language Modeling for Information Retrieval* pp. 11–56 (2003)
 17. Lü, L., Medo, M., Yeung, C.H., Zhang, Y.C., Zhang, Z.K., Zhou, T.: Recommender systems. *Physics Reports* 519(1), 1–49 (2012)
 18. Myerson, R.B.: *Game Theory: Analysis of Conflict*. Harvard University Press (1991), <http://www.jstor.org/stable/j.ctvjjsf522>
 19. Nisan, N., Roughgarden, T., Tardos, É., Vazirani, V.V.: *Algorithmic game theory*, vol. 9780521872829. Cambridge University Press, Cambridge (2007)
 20. Nussbaumer, A., Dahrendorf, D., Schmitz, H.C., Kravčík, M., Berthold, M., Albert, D.: Recommender and guidance strategies for creating personal mashup learning environments. *Computer Science and Information Systems* 11, 321–342 (1 2014)
 21. Quadrana, M., Karatzoglou, A., Hidasi, B., Cremonesi, P.: Personalizing session-based recommendations with hierarchical recurrent neural networks. *RecSys 2017 - Proceedings of the 11th ACM Conference on Recommender Systems* pp. 130–137 (2017)
 22. Rendle, S., Zhang, L., Koren, Y.: On the difficulty of evaluating baselines: A Study on Recommender Systems. *arXiv* pp. 1–19 (2019)
 23. Resnick, P., Varian, H.R.: Recommender systems. *Commun. ACM* 40(3), 56–58 (Mar 1997), <https://doi.org/10.1145/245108.245121>
 24. Robertson, S.E., Jones, K.S.: Relevance weighting of search terms. *Journal of the American Society for Information Science* 27(3), 129–146 (1976), <https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.4630270302>
 25. Robertson, S., Zaragoza, H.: The probabilistic relevance framework: Bm25 and beyond. *Foundations and Trends in Information Retrieval* 3, 333–389 (01 2009)
 26. Rocchio, J.J.: Relevance feedback in information retrieval. In: Salton, G. (ed.) *The Smart retrieval system - experiments in automatic document processing*, pp. 313–323. Englewood Cliffs, NJ: Prentice-Hall (1971)
 27. Ross, P., Corne, D.: Applications of genetic algorithms. In: *ON TRANSCOMPUTER BASED PARALLEL PROCESSING SYSTEMS,” LECTURE*. University of Edinburgh (1995)
 28. Shani, G., Gunawardana, A.: Evaluating Recommendation Systems, vol. 12, pp. 257–297 (01 2011)
 29. Ter Hofstede, A.H., Proper, H.A., Van Der Weide, T.H.: Query formulation as an information retrieval problem. *Computer Journal* 39(4) (1996)
 30. Wang, J., de Vries, A.P., Reinders, M.J.T.: Unifying user-based and item-based collaborative filtering approaches by similarity fusion. In: *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. p. 501–508. SIGIR '06, Association for Computing Machinery, New York, NY, USA (2006), <https://doi.org/10.1145/1148170.1148257>
 31. Xu, J., Croft, W.B.: Query expansion using local and global document analysis. *SIGIR Forum* 51(2), 168–175 (Aug 2017), <https://doi.org/10.1145/3130348.3130364>
 32. Xu, Z., Jiang, H., Kong, X., Kang, J., Wang, W., Xia, F.: Cross-domain item recommendation based on user similarity. *Computer Science and Information Systems* 13, 359–373 (6 2016)

33. Zhai, C.: Towards a game-theoretic framework for text data retrieval. *IEEE Data Eng. Bull.* 39, 51–62 (2016)
34. Zhai, C., Lafferty, J.: A study of smoothing methods for language models applied to information retrieval. *ACM Trans. Inf. Syst.* 22(2), 179–214 (Apr 2004), <https://doi.org/10.1145/984321.984322>
35. Zhang, S., Yao, L., Sun, A., Tay, Y.: Deep learning based recommender system: A survey and new perspectives. *ACM Comput. Surv.* 52(1) (Feb 2019), <https://doi.org/10.1145/3285029>
36. Zou, S., Tao, G., Wang, J., Zhang, W., Zhang, D.: On the equilibrium of query reformulation and document retrieval. *ICTIR 2018 - Proceedings of the 2018 ACM SIGIR International Conference on the Theory of Information Retrieval* pp. 43–50 (2018)

Evangelos Sofikitis is currently working as a Junior Software Engineer at Oxa, a highly efficient, analytical, distributed Database. As of 2021, he is a graduate of the Department of Computer Engineering and Informatics, University of Patras. His main field of interest during his studies was Machine Learning and Recommendation Systems.

Christos Makris is an Associate Professor in the University of Patras, Department of Computer Engineering and Informatics, from September 2017. Since 2004 he served as an Assistant Professor in CEID, UoP, tenured in that position from 2008. His research interests include Data Structures, Information Retrieval, Data Mining, String Processing Algorithms, Computational Geometry, Internet Technologies, Bioinformatics, and Multimedia Databases. He has published over 100 papers in refereed scientific journals and conferences and has more than 600 citations excluding self-citations (h-index: 18).

Received: September 30, 2021; Accepted: May 10, 2022.

Effective Methods for Email Classification: Is it a Business or Personal Email?

Milena Šošić and Jelena Graovac

Faculty of Mathematics, University of Belgrade
Studentski Trg 16, 11000 Belgrade, Serbia
pd202030@alas.matf.bg.ac.rs
jgraovac@matf.bg.ac.rs

Abstract. With the steady increase in the number of Internet users, email remains the most popular and extensively used communication means. Therefore, email management is an important and growing problem for individuals and organizations. In this paper, we deal with the classification of emails into two main categories, Business and Personal. To find the best performing solution for this problem, a comprehensive set of experiments has been conducted with the deep learning algorithms: Bidirectional Long-Short Term Memory (BiLSTM) and Attention-based BiLSTM (BiLSTM+Att), together with traditional Machine Learning (ML) algorithms: Stochastic Gradient Descent (SGD) optimization applied on Support Vector Machine (SVM) and Extremely Randomized Trees (ERT) ensemble method. The variations of individual email and conversational email thread arc representations have been explored to reach the best classification generalization on the selected task. A special contribution of this paper is the extraction of a large number of additional lexical, conversational, expressional, emotional, and moral features, which proved very useful for differentiation between personal and official written conversations. The experiments were performed on the publicly available Enron email benchmark corpora on which we obtained the State-Of-the-Art (SOA) results. As part of the submission, we have made our work publicly available to the scientific community for research purposes.

Keywords: Email classification, business, personal, deep learning, BiLSTM, SGD, BERT embeddings, Tf-Idf, lexicons, NLP.

1. Introduction

In the last decade, emails have become one of the crucial media for both personal and business communication. Despite the rise of social media and instant messaging, email usage is steadily growing, with more than 4 billion users worldwide in 2021 and about 6.8 billion email accounts – and it continues to grow [21]. This is mainly due to their efficiency, low cost, and compatibility with diversified types of information. In the last decade, emails have become one of the crucial media for both personal and business communication. Despite the rise of social media and instant messaging, email usage is steadily growing, with about 6.8 billion email accounts, more than 4 billion users and about 320 billion sent and received emails per day worldwide in 2021 – with expectations for numbers to further increase by 2025 [21]. This is mainly due to their efficiency, low cost, and compatibility with diversified types of information. Observed trend has made

the automatic processing of emails more than desirable. For example, the classification of emails into Business and Personal categories can help a lot in better handling the email inbox and decreasing the time spent managing emails every day. This trend has made the automatic processing of emails more than desirable. For example, the classification of emails into Business and Personal categories can help a lot in better handling the email inbox and decreasing the time spent managing emails every day.

To facilitate usage of emails and explore business potentials in emailing, various studies have been proposed such as spam-filtering [17], multi folder classification [25], phishing email classification [1], etc. In this paper, we focus on the classification of emails into two main categories, Business and Personal, which belongs to the text classification task [7], [8].

Unlike other email processing tasks, such as spam filtering, this problem has not received much attention, and it remains a challenging task. One of the reasons for that is a lack of data - personal emails are often highly private, and they are usually unavailable for research purposes. In this study for training and testing purposes, we used two different distributions of the Enron email corpus [14], the sole email corpus that is freely available (public and not licensed).

The main contributions of this paper are as follows:

- Conducting a comprehensive set of experiments using advanced deep learning and traditional machine learning (ML) techniques.
- Experimentation with different variants of individual emails, and conversational thread arcs of emails.
- Experimentation with different text representation techniques on words, word n-grams, character n-grams, and BERT embeddings.
- Extraction of different lexical, conversational, expressional, emotional, and moral features using a diverse set of lexicons and email content characteristics.
- Extensive comparison and evaluation of the obtained results.
- Production of the State-Of-the-Art (SOA) results.

The paper continues with a review of the related work in section 2. It is followed by the presentation of our approach in section 3 including preprocessing, features extraction, and used traditional and deep learning ML techniques. After that, in section 4, the experimental framework is described. The obtained results are expounded in section 5, while section 6 presents the results of comparison with previously published SOA techniques. Section 7 concludes the paper.

We make our work publicly available and reproducible ¹.

2. Related Work

Since Enron is the only freely available email data set, many researchers have worked on it with different tasks. To our knowledge, the previous efforts most closely related to our research are [12], [2], and [3]. They all have worked on the same problem: classification of emails into Business or Personal category and used the same Enron data set for training and testing.

¹ <https://github.com/milena-sosic/Email-Business-Personal>

First attempts to categorize corporate emails into Business and Personal categories were made by [12]. The main contribution of this paper is the largest scale annotation project involving the Enron email data set. Over 12,500 emails were classified by humans, into the Business and Personal categories. They used inter-annotator agreement to evaluate how well humans perform this task. They also used a probabilistic classifier based upon the distribution of distinguishing words, to determine the feasibility of separating business and personal emails by machine.

In [2] and [3], the authors trained their models on the Enron data set, and tested them on the Enron and Avocado data sets. In [2], the authors represented the email exchange networks as social networks with graph structures. They used social networks features from the graphs in addition to pre-trained GloVe embedding vectors as lexical features from email content to improve the performance of SVM and Extra-Trees classifiers. As a supplementary contribution to this paper, the authors also provided manually annotated sets of the Enron and Avocado email corpora. In [3], the same authors additionally considered the thread structure of emails which improved the performance further. They also used node embedding based on both lexical and social network information. All results presented here are used in Section 6 for comparison purposes.

There are a lot of other research papers that cover solving different email processing tasks, such as spam-filtering, multi folder classification, phishing email classification [26], but we have focused here only on the papers most related to our research.

3. Our Approach

The rich textual structure of email has a predefined format in which two main segments have been identified: a header and conversational content. Our approach exploits useful information from both of them. The textual features used in the classification process are based on the conversational content only, ignoring the content of email headers, e.g. dates, personal names, etc. Email domains identified with regular expressions from the headers are added to the end of the email content (the most recent email or the most recent email with quote messages from the same thread arc). We have found that the result of this is that the words such as 'hotmail' and 'yahoo' are characteristics of the Personal class (see Table 2). Personal communication often happens between people outside the organization and email domains could be an indicator of it. The architecture of our approach is presented in Fig. 1.

Based on the fact that emails from the same thread, and especially from the thread arc, usually belong to the same Business/Personal class [3], we have split our experiments based on the content used as follows:

- The most recent email (E – baseline)
- The most recent email with domains found in headers (ED)
- The most recent email with quote messages from the same thread arc (EQ)
- The most recent email with quote messages from the same thread arc and domains found in headers (EQD)
- A whole email thread arc found in the body field (B – baseline)

In all mentioned cases, the subject is added to the email content. To obtain some new/fresh insights and results, we have analysed user writing behavior in the business environ-



Fig. 1. The architecture of the proposed approach for effective emails classification into Business and Personal classes

ment and conversational context from a lexical, conversational, expressive, emotional, and moral perspectives.

3.1. Data

The Enron email data set consists of both personal and business emails from over a hundred Enron employees over a period for 3.5 years (1998 to 2002). It was made publicly available by the Federal Energy Regulatory Commission (FERC) during the legal investigation of the company's collapse. The corpus was first processed and released by Klimt and Yang (2004) at Carnegie Mellon University (CMU), and this CMU data set has later been re-processed by several other research groups. In our experiments, we use the version of annotated Enron data set presented in the article [2] which we call *Enron Columbia* and denote with *Enron_C*. This data set is annotated as follows:

- Business, with clearly professional content;
- Somehow Business, with professional content but with some personal parts;
- Mixed, with combined both professional and personal content;
- Somehow Personal, with personal content but with some business-related parts;
- Personal, with clearly personal content;
- Cannot Determine, with not enough content to determine the category.

In our experiments, these categories were merged into two categories: Business and Personal in the following way:

- Personal: Personal, Somehow Personal, Mixed;
- Business: Business and Somehow Business.

Fig. 2 presents *Enron_C* data set through a frequency scatter plot² of the words present in the Business and Personal categories. The word frequency metric is what scattertext uses as the coordinates for each point. The x-axis indicates the frequency in the Business category: if a word frequently appears in business emails, it is placed on the right. Similarly, the y-axis encodes the frequency in the Personal category. A word that frequently appears in personal emails will be placed on the top area. Consequently, the areas where the more frequent words appear is of particular interest: top left (frequent in personal emails), bottom right (frequent in business emails), and top right of the figure (frequent in both personal and business emails). These areas offer a view of how the words are distributed in these two categories. For example, common Business words such as 'agreement', 'energy', and 'attachment', stress the official tone that can be found in the narratives of the business emails. In contrast, the emails found in the Personal category have a more relaxing tone, frequently using words such as 'love', 'weekend', and 'fun'. The colors express the value of the score called "scaled F-Score" introduced by the authors in [13]. Words with scores near zero, colored in yellow and orange in the plot, have frequencies that are similar for both classes. These words are not of great importance. When the frequency of the word is dominated by one class, scores are shifting to -1 (Business) or 1 (Personal), marked with red or blue color respectively. The darker the color of red or blue indicates the higher dominance of the word for the corresponding class. Another version of the Enron data set is annotated by Berkeley students in Applied Natural Language Processing Course (ANLP), so we call it *Enron Berkeley* and denote it with *Enron_B*. They developed a set of hierarchical categories and the selected subset of emails focusing on business-related emails. Each email message was annotated by two people and got assigned multiple labels at once. The data set contains 1702 emails that were categorized into 53 topic categories, such as company strategy, humor, and legal advice. It has been mainly used as a benchmark data set for multi-label classification. In our experiments these categories were merged as:

- Personal: Purely Personal, Personal in a professional context and Private humor;
- Business: all other categories which are related to business policies, strategy, legal notes or regulations.

The numbers of emails for each category in both Enron distributions (*Enron Columbia* and *Enron Berkeley*) are presented in the Table 1, while top business and top personal words in both data sets are presented in Table 2.

3.2. Preprocessing and Text Representation

The content of the emails is represented in the forms of vectors of word frequencies (or Bag of Words denoted as BoW in the following part of the text), Tf-Idf vectors on n-grams and n-gram characters as well as BERT embeddings. Elements of Tf-Idf matrix are calculated using the formula presented in equation 1:

² <https://github.com/JasonKessler/scattertext>

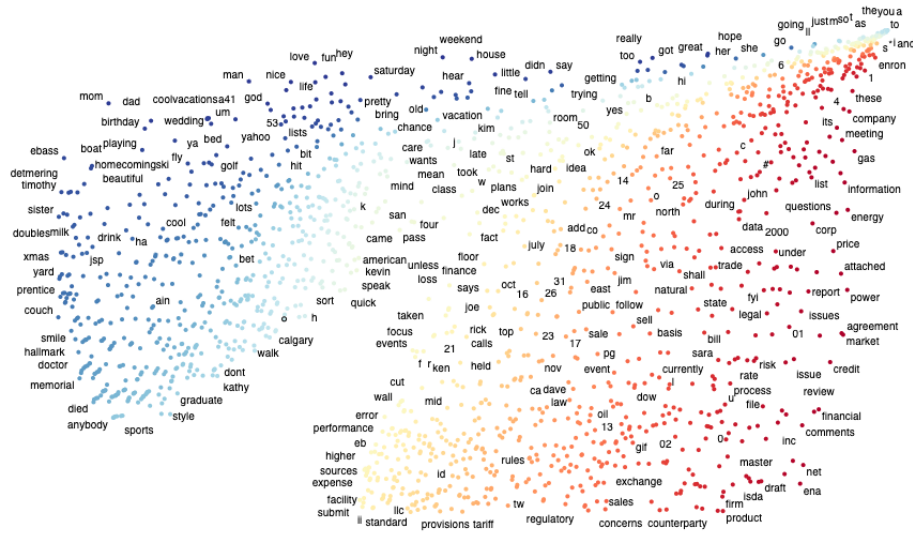


Fig. 2. Characteristic words for Business and Personal classes. 'Love', 'weekend', 'fun' in Personal and 'agreement', 'energy', 'attachment', words containing numbers in Business are among the most dominant words

Table 1. Summary of Enron data sets before and after processing empty and duplicate emails

Data Set	Business	Personal	Total
<i>Enron_C</i>	9738 (86.5%)	1523 (13.5%)	11261
<i>Enron_{Cp}</i>	8651 (86.6%)	1340 (13.4%)	9991
<i>Enron_B</i>	1491 (87.6%)	211 (12.4%)	1702
<i>Enron_{Bp}</i>	1276 (88.0%)	173 (11.9%)	1449

Table 2. Dominant words for Business and Personal class and characteristic words for the whole corpus

data set	Top Business	Top Personal	Corpus Characteristic
<i>Enron_C</i>	energy, agreement, informa- tion, power, market, attached, end, hey, msn, man, mom, ya- gas, price, trading, issues, hoo, fun, god, really, game, terparty, ect, cpuc, com, org, trading, credit, review, questions, con- tract	love, hotmail, night, week- end, hey, msn, man, mom, ya- house	enron, ferc, skadden, hotmail, isda, http, attached, dynegy, aol, fyi, carrfut, coun- nymex, eol, thanks, www, enrononline, gas, tomorrow, calpine, pge
<i>Enron_B</i>	state, gas, price, energy, market, electricity, utility, congratulations, london, stu- power, blackouts, billion, dio federal, percent	energy, thanks, sorry, great, love, life, 2001, blackouts, enron, dynegy, edison, generators, electricity, 2000, megawatt, deregulation, gov	

$$w_{i,j} = tf_{i,j} * \log\left(\frac{N}{df_i}\right) \quad (1)$$

where $w_{i,j}$ is Tf-Idf weight for token i in document j , tf_i is the number of occurrences of token i in document j , df_i is the number of documents that contain token j and N is the total number of emails in the training set. Vocabulary size for different BoW and Tf-Idf text representations and experiments is presented in the Table 3.

Lemmatization was included in the data preprocessing stage for verbs, nouns, adjectives, and adverbs. For lemmatization we used WordNetLemmatizer from NLTK³ python package. We had special treatment of numbers, personal names, punctuations, spaces, and contractions. Also, we defined corpus-specific stop words. The author in [13] introduces measures of precision and recall for the words in the corpus and explains their inverse relationship. Precision is a word's discriminative power regardless of its frequency, while recall denotes the frequency at which a word appears in a particular class, or $P(word|class)$. For visual interpretation, the words with high recall values tend toward the top right-hand corner of the chart, while the words with high precision values tend toward the axes (See Fig. 2). The revelation that extremely high recall words tend to be stop words is used for the creation of the corpus specific list of stop words.

To compare different preprocessing and text representation techniques, we performed a large set of experiments using SGD-SVM ML algorithm. As it is presented in the Fig. 3.2, we came to conclusion that word 2-grams outperforms word n-grams of other lengths ($n = 1$ or $n > 2$). Moreover, Tf-Idf weights outperform frequency weights which were widely used in the previous publications on the same task. Additionally, word n-grams outperform character n-grams. Limiting minimum or maximum number (or percentage) of allowed token appearance across the corpus does not improve model performance by any means. Incorporation of lemmatization and custom defined stop words plays an important role in improving the model performance, together with the removal of personal names and punctuation tokens. However, limitation of vocabulary size decreases model performance. From all of these points, best resulting preprocessing actions have been applied on the raw text, resulting in the text representation used in the following experimentation steps.

Table 3. Vocabulary size for different text representations and experiments - E - the most recent email, ED - email with domains, EQ - email with quotes, EQD - email with quotes and domains, B - body used as email content

Experiment	BoW/Tf-Idf (1,1)	Tf-Idf-Ngram (1,2)	Tf-Idf-Ngram-Char (1,4)	BERT embeddings
E	22381	217186	100417	30522
ED	23624	223595	105819	30522
EQ	29074	320018	123615	30522
EQD	30257	325929	128274	30522
B	33099	362959	138409	30522

Another technique for emails representation used in our work is BERT embeddings vectors. Word embeddings techniques aim to use continuous low-dimension vectors rep-

³ <https://nltk.org/>

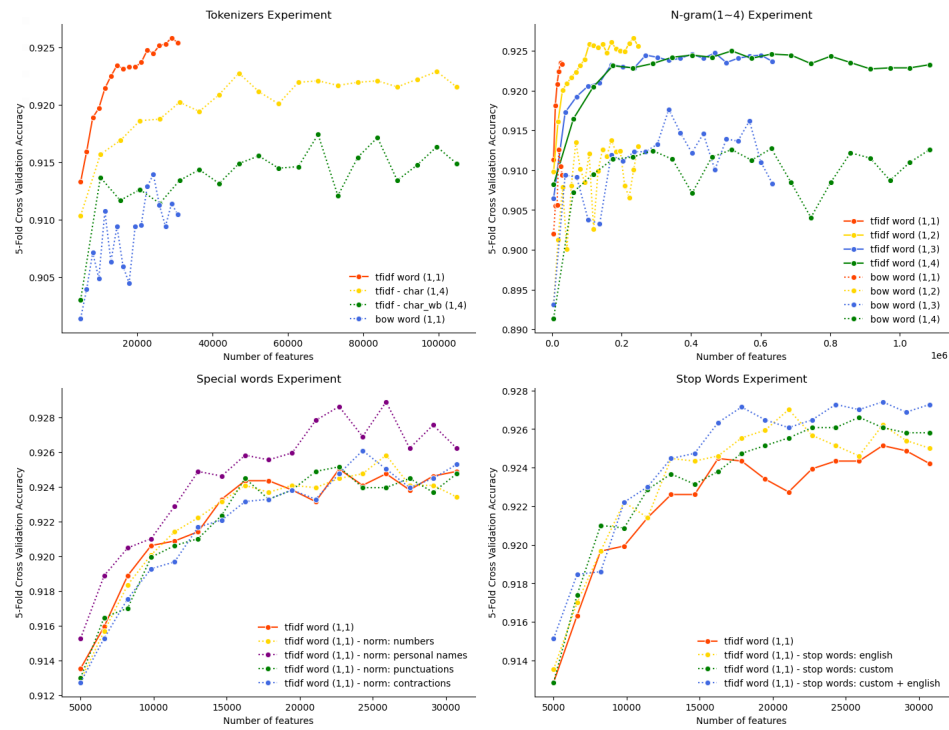


Fig. 3. The results of testing preprocessing steps for different weight and token types, n-gram length, stop words and special words selection. Used SGD classifier with default set of parameters in email with domains (ED) experiment

representing the features of the words captured in context [15]. A compact pre-trained BERT word embedding model, pre-trained on Wikipedia and BookCorpus, was selected from Google’s TensorFlow Hub⁴ repository. This model, with L=2 hidden layers (i.e., transformer blocks), a hidden size of H=256, and A=4 attention heads, was used for initializing the word embedding layer (the input layer) of all our deep learning models. Model architecture and training objectives from the standard BERT model is replicated to a wide range of model sizes, making smaller BERT models applicable for environments with restricted computational resources. They can be fine-tuned in the same manner as the original BERT models [27]. Descriptive statistics of email content lengths for different input emails (E, ED, EQ, EQD, B) and Tukey’s outliers rule, helped us to set appropriate thresholds for maximum sequence length. In the case of E and ED, it is set to 256, while in the case of EQ, EQD and B, it is set to 512 to gather most of the information from the data.

⁴ <https://tfhub.dev/s?module-type=text-embedding>

3.3. Additional Features

Set of additional lexical (including punctuation-based and NER-based), conversational, expressional, emotional, and moral features has been extracted to analyse conversational context of exchanged emails.

Lexical Features (Lex) capture various counts and ratios associated with the subject and content of the email. Text classification extensively relies on such features and hence we hypothesize that the lexical properties will contribute to our task. Syntactic features include NER-based features, number of lines, number of noun phrases, syllables, difficult words which contain more than one syllable, average syllable per word (ASPW) and sentence (ASPS), sentence and word density. ASPW, sentence and word densities are defined with equations 2, 3 and 4 respectively:

$$ASPW = \frac{\#syllables}{\#words} \quad (2)$$

$$sentence\ density = \frac{\#sentences}{1 + \#lines} \quad (3)$$

$$words\ density = \frac{\#words}{1 + \#spaces} \quad (4)$$

where #sentences, #words, #lines and #spaces denote number of, sentences, words, lines (including blank lines) and blank spaces in email content respectively. To identify syllables, *pyphen* python package is used, while remaining features in this group are calculated with *textstat* package.

The business indicator is a numerical feature representing the ratio of business terms in the content. Business terms are identified using Business Thesaurus⁵ dictionary containing terms, expressions, and terminology used in business conversations. The ratio of abbreviations in the content is noted as an acronyms indicator. Abbreviations are identified using Abbreviations and Acronyms Dictionary⁶ together with regular expressions to fine tune their finding in the email content.

Punctuation-based features (Punct) measure the presence of dots, question marks, exclamation marks, hash and reference tags with their ratio among the whole punctuation characters found in the email content.

NER-based features (NER) are numerical representations of the NER tags presence in the content. The ratio of personal names, organization names, words containing numbers, words in English language marked as connectors (e.g. 'in', 'the', 'all', 'for', 'and', 'on', 'but', 'at', 'of', 'to', 'a'), month and day names, and a valid email and URL addresses are incorporated in the list of features.

Conversational Features (Conv) are extracted from email signatures containing the number of mail recipients and information about a conversation with external email domains. For that purpose we use free email domains dictionary⁷. The free domains ratio is the ratio of free domains presented in the email content including signature among all

⁵ <https://www.businessballs.com/glossaries-and-terminology/business-thesaurus-290/>

⁶ <https://abbreviations.yourdictionary.com/>

⁷ <https://github.com/willwhite/freemail>

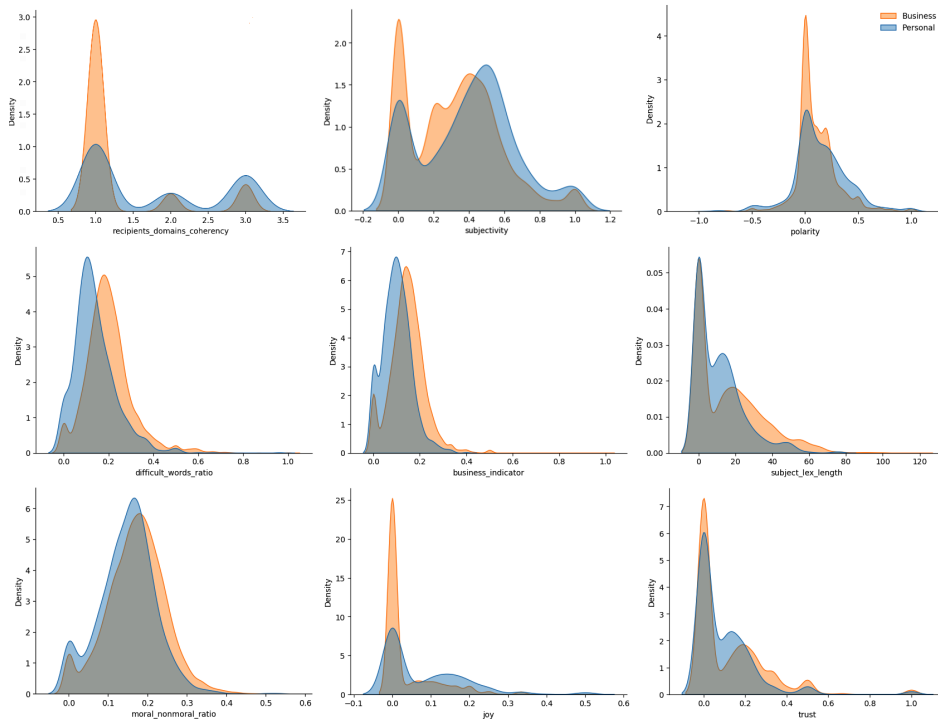


Fig. 4. Variation in the features distributions in the Business and Personal classes

domains found there. Recipients domains coherency is the feature created to capture coherency between recipients domains e.g. if they all belong to the default company domain, external domains or if recipient domains are of a mixed structure.

Expressional Features (Expr) capture information such as readability of text, subjectivity and polarity. Subjectivity and polarity are based on the TextBlob⁸ implementation. Polarity is a float that lies in the range of $[-1,1]$ where 1 denotes a positive statement and -1 denotes a negative statement. Subjective sentences refer to personal opinion, emotion or judgment, whereas objective ones refer to factual information. Subjectivity is presented as a float number that lies in the range of $[0,1]$ closer to 1 in a more subjective context. Readability is measured based on Automated readability index (ARI) and Flech Reading Ease Score (FRES), which are calculated by equations 5 and 6:

$$ARI = 4.71\left(\frac{\#characters}{\#words}\right) + 0.5\left(\frac{\#words}{\#sentences}\right) - 21.43 \quad (5)$$

$$FRES = 206.835 - 1.015\left(\frac{\#words}{\#sentences}\right) - 84.6\left(\frac{\#syllables}{\#words}\right) \quad (6)$$

where $\#characters$, $\#words$, $\#sentences$ and $\#syllables$ denote the number of letters and numbers, words, sentences and syllables in the text.

⁸ <https://textblob.readthedocs.io/en/dev/>

The LWM Algorithm is giving a 'grade level' measure, reflecting the estimated years of education needed for reading the text fluently. ARI and FRES scores measure how easy it is to read a text. We use textstat⁹ python package for the implementation of that.

Table 4. Summary of features. Meta refers to all extracted features combined together

Feature Group	Features List	# of Features
Lexical	Number of characters and words in content and subject, sentences count, average sentence length, average word length, noun phrases, average syllables per word, average syllables per sentence, sentence and word density, difficult words, business indicator, acronyms indicator	16
	NER-based	8
	Punctuation-based	5
Conversational	Free domains in headers ratio, number of recipients, recipients domains coherency	3
Expressional	Automated Readability Index(ARI), Flech Reading Ease Score(FRES), Linsear Write Metric(LWM), content subjectivity, content polarity	5
Moral	Probability measures of care, sanctity, authority, loyalty and fairness on word and sentence, moral/non-moral ratio	11
Emotional	Measures of trust, joy, anger, disgust, sadness, fear, surprise, positive, negative	9
All features (Meta)		57

Emotional Features (Emo) use the Plutchik's approach [19] which postulates the following eight basic human emotions: joy, sadness, anger, fear, trust, disgust, anticipation, and surprise, extending a simple positive-negative dichotomy to capture the full range of emotions. There have been extensive applications of this approach, for example, the National Research Council (NRC) Word-Emotion Association Lexicon which contains 10,170 lexical items that are coded for Plutchik's basic human emotions [16]. Plutchik's categories also have the advantage of providing a balanced list of positive (trust, joy, anger, and anticipation) and negative (disgust, sadness, fear, and surprise) emotions. To the best of our knowledge, they have not been applied in business conversation analysis in general, or email content analysis in particular. We use the python NRCLex¹⁰ package which expands the lexicon to 27,000 words based on WordNet synonyms and effectively measures the emotional effect on the categories.

Moral Features (Mor) are based on Moral Foundations Theory (MFT), a framework for explaining variation in people's moral reasoning [6]. The framework decomposes the types of moral evaluations people make into five foundations: Authority/Subversion, Care/Harm, Fairness/Cheating, Loyalty/Betrayal, Sanctity/Purity. The emphasis on moral

⁹ <https://pypi.org/project/textstat/0.1.6/>

¹⁰ <https://pypi.org/project/NRCLex/>

foundations is most commonly inferred from written text (speech acts) by flagging combinations of words that have validated connections to each foundation. Recent behavioral research has focused on developing extended vocabulary sets with human ratings for mapping large sets of terms onto various moral foundations [11].

In this paper, we use an eMFD¹¹ python package to calculate the moral sentiment in emails. Each email document is assigned to five foundation probabilities that denote the average probability of each document belonging to one of the five moral foundations and five sentiment scores that describe the average sentiment of detected moral words for that foundation. In addition, the moral-to-nonmoral word ratio has been added to the list of moral features for that document.

All groups of features are merged into a single list, denoted with Meta, which contains 57 features in total (see Table 4). L2 normalization technique is applied to all constructed features, rescaling the features vector representation of each document to have Euclidean norm equals to 1. The graphs on Fig. 4 show how specific features vary in their distributions for the Business and Personal classes. Subject length, difficult and moral words ratio tend to have higher values for Business class. The Personal class has higher subjectivity and polarity scores, as well as words assigned to joy. Recipients domain coherency has higher values in the Personal class for the recipients domains outside the default organization domain (value 3).

When features are collinear, dropping one feature will have little effect on the model's performance because it can get the same information from a correlated feature. Multicollinear and correlated features from our set of Meta features were removed by performing hierarchical clustering on the Spearman rank-order correlations, with a threshold of 0.7, and a single feature from each cluster is kept. On the retrieved 'clean list' of features, we performed drop-column algorithm to measure the impact of each feature on the variance of the default model accuracy and use this measure as the driver for feature selection. With a threshold of 0.01, the final set of uncorrelated and the most important Meta features for each experiment was selected. We can observe that features from each predefined group of features take their role among the most important features on average, but also in each of the examined experiments. The most important features from each group are: joy, fear and trust (Emo); fairness, authority and loyalty (Mor); names, numbers and connector ratios (NER); exclamation and dots ratio (Punct); recipient domains coherency and free domains ratio (Con); subjectivity, polarity, FRES, LWF and ARI (Expr); business and acronyms indicators, average syllables per word, average word length, subject length, sentence and word density (Lex). Moreover, Expr, Mor and Emo features has the highest tendency towards the top. It is noticeable that positive/negative (NRCLex) and polarity (TextBlob) features which are extracted using different packages and their lexicons have different scores in B experiment, which is not expected. It could be due to the differences in lexicons for these categories as well as the characteristics of particular content experiment. Moreover, in B experiment, some of the most important features such as names and connector ratios, FRES, difficult words, and free domains ratios indicate that header content which is included in the email body has an important role in distinguishing Business/Personal classes. Personal names, dates, and domain names (especially corporate domain enron.com), together with connectors and punctuation characters which can be found in reply and forward email headers are data set dependant. We have tried to avoid

¹¹ <https://github.com/medianeuroscience/emfd>

this dependency by using another email content structure through our experiments with an ambition to improve classification generalization on other data sets.

3.4. Machine Learning Techniques

Traditional Learning. The learning process in ML is producing the function f by processing the samples of the training set (E). The function f maps email content E_n to one of the classes $C_k, k = 2$. The email content for each E_k is represented with the numeric features vectors. Therefore, as it is described with equation 7, feature vector extractor ϕ computes vectors of features for each email from E :

$$\phi(E_k) = (\phi_1(E_k), \dots, \phi_d(E_k)), \phi(E) \in R^d \quad (7)$$

representing a point in the d dimensional feature space. Moreover, the parameter vector that specifies the contributions of feature vectors to the prediction output is given with equation 8:

$$P = P_1, \dots, P_d, P \in R^d \quad (8)$$

Consequently, in equation 9, we mathematically express f by assembling both $\phi(E)$ and P :

$$f = \phi(E) * P \quad (9)$$

The Gradient Descent optimization algorithm aims to find the coefficient of f with a condition that minimizes the cost of the inaccuracy of predictions. It uses different coefficient values and the cost function estimates their values through the predicted results for each sample of the training set. The aforementioned process occurs by comparing the prediction result with the actual value to choose the lowest loss. The algorithm tries different coefficient values to look for lower loss. The learning rate is used to update the coefficients for the next iteration. Such calculation is very expensive as the cost is computed over the entire training data set in each iteration. On the other hand, Stochastic Gradient Descent updates the coefficient for each training sample instead at the end of the iteration over all samples of the training set. We will apply the Stochastic Gradient Descent optimization technique in section 4 on a diverse set of linear classifiers and choose the best one for our task.

Extremely Randomized Trees (ERT) is an algorithm for building decision tree ensembles, for both supervised classification and regression problems. The best splitting attribute is selected for each node from a random subset of attributes. Including randomness in the cut-point choice, the algorithm builds an ensemble of decision trees whose structure is independent of the output values [5].

A comparison of the SGD and ERT classifiers performances was made by using their implementations from scikit-learn python library [18].

Deep Learning. With a successful initial application to computer vision problems, Convolutional Neural Networks (CNNs) confirmed their good performance in NLP [29]. CNNs are able to extract the local n-gram features, having difficulty with capturing long-distance dependencies.

Recurrent Neural Networks (RNNs) can capture dynamic information in serial data by recurrently connecting the hidden layer nodes. RNNs can store a state of context, learn

and express relevant information in any long context window, unlike CNN’s fixed-input formation. An RNN can overcome the problem of a long-distance dependency. However, it is difficult to train because gradients may explode or vanish over long sequences [10].

One way to address this problem is by employing a variant of the regular RNN, the LSTM [9]. LSTMs have a more complex internal structure with cells replacing RNN nodes, which allows LSTMs to remember information for either a long or short time. A regular LSTM tends to ignore future contextual information while processing sequences.

The Bidirectional LSTM (BiLSTM) is able to use both past and future contexts by processing the text from both directions [24].

Employing an attention mechanism between sequences (BiLSTM+Att), BiLSTM shows a considerable improvement by changing the contribution of each word to the analysis of the whole text [23], [22]. Before the RNN model summarizes the hidden states for the output, an attention mechanism amplifies the results by aggregating the hidden states (See equations 10 and 11) and weighting their relative importance (See equation 12), where W_h and b_h are the weight and bias from the attention layer.

$$e_i = \tanh(W_h h_i + b_h), e_i \in [-1, 1] \quad (10)$$

$$w_i = \frac{\exp(e_i)}{\sum_{t=1}^N \exp(e_t)}, \sum_{i=1}^N w_i = 1 \quad (11)$$

$$r = \sum_{i=1}^N w_i h_i, r \in R^{2L} \quad (12)$$

Not all words make the same contribution to the business vs. personal categorization of the text. The attention mechanism is able to shuffle the word annotation weights according to their importance to the meaning of a sentence.

4. Experimental Framework

We compare the accuracy of the two traditional (SGD and ERT) and two deep learning (BiLSTM, BiLSTM+Att) algorithms on different vector representations of email content. The best parameters for both classical learners were selected by grid search algorithm as it is presented in Table 5.

Selected modified huber loss is equivalent to quadratically smoothed SVM with $\gamma = 2$ [28]. In the following part of the text, SGD-SVM will denote modified huber loss. For deep learning models, parameters were selected manually using extensive experimentation. For all models, we use the binary cross-entropy loss function and the same optimizer that BERT was originally trained with: the ‘Adaptive Moments’ (Adam). This optimizer minimizes the prediction loss and does regularization by the weight decay, which is also known as AdamW. For the learning rate, we use the same schedule as BERT pre-training: the linear decay of a notional initial learning rate, prefixed with a linear warm-up phase over the first 10% of the training steps known as the number of the warm-up steps. The learning rate is set on $3e-5$, being in line with the BERT paper [4], which specifies the initial learning rate values for fine-tuning. An early stopping strategy is used to prevent over-fitting [20]. All models use gradient descent with mini-batches of

Table 5. Grid-search parameter selection. B: Business, P: Personal. Balanced: class weights are adjusted inversely proportional to class frequencies in the training set

Classifier	Parameter	Search Parameter Space	Best Performing Values
SGD	loss	hinge, log, modified.huber, squared.hinge, perceptron	modified.huber
	penalty	l1, l2, elasticnet	l2
	alpha	0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000	0.0001
	learning_rate	constant, optimal, invscaling, adaptive	optimal
	class_weight	{P: 0.5, B: 0.5}, {P: 0.6, B: 0.4}, {P: 0.7, B: 0.3}, balanced	{P: 0.7, B: 0.3}
	eta0	1, 10, 100	10
ERT	n_estimators	10, 20, 30, 50, 100, 200	10
	criterion	gini, entropy	entropy
	min_samples_leaf	1, 3, 10	3
	class_weight	{P:0.5, B:0.5}, {P:0.6, B:0.4}, {P:0.7, B:0.3}, balanced	{P:0.7, B:0.3}

size 64, ReLU activation function, dimension of word embedding equal to 256, maximum sequence length equal to 256 (E, ED) and 512 (EQ, EQD, B), the number of LSTM equal to maximum sequence length, dropout ratio of 0.1 for LSTM, and 0.4 for Dense layers for all models. Models are trained on 30 epochs.

Models are trained on the training set and evaluate the prediction with the best scores retrieved on the validation and test sets. For traditional models (SGD-SVM, ERT), we use cross validation with 5 folds. For deep learning models, the split ratio for training, validation, and test sets is 50:25:25. In order to illustrate the good performance of our approach, we compare the results with baseline models built on the most recent email (E) and the whole thread arc from the body field of the data set (B).

For evaluating the performance of the techniques, we use the typical evaluation metrics that come from information retrieval - precision, recall and F1 measure, accuracy and balanced accuracy. We aim to improve both the general and balanced accuracy of the classification model as well as F1 measure on minority Personal class.

5. Experimental Results

The results of the model comparison of BoW, Tf-Idf and BERT word embedding with and without Meta features included for SGD-SVM and ERT classifiers in ED experiment are presented in Table 6. Our results show that using Tf-Idf weights for unigrams, $n = 1$ (Tf-Idf-Unigram), unigrams and bigrams, $n \in [1, 2]$ (Tf-Idf-Ngram) and ngram characters of length 1-4 (Tf-Idf-Ngram-Char) as features significantly improves model performances compared to BoW weights on unigrams, $n = 1$ used as features. Moreover, Tf-Idf-Ngram weights generally give the best performance across the experiments and measures.

Traditional learners on all Tf-Idf weights have comparable metric values with deep learning learners and even overcome them at the learner general accuracy, while the later give better balanced accuracy and F1 score on minority Personal class. ERT classifier presents lower values across the measures compared with SGD-SVM classifier. From the results shown, we can also observe that the BiLSTM+Att obtains higher scores than the BiLSTM without the attention mechanism. Moreover, all models with additional Meta features are showing better results improving it by at least 0.1% across the experiments.

Table 6. Comparison between traditional (SGD-SVM, ERT) and deep learning algorithms (BiLSTM, BiLSTM+Att) for different email content representations with and without additional email features included for emails content with domains experiment (ED)

Algorithm	Features	Accuracy	Business			Personal			
			Balanced Accuracy	Precision	Recall	F1	Precision	Recall	F1
ERT	BoW	90.1%	68.9%	91.4%	97.8%	94.5%	73.9%	39.9%	51.9%
	BoW + Meta	90.7%	70.5%	91.8%	98.0%	94.8%	76.9%	42.9%	55.1%
	Tf-Idf-Unigram	90.2%	69.5%	91.5%	97.8%	94.6%	74.1%	41.1%	52.9%
	Tf-Idf-Unigram + Meta	90.6%	70.5%	91.8%	97.8%	94.7%	75.4%	43.2%	55.0%
	Tf-Idf-Ngram	90.1%	67.2%	90.9%	98.4%	94.5%	77.4%	36.0%	49.2%
	Tf-Idf-Ngram + Meta	90.3%	68.9%	91.4%	98.1%	94.6%	76.3%	39.6%	52.2%
	Tf-Idf-Ngram-Chr	89.8%	66.2%	90.6%	98.4%	94.4%	76.9%	33.9%	47.1%
	Tf-Idf-Ngram-Chr + Meta	90.1%	67.6%	91.0%	98.3%	94.5%	76.9%	36.9%	49.9%
SGD-SVM	BoW	90.0%	79.9%	94.7%	93.7%	94.2%	62.0%	66.0%	64.0%
	BoW + Meta	90.2%	80.2%	94.7%	94.0%	94.3%	63.0%	66.4%	64.6%
	Tf-Idf-Unigram	92.0%	82.6%	95.3%	95.5%	95.4%	70.7%	69.8%	70.2%
	Tf-Idf-Unigram + Meta	92.5%	82.1%	95.1%	96.4%	95.7%	74.4%	67.9%	71.0%
	Tf-Idf-Ngram	92.8%	83.2%	95.4%	96.4%	95.9%	75.0%	70.1%	72.5%
	Tf-Idf-Ngram + Meta	92.9%	83.8%	95.6%	96.3%	95.9%	74.8%	71.3%	73.0%
	Tf-Idf-Ngram-Chr	92.3%	82.3%	95.2%	96.0%	95.6%	72.6%	68.5%	70.5%
	Tf-Idf-Ngram-Chr + Meta	92.2%	83.0%	95.4%	95.6%	95.5%	71.1%	70.4%	70.7%
BiLSTM	BERT-Embd	91.4%	81.1%	94.5%	95.6%	95.0%	71.5%	66.7%	69.0%
	BERT-Embd + Meta	91.5%	81.8%	95.3%	94.9%	95.1%	67.0%	68.6%	67.8%
BiLSTM+Att	BERT-Embd	92.1%	83.4%	95.9%	95.1%	95.5%	67.6%	71.7%	69.6%
	BERT-Embd + Meta	92.3%	83.4%	95.7%	95.4%	95.5%	70.3%	71.3%	70.8%

The best models from traditional and deep learning streams, SGD-SVM and BiLSTM+Att on Tf-Idf-Ngram + Meta and BERT-Embd +Meta vector spaces from the previous results have been selected for comparison of the email content experiments. The ED and EQD experiments were able to capture additional knowledge of each email content, so that the whole system slightly improved accuracy compared with the E and EQ experiments respectively, but for such high accuracy values, any improvement becomes significant. The EQD experiment made full use of the associated data available in each email (quotes and recipient email domains) with retrieved improvement in Accuracy score for 3.2% and for 0.6% in SGD-SVM classifier compared with the baseline E and B experiments respectively, as it is presented in Table 7.

Testing approach generalization has been performed using the models built on $Enron_{C_p}$ and $Enron_{B_p}$ data sets independently. For the model trained on $Enron_{C_p}$, the whole $Enron_{B_p}$ data set has been used for testing. In the $Enron_{B_p}$ based model, a data set is firstly split on training, validation, and test data sets. The results from this test, on all different text representations on SGD-SVM and BiLSTM classifiers, confirm that a model can capture important information and transfer the knowledge to differently annotated data sets (see Table 8). Even more, the ED experiment better generalizes the learning process than the B experiment, with a lower difference on all measures between the models. We observe a slight decrease in the test results on $Enron_{B_p}$ since our model parameters are optimized on the $Enron_{C_p}$ data set. Moreover, the size of $Enron_{B_p}$ is much smaller, it is not intentionally annotated for business/personal categorization and it

Table 7. Comparison between different email content representations (Experiments - E, ED, EQ, EQD, B) with additional email features included. SGD-SVM and Bi-LSTM+Attention algorithms are used for models building and testing

Experiment	Algorithm	Features	Business				Personal			
			Accuracy	Balanced Accuracy	Precision	Recall	F1	Precision	Recall	F1
E	SGD-SVM	Tf-Idf-Ngram + Meta	92.6%	81.7%	94.9%	96.7%	95.8%	75.6%	66.7%	70.9%
	BiLSTM+Att	BERT-Embd + Meta	92.3%	86.0%	97.5%	93.8%	95.6%	58.3%	78.2%	66.8%
ED	SGD-SVM	Tf-Idf-Ngram + Meta	92.9%	83.8%	95.6%	96.3%	95.9%	74.8%	71.3%	73.0%
	BiLSTM+Att	BERT-Embd + Meta	92.3%	83.4%	95.7%	95.4%	95.5%	70.3%	71.3%	70.8%
EQ	SGD-SVM	Tf-Idf-Ngram + Meta	95.7%	90.3%	97.4%	97.7%	97.5%	85.0%	82.9%	83.9%
	BiLSTM+Att	BERT-Embd + Meta	94.1%	87.2%	96.6%	96.6%	96.6%	77.8%	77.8%	77.8%
EQD	SGD-SVM	Tf-Idf-Ngram + Meta	95.8%	91.3%	97.7%	97.5%	97.6%	84.0%	85.0%	84.5%
	BiLSTM+Att	BERT-Embd + Meta	94.0%	87.7%	97.0%	96.0%	96.5%	73.9%	79.4%	76.5%
B	SGD-SVM	Tf-Idf-Ngram + Meta	95.2%	90.2%	97.4%	97.1%	97.3%	81.9%	83.2%	82.5%
	BiLSTM+Att	BERT-Embd + Meta	93.9%	86.0%	95.5%	97.4%	96.4%	83.8%	74.5%	78.9%

contains initial categories such as 'personal in professional context' included in the final Personal class email categorization.

6. Comparison with Other SOA Methods

To the best of our knowledge, there have been three attempts in research papers published so far to classify emails in Business and Personal categories. All of them have used their own annotated emails of the Enron corpus with different classification strategies and compared obtained results with other available annotated email data sets (Enron, Avocado). Since our work is based only on the Enron data set, we will compare the retrieved results with the same and differently annotated Enron data sets. The results presented in the paper [12] are based on the Enron data set annotated by the authors, usually denoted as the Sheffield Enron data set in the research papers. It is not obvious, as it is also noted by [3], which training/test ratio was used for obtaining these results. Moreover, the structure of the email content used for email annotation and classification is not known to us. For that reason, we can only treat results from [12] as general points for our classification results comparison. The results obtained after the application of the models from our approach outperform the reported results in the overall Accuracy, Recall, and F1 score on minority (Personal) class in the EQ, EQD and B experiments.

On the other hand, the annotated data set presented in [2] was used in our work. When compared, the results obtained in our baseline experiment E outperform the results reported in the papers [2] and [3] in the overall Accuracy score (+1.4/+1.6%). Macro F1 score on minority Personal class in the E and ED experiments (+0.4% and +2.5% respectively) is better than the one presented in [3]. By comparing other measures from the classification report, they outperform results reported in both of these papers in overall Accuracy score (+4.6%), Recall (+4.9%) and Macro F1 (+2.9%) on Business and Macro F1 (+6.4%) on Personal class across the EQ, EQD and B experiments (see Table 9). Although it is not noted if the authors treated only the most recent email or the whole

Table 8. Results of testing the models for emails content with domains (ED) and body (B) experiments on Berkeley data set - $Enron_{BP}$

Train Data	Algorithm	Features	Exp=ED				Exp=B					
			Accuracy	Balanced F1 Accuracy	Macro F1	Weighted Accuracy	Accuracy	Balanced F1 Accuracy	Macro F1	Weighted Accuracy		
$Enron_C$	SGD-SVM	BoW	83.7%	60.0%	60.3%	83.5%	87.1%	54.4%	55.2%	83.8%		
		BoW + Meta	85.9%	57.0%	58.3%	84.1%	87.8%	56.5%	58.3%	84.9%		
		Tf-Idf-Unigram	87.9%	59.2%	61.6%	85.8%	88.5%	57.8%	60.2%	85.8%		
		Tf-Idf-Unigram + Meta	87.3%	58.1%	60.0%	85.2%	88.5%	58.8%	61.4%	86.0%		
		Tf-Idf-Ngram	88.0%	58.0%	60.2%	85.5%	88.4%	54.7%	55.7%	84.7%		
		Tf-Idf-Ngram + Meta	89.0%	58.0%	60.6%	86.1%	88.4%	56.4%	58.3%	85.3%		
		Tf-Idf-Ngram-Chr	87.2%	59.3%	61.3%	85.4%	88.1%	54.5%	55.4%	84.5%		
		Tf-Idf-Ngram-Chr + Meta	88.3%	59.1%	61.7%	86.0%	87.7%	54.1%	54.7%	84.1%		
		BiLSTM	BERT-Embd	87.8%	69.2%	61.3%	89.1%	87.3%	65.6%	57.1%	90.1%	
			BERT-Embd + Meta	87.9%	69.4%	61.6%	90.1%	87.9%	69.4%	54.6%	91.7%	
		BiLSTM+Att	BERT-Embd	88.7%	73.2%	64.1%	90.6%	87.5%	66.6%	56.5%	90.7%	
			BERT-Embd + Meta	89.2%	78.4%	62.3%	91.9%	88.1%	70.9%	58.8%	91.0%	
		$Enron_{BP}$	SGD-SVM	BoW	86.5%	70.9%	67.4%	87.5%	87.5%	67.1%	69.1%	86.6%
				BoW + Meta	89.0%	69.7%	69.2%	89.1%	87.7%	69.5%	71.0%	87.2%
Tf-Idf	88.7%			61.8%	63.3%	88.0%	90.6%	70.6%	71.7%	90.4%		
Tf-Idf + Meta	89.8%			63.7%	65.9%	88.9%	90.9%	69.4%	71.3%	90.5%		
Tf-Idf-Ngram	90.4%			58.9%	61.7%	88.4%	90.4%	65.3%	67.7%	89.5%		
Tf-Idf-Ngram + Meta	90.4%			60.2%	63.1%	88.7%	90.9%	68.2%	70.5%	90.3%		
Tf-Idf-Ngram-Chr	89.3%			59.6%	61.6%	87.9%	90.9%	73.3%	73.6%	90.9%		
Tf-Idf-Ngram-Chr + Meta	90.1%			61.3%	64.0%	88.7%	92.0%	67.5%	71.5%	91.0%		
BiLSTM	BERT-Embd			89.7%	55.4%	50.4%	93.0%	89.4%	72.5%	55.6%	92.3%	
	BERT-Embd + Meta			90.7%	71.1%	61.0%	92.7%	89.7%	69.5%	52.4%	92.6%	
BiLSTM+Att	BERT-Embd			90.3%	45.3%	47.5%	94.6%	89.7%	74.5%	52.5%	93.0%	
	BERT-Embd + Meta			90.7%	70.8%	56.6%	93.5%	90.4%	88.4%	56.9%	93.9%	

thread arc stored in the Body field of $Enron_C$ data set as individual email, the latest observation has confirmed the strength of our approach in both of these cases.

7. Conclusion and Future Work

The importance and usage of emails by both personal and business users are continuously growing despite the prevalence of alternative means, such as instant mobile and social network messaging. Therefore, email management is an important and growing problem for individuals and organizations. In this paper, we have explored the classification of emails into two main categories, business and personal.

During our work, a comprehensive set of experiments was conducted to find the best solution or this task. We used different traditional and deep learning ML techniques including SGD-SVM, ERT, BiLSTM, and BiLSTM+Att together with different text representation techniques such as BoW and Tf-Idf, word and character n-grams, as well as BERT embeddings. The experimental results showed that traditional ML techniques with Tf-Idf text representation techniques slightly outperformed deep learning approach on this task. The reason for that may be the limitations in the research computational environment we used. Additionally, we put a lot of effort into introducing and experimenting with various additional features. To achieve the best possible generalization of the model, we excluded from the email all specificity of the training data set and focused only on the part of the email that contains the conversation itself.

Based on this work, we plan to expand our research in several different directions. First, data sets that differ in many aspects should be incorporated, including email data

Table 9. Comparison of results with other SOA methods. Accuracy, F1, Precision, and Recall measures were taken from the best experiments on test sets reported in the papers

Paper		Business				Personal		
		Accuracy	Precision	Recall	F1	Precision	Recall	F1
[12]		93.0%	92.0%	99.0%	95.0%	95.0%	69.0%	80.0%
[2]		91.2%	96.7%	92.1%	94.4%	73.5%	87.5%	79.9%
[3]		91.0%	96.6%	92.9%	94.7%	63.4%	79.3%	70.5%
Our Approach	E	92.6%	94.9%	96.7%	95.8%	75.6%	66.7%	70.9%
	ED	92.9%	95.6%	96.3%	95.9%	74.8%	71.3%	73.0%
	EQ	95.7%	97.4%	97.7%	97.5%	85.0%	82.9%	83.9%
	EQD	95.8%	97.7%	96.1%	97.6%	84.0%	85.0%	84.5%
	B	95.2%	97.4%	97.1%	97.3%	81.9%	83.2%	82.5%

sets in languages other than English and other conversational data sets, such as short messages. Further, different weighting schemes for additional features used in our research (such as NER tokens) should be investigated. Some of the lexicons, such as acronyms, business words and personal names, should be further analysed and improved. By using pre-trained BERT models based on the conversational data sets from business environments, as well as the sentence instead of the word embedding space for text representation could give significant value. Also, we plan to extend the research on the prediction of the hierarchical organizational structure by analyzing only business emails exchanged through the organization. One of our goals will be to examine extraction of the signatures from the business emails, as well as entities from their signatures. Our approach raises questions about the significance of different ways of expression in email communication, and how they can be used to better understand human behavior in a business environment. By understanding more deeply the emotional and moral framework of correspondents, organizers can better anticipate their response to certain requests and predict the outcome of the planned activities.

Acknowledgments. The work presented has been supported by the Ministry of Science and Technological Development, Republic of Serbia, through Projects No. 174021 and No. III47003.

References

1. Alhogaïl, A., Alsabih, A.: Applying machine learning and natural language processing to detect phishing email. *Computers & Security* 110, 102414 (2021)
2. Alkhereyf, S., Rambow, O.: Work hard, play hard: Email classification on the avocado and enron corpora. In: *Proceedings of TextGraphs-11: the Workshop on Graph-based Methods for Natural Language Processing*. pp. 57–65 (2017)
3. Alkhereyf, S., Rambow, O.: Email classification incorporating social networks and thread structure. In: *Proceedings of The 12th Language Resources and Evaluation Conference*. pp. 1336–1345 (2020)
4. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018)
5. Geurts, P., Ernst, D., Wehenkel, L.: Extremely randomized trees. *Machine learning* 63(1), 3–42 (2006)

6. Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S.P., Ditto, P.H.: Moral foundations theory: The pragmatic validity of moral pluralism. In: *Advances in experimental social psychology*, vol. 47, pp. 55–130. Elsevier (2013)
7. Graovac, J.: A variant of n-gram based language-independent text categorization. *Intelligent Data Analysis* 18(4), 677–695 (2014)
8. Graovac, J., Kovačević, J., Pavlović-Lažetić, G.: Hierarchical vs. flat n-gram-based text categorization: can we do better? *Computer Science and Information Systems* 14(1), 103–121 (2017)
9. Graves, A.: Long short-term memory. In: *Supervised sequence labelling with recurrent neural networks*, pp. 37–45. Springer (2012)
10. Hochreiter, S.: The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International JOURNAL of Uncertainty, Fuzziness and Knowledge-Based Systems* 6(02), 107–116 (1998)
11. Hopp, F.R., Fisher, J.T., Cornell, D., Huskey, R., Weber, R.: The extended moral foundations dictionary (emfd): Development and applications of a crowd-sourced approach to extracting moral intuitions from text. *Behavior Research Methods* 53(1), 232–246 (2021)
12. Jabbari, S., Allison, B., Guthrie, D., Guthrie, L.: Towards the orwellian nightmare: separation of business and personal emails. In: *Proceedings of the COLING/ACL 2006 Main conference poster sessions*. pp. 407–411 (2006)
13. Kessler, J.S.: Scattertext: a browser-based tool for visualizing how corpora differ. arXiv preprint arXiv:1703.00565 (2017)
14. Klimt, B., Yang, Y.: The enron corpus: A new dataset for email classification research. In: *European Conference on Machine Learning*. pp. 217–226. Springer (2004)
15. Mikolov, T., Yih, W.t., Zweig, G.: Linguistic regularities in continuous space word representations. In: *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: Human language technologies*. pp. 746–751 (2013)
16. Mohammad, S.M.: Word affect intensities. arXiv preprint arXiv:1704.08798 (2017)
17. Nisar, N., Rakesh, N., Chhabra, M.: Review on email spam filtering techniques. *International JOURNAL of Performability Engineering* 17(2) (2021)
18. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al.: Scikit-learn: Machine learning in python. *the JOURNAL of machine Learning research* 12, 2825–2830 (2011)
19. Plutchik, R.: The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist* 89(4), 344–350 (2001)
20. Prechelt, L.: Early stopping-but when? In: *Neural Networks: Tricks of the trade*, pp. 55–69. Springer (1998)
21. Radicati, S.: *Email market, 2021-2025*. The Radicati Group, Inc., Palo Alto, CA (2021)
22. Raffel, C., Ellis, D.P.: Feed-forward networks with attention can solve some long-term memory problems. arXiv preprint arXiv:1512.08756 (2015)
23. Rocktäschel, T., Grefenstette, E., Hermann, K.M., Kočiskỳ, T., Blunsom, P.: Reasoning about entailment with neural attention. arXiv preprint arXiv:1509.06664 (2015)
24. Schuster, M., Paliwal, K.K.: Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing* 45(11), 2673–2681 (1997)
25. Sharaff, A., Nagwani, N.K.: Identifying categorical terms based on latent dirichlet allocation for email categorization. In: *Emerging Technologies in Data Mining and Information Security*, pp. 431–437. Springer (2019)
26. Shroff, N., Sinhgala, A.: Email classification techniques—a review. *Data Science and Intelligent Applications* pp. 181–189 (2021)
27. Turc, I., Chang, M.W., Lee, K., Toutanova, K.: Well-read students learn better: On the importance of pre-training compact models. arXiv preprint arXiv:1908.08962 (2019)

28. Zhang, T.: Solving large scale linear prediction problems using stochastic gradient descent algorithms. In: Proceedings of the twenty-first international conference on Machine learning. p. 116 (2004)
29. Zhang, Y., Wallace, B.: A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. arXiv preprint arXiv:1510.03820 (2015)

Milena Šošić is a third-year PhD student under the supervision of dr Jelena Graovac at the University of Belgrade, Faculty of Mathematics. Her doctoral work explores the significance of different machine learning techniques in natural language processing and understanding fields with a focus on conversational text analysis. She holds a magister (mr) and graduation degrees in Computer Science, both from University of Belgrade, Faculty of Mathematics. She can be contacted at: pd202030@alas.matf.bg.ac.rs.

Jelena Graovac is Assistant Professor in the Department of Computer Science, Faculty of Mathematics, University of Belgrade. She received M.Sc. (2008, Computer Science) and Ph.D. (2014, Computer Science) degrees from the Faculty of Mathematics, University of Belgrade. The courses she taught at the University of Belgrade include Database Design, Information Systems, Introduction to Computer Organization and Architecture, Web Programming, Introduction to Programming, etc. (Faculty of Mathematics), and Intelligent Search (Intelligent Systems - the Ph.D. program of academic studies). Her research interests include Natural Language Processing, Information Retrieval, and Text Classification using Machine Learning and Knowledge-Based approaches. She co-authored many scientific papers as book chapters and articles in journals and conference proceedings.

Received: February 12, 2022; Accepted: Jun 01, 2022.

Re-evaluation of the CNN-based State-of-the-art Crowd-counting Methods with Enhancements

Matija Teršek*, Maša Kljun*, Peter Peer, and Žiga Emeršič

Faculty of computer and information science
Večna pot 113, SI-1000 Ljubljana
{matija.tersek, masa.kljun}@student.uni-lj.si,
{peter.peer, ziga.emersic}@fri.uni-lj.si

Abstract. Crowd counting has a range of applications and it is an important task that can help with the accident prevention such as crowd crushes and stampedes in political protests, concerts, sports, and other social events. Many crowd counting approaches have been proposed in the recent years. In this paper we compare five deep-learning-based approaches to crowd counting, reevaluate them and present a novel CSRNet-based approach. We base our implementation on five convolutional neural network (CNN) architectures: CSRNet, Bayesian Crowd Counting, DM-Count, SFA-Net, and SGA-Net and present a novel approach by upgrading CSRNet with application of a Bayesian crowd counting loss function and pixel modeling. The models are trained and evaluated on three widely used crowd image datasets, ShanghaiTech part A, part B, and UCF-QNRF. The results show that models based on SFA-Net and DM-Count outperform state-of-the-art when trained and evaluated on the similar data, and the proposed extended model outperforms the base model with the same backbone when trained and evaluated on the significantly different data, suggesting improved robustness levels.

Keywords: Crowd counting, convolutional neural networks, deep learning.

1. Introduction

Automatic estimation of a number of people in a crowd as illustrated in Figure 1 is an important technique with applications in many fields. Political protests, rallies, concerts, religious events, etc., are just some of the situations that can benefit from the automatic crowd counting, since having a good estimate of the crowd can help prevent crowd crushes, stampedes, and other accidents. Furthermore, in the light of the recent pandemic of the COVID-19, crowd counting and crowd analysis can help prevent the spread of the virus by ensuring enough physical distance between people in some usually crowded public places, such as stores, cinemas, recreational areas, etc.

In addition to the mentioned applications, crowd counting is popular as it can be easily extended to a task of counting objects in other fields. Some of them include counting vehicles for traffic control [37, 49], monitoring discarded fish catch and counting animals for environmental control [2, 13, 52, 50], counting leaves for plant phenotyping [1], estimating the number of cells in microscopic images [27] or more generally detecting moving objects [6, 17, 23]. Counting of objects is crucial in such tasks as it automates and speeds up otherwise tedious processes.

* Both authors contributed equally



Fig. 1. Figure shows 4 randomly chosen images from the ShanghaiTech part A train set. We can see that the images are cropped to contain dense crowds only.

Because of the wide variety of applications of crowd counting methods, a lot of research has been made and many different algorithms have been proposed. Different approaches to crowd counting exist, and they can be roughly divided into 3 groups—detection, regression, and density based. While some related works include overviews of existing crowd analysis methods [16, 32, 44, 48, 65], the other focus more on discovering the new approaches [29, 33, 57, 59, 71].

In this paper we focus on CNN-based approaches, as they recently began to gain in the popularity. We briefly describe and provide key features of five state-of-the-art models.

Unlike some of the related works, which only gather the results from authors' papers, we try to train and evaluate the models ourselves on three popular crowd counting datasets. Furthermore, we propose an improvement for one of the models and compare it to the others. Source code and pretrained weights are available at our GitHub¹. To summarize our key contributions:

- Direct comparison of some the most popular state-of-the-art models and their re-implementation. To the best of our knowledge, no comparison to this extent has not been made in literature yet.
- Presentation and comparison of our own model. The model's architecture is based on CSRNet with dilated convolutions, with added pixel modeling and enhanced Bayesian loss function.
- We make our implementations freely available for other researchers to use and modify.

This paper is organized as follows: In Chapter 2 we provide the most common approaches to crowd counting. In Chapter 3 we describe five state-of-the-art CNN models and our suggested improvement. In Chapter 4 we describe the three datasets on which we evaluate the models and discuss the results of our evaluation.

2. Crowd counting approaches

The goal of crowd (of people) counting methods is to determine the number of people present in a particular area. There exist many different approaches of doing this and we can divide the traditional approached into 3 main categories - detection, regression, and density based approaches [16, 48]. CNNs dominate the more recent approaches, which can be categorized into its own group. We would like to emphasize that despite the fact that the term crowd could be used for any type of crowd of objects, all the mentions of a crowd refer to the crowd of people.

For the most comprehensive overview of the area we refer the reader to some of the surveys on crowd counting approaches, such as [11]. Here we only skim through the more popular approaches from the recent years.

Detection based approaches: This is the most straight-forward approach that can use whole bodies (Monolithic detection [9, 10, 15, 26, 43, 53–55, 63]) or just parts of it (Part-based detection [12, 28, 30, 62]), e.g., the combination of head and hands. Approaches in the first group use features such as Haar wavelets or histogram of oriented gradient (HOG) to represent the body, and then use a classifier with the sliding window approach across the image to detect person candidates. Models can be then learned using support vector machines, boosting, random forests, etc.

In the recent years many object detectors based on CNNs were also presented. YOLO network [40] applies a single neural network to the full image, divides it into region, and predicts bounding boxes and probabilities for each region. Other CNN approaches include Fast R-CNN [19] and Faster R-CNN [41].

¹ <https://github.com/tersekmatija/crowd-counting-cnns>

Another approach uses shape learning, where humans are modeled with 3d shapes composed of ellipsoids. A stochastic process is then employed to estimate the number and shape configuration which best explains a given foreground mask in a scene [18, 69]. The drawback of detection based approaches is that they fail in high occlusion situations or in highly crowded spaces [32].

Regression based approaches: The idea of this group’s methods is not to count individuals, but to estimate the crowd density, which is specifically useful in more crowded places [4, 5, 7–9, 20, 21, 31, 34, 36, 38, 42, 51]. Methods in this group first encodes low-level information with the help of foreground, edge, texture, and/or gradient features. Then, with the help of a regression model, a mapping between low-level features and people count is made. Different regressions, such as linear regression, ridge regression, neural network, etc., can be used. The drawback of regression based approaches is that when the same object is placed in different depths in the image, the values of features extracted from those objects can vary upon the depth of where the object was placed. However, this problem can be tackled by geometric correction [32].

Density based approaches: The idea of this group’s methods, such as [39, 61, 64, 67] in its most simplistic form is to obtain a density map from an image and then integrate it in order to get the estimation of people in the image. Contrary to the previous approaches, these also consider the spatial information. The pioneering work include [27], who suggest learning a linear mapping between local patch features and corresponding object density maps. The methods differ in the choice of a training loss function (e.g., squared error between the predicted density values and the ground truth) and in the choice of a density map prediction method (e.g., with the help of a linear model) [24].

CNNs: In 2015 the pioneering work with deep networks in crowd counting was introduced in [58], introducing CNN approaches to the crowd counting. Since then many of CNN based approaches were proposed. The basic idea behind CNN based approaches is that they normally try to predict the density map from the input image and infer the count from it. This also means they are the most similar to the traditional density based approaches. Models that are based on CNNs differ in the usage of different backbones (e.g., VGG-16, VGG-19, Inception v3), loss functions, additional maps (e.g., attention map), and model structure (e.g., single or multi column).

In recent surveys [48, 11] authors classify CNN-based approaches into four categories, based on the property of the networks: *Basic CNNs* include networks with basic CNN layers and represent initial deep learning approaches for crowd counting [14, 35, 56, 58, 67], *scale-aware models* that leverage multi-column or multi-resolution architectures to achieve scale robustness [3, 25, 37, 68], *context-aware models* that incorporate global and local contextual information to improve performance [45, 46], and *multi-task frameworks* that combine crowd counting with tasks such as crowd velocity estimation, etc. [2, 47, 66, 70] Based on the inference methodology, they also classify them into *patch-based*, where models are trained using patches from the image and the inference is done using sliding window approach [2, 3, 14, 25, 35, 37, 56, 58, 66, 70], and whole image-based [45–47, 68, 60].

We describe five CNN models for crowd counting along with their key features and one improved model in the next chapter.

3. CNN models

In this section we shortly describe each of the models. Note that we put the main focus on their features, where they differ the most from each other. The models were chosen as each of them made a significant contribution in the field, as well as based on their popularity in literature, at the time of writing.

3.1. CSRNet

The architecture of this model is divided into 2 parts: a CNN at the front-end and a dilated CNN at the back-end. The basis of CSRNet front-end is build on VGG-16 model with the fully-connected layers removed [29]. Ten layers of VGG-16 are kept, with only three pooling layers instead of five. The back-end consists of six dilated convolutional layers, for which the authors suggest that it represents a good alternative to the pooling layers. Dilated convolution can be used instead of the pooling layer, since it maintains the resolution of feature map and contains more detailed information. Another 1×1 convolutional layer is added as the output layer.

Authors suggest different models, which are determined by different back-end settings that vary in the dilation rate. We use model B in our experiments, as it is the most successful [29], where the dilation rate is set to 2 for all the back-end layers.

Dilated convolution. The idea of the dilated convolution is that it uses sparse kernels, which enlarge the receptive field. The same can be achieved by adding more convolutional layers, however, that increases the computational cost.

For input $x(m, n)$ and filter $w(i, j)$, of length M , width N , and the dilation rate r ($r = 1$ results in a normal convolution), we can define output $y(m, n)$ of the dilated convolution as

$$\sum_{i=1}^M \sum_{j=1}^N x(m + r \times i, n + r \times j) w(i, j). \quad (1)$$

Loss function and training. Loss function is derived from the Euclidean distance between the ground truth and estimated density map. The loss function is defined as

$$\mathcal{L} = \frac{1}{2N} \sum_i^N \|D_i^{est} - D_i^{gt}\|_2^2, \quad (2)$$

where N is the size of the training batch, D_i^{est} the density map generated by the CSRNet, and D_i^{gt} the ground truth density map of the input image.

In training the first 10 convolutional layers are fine-tuned from a trained VGG-16. Initial settings for other layers are set with the help of a Gaussian distribution with 0.01 standard deviation, and stochastic gradient descent (SGD) with rate $1e - 6$ is applied during training.

3.2. Bayesian crowd counting

The Bayesian model uses VGG-19 as the backbone, with the last pooling and the subsequent fully connected layers removed. The output of the backbone is upsampled to $\frac{1}{8}$ of the input image size by bilinear interpolation and fed to a regression header. The regression header consists of two 3×3 convolutional layers, one with 256 and the other with 128 channels, and one 1×1 convolutional layer. The produced output is a density map [33].

Bayesian crowd counting model differs from other CNN based models in the utilization of a loss function. Opposed to the previous models, which use a Gaussian kernel to obtain the ground truth density map and define loss function as a sum of pointwise distances between ground truth and estimated density maps, it uses a novel Bayesian loss function.

Bayesian Loss function and training. We can derive the loss function as follows. Let x be a random variable describing the spatial location, and y be a random variable representing the annotated head point. Let $m = 1, \dots, M$ where M is the number of pixels in the density map and let $n = 1, \dots, N$, where N is the total crowd count. Let z_n be a head position and y_n be a corresponding label. The likelihood function of location x_m given the label y_n can be defined as

$$p(x_m|y_n) = N(x_m; z_n, \sigma^2 \mathbf{1}_{2 \times 2}), \quad (3)$$

where $N(x_m; z_n, \sigma^2 \mathbf{1}_{2 \times 2})$ is a 2D Gaussian distribution evaluated at x_m , with the mean at the annotated point z_n and an isotropic covariance matrix $\sigma^2 \mathbf{1}_{2 \times 2}$.

Using Bayes we can then compute

$$p(y_n|x_m) = \frac{p(x_m|y_n)p(y_n)}{p(x_m)} = \frac{N(x_m; z_n, \sigma^2 \mathbf{1}_{2 \times 2})}{\sum_{n=1}^N N(x_m; z_n, \sigma^2 \mathbf{1}_{2 \times 2})}. \quad (4)$$

The Bayesian loss function can be defined as

$$\mathcal{L}^{\text{Bayes}} = \sum_{n=1}^N \mathcal{F}(1 - E[c_n]), \quad (5)$$

where \mathcal{F} is a distance function (ℓ_1) and $E[c_n]$ is the expected value of a total count associated with y_n , that can be computed as

$$E[c_n] = \sum_{m=1}^M p(y_n|x_m) D^{\text{est}}(x_m). \quad (6)$$

When inferring, the total count is just a sum over an estimated density map.

Additionally, authors introduce the background pixel modeling for background pixels that are far away from any of the annotation points. They introduce an additional background label $y_0 = 0$ in addition to the head labels, as it makes no sense to assign the background pixels to any of the head labels. The posterior label probability is then rewritten and additional expected count for the entire background $E[c_0]$ is introduced. Pixel modeling defines a new, enhanced loss function, as also described in Equation 7.

MSRA initializer is used for the initialization of the regression header, whereas the backbone is pre-trained on ImageNet. Parameters are updated with the help of the Adam optimizer with an initial learning rate $1e - 5$.

3.3. Our proposed model

Concepts such as dilated convolution, use of Bayesian loss instead of Gaussian kernel and using pixel modeling to suppress background pixels have been shown to improve crowd counting performance [29]. We infer that combining all of these key concepts should result in a more robust and better performing crowd counting model. Therefore we base our proposed model on the CSRNet and Bayesian crowd counting loss function and pixel modeling [33, 33], with the goal of improving performance.

The basic structure of our model is the same as the one of the CSRNet, described in Subsection 3.1. We use the first ten layers of the VGG-16 with 3 pooling layers for the front-end, 6 convolutional layers with the dilation rate set to 2 as the back-end, and an additional 1×1 layer as the output layer.

Loss function and training. Instead of CSRNet’s loss function provided in Equation 2, we use the enhanced loss function defined as:

$$\mathcal{L}^{Bayes+} = \sum_{n=1}^N \mathcal{F}(1 - E[c_n]) + \mathcal{F}(0 - E[c_0]). \quad (7)$$

The weights are initialized in the same way as in the CSRNet. The first 10 convolutional layers are fine-tuned from a trained VGG-16, whereas initial settings for other layers are obtained with the help of a Gaussian distribution with 0.01 standard deviation. Parameters are updated with the help of the Adam optimizer with an initial learning rate of $1e - 6$.

3.4. SFANet

The next model we analyse is SFANet [71]. It uses the first 13 layers of a pre-trained VGG-16-bn (VGG-16 with batch normalization) as the front-end feature map extractor. It is suitable as it has a strong ability to represent features and can be easily concatenated by the back-end dual path networks. Four source layers (conv2-2, conv3-3, conv4-3, and conv5-3) are then connected to a dual multi-scale fusion networks with attention (density map path and attention map path), which represent the back-end. Attention map path is incorporated to tackle the background noise and non-uniformity of crowd distributions.

Loss function and training. In most models an Euclidean loss is used for measuring estimation error, which is defined as:

$$\mathcal{L}^{\text{DEN}} = \frac{1}{N} \sum_{i=1}^N \|D_i^{\text{est}} - D_i^{\text{gt}}\|^2, \quad (8)$$

where D_i^{est} is the estimated density map of i -th input image, D_i^{gt} represents the ground truth density map, and N is the batch size. SFANet also uses the described loss function. In addition, the model uses the attention map loss function, a binary class entropy defined as

$$\mathcal{L}^{ATT} = -\frac{1}{N} \sum_{i=1}^N (A_i^{gt} \log(P_i) + (1 - A_i^{gt}) \log(1 - P_i)), \quad (9)$$

where A_i^{gt} is the attention map ground truth, and P_i probability of each pixel in predicted attention map activated by sigmoid function.

The unified loss function is then defined as

$$\mathcal{L} = \mathcal{L}^{DEN} + \alpha \mathcal{L}^{ATT}, \quad (10)$$

with α weighting weight set to 0.1.

The first 13 layers of a pre-trained VGG-16-bn are applied as the front-end feature extractor. Other parameters are randomly initialized with a Gaussian distribution with a standard deviation 0.01. Parameters are updated with the help of Adam optimizer with learning rate of $1e - 4$ and weight decay of $5e - 3$.

Ground truth. Density map ground truth D^{gt} is obtained similarly as in most models, with the use of Gaussian kernels.

Attention map ground truth is obtained from D^{gt} and Gaussian kernel as

$$\begin{aligned} \mathbb{Z} &= D_i^{gt} \times G_{\mu, \sigma^2}(x), \\ A_i^{gt}(x) &= \begin{cases} 0, & x < thresh \\ 1, & x \geq thresh \end{cases}, \forall x \in \mathbb{Z}, \end{aligned} \quad (11)$$

with *thresh* set to 0.001.

3.5. DM-Count

DM-Count model considers crowd counting as a distribution matching problem [57]. The architecture of the model is based on the VGG-19 and is the same as in the Bayesian Crowd Counting model (see Subsection 3.2). Different to the previous models, who use density map estimations that are computed with the help of Gaussian kernels, DM-Count can preprocess ground truth annotations without the use of a Gaussian. Instead it uses Optimal Transport (OT) to measure the similarity between the normalized predicted density map and the normalized ground truth density map. OT computation is then stabilized with the help of a Total Variation (TV) loss.

Loss function and training. The loss function is the combination of the counting loss, optimal transport loss, and the total variation loss. Let $h \in \mathbb{R}_+^n$ be a vectorized binary map for dot annotation, and $\hat{h} \in \mathbb{R}_+^n$ a vectorized predicted density map.

Counting loss.

$$\mathcal{L}^{COUNT}(h, \hat{h}) = |||h||_1 - ||\hat{h}||_1|, \quad (12)$$

where $|| \cdot ||$ denotes the L_1 norm.

Optimal transport loss. Since h and \hat{h} are both unnormalized density functions, they can be turned into the probability density functions (PDF) with dividing them by their respective total mass. Optimal transport loss is then defined as

$$\begin{aligned}\mathcal{L}^{\text{OT}}(h, \hat{h}) &= \mathcal{W}\left(\frac{h}{\|h\|_1}, \frac{\hat{h}}{\|\hat{h}\|_1}\right) \\ &= \left\langle \alpha^*, \frac{h}{\|h\|_1} \right\rangle + \left\langle \beta^*, \frac{\hat{h}}{\|\hat{h}\|_1} \right\rangle,\end{aligned}\quad (13)$$

where \mathcal{W} is a Monge-Kantorovich's Optimal Transport cost (see [57] for the definition), with α^* and β^* being solutions of the optimal transport problem.

Authors suggest the use of OT instead of some other measure of similarity between two PDFs, such as Kullback-Leibler divergence or Jensen-Shannon divergence, as it provides a valid gradient to train a network. The gradient with respect to \hat{h} can be obtained as

$$\frac{\partial \mathcal{L}^{\text{OT}}(h, \hat{h})}{\partial \hat{h}} = \frac{\beta^*}{\|\hat{h}\|_1} - \frac{\langle \beta^*, \hat{h} \rangle}{\|\hat{h}\|_1^2}, \quad (14)$$

which can be back-propagated to learn the parameters of the density estimation network.

Total variation loss. OT loss is optimized with Sinkhorn algorithm for approximating α^* and β^* in each training iteration. Due to this optimization, OT loss approximates well more dense areas, but it performs poorer for the low density areas. To cope with that, Total variation loss is additionally used and can be defined as

$$\mathcal{L}^{\text{TV}}(h, \hat{h}) = \frac{1}{2} \left\| \frac{h}{\|h\|_1} - \frac{\hat{h}}{\|\hat{h}\|_1} \right\|_1. \quad (15)$$

3.6. SGANet

The SGANet model is the first model that investigates Inception-v3 as a backbone network instead of VGG-16, VGG-19, or ResNet, as in the most state-of-the-art models [59]. Fully-connected layers and two maxpooling layers are removed. Before the last Inception Module an upsampling layer is added, which is connected to both, the attention layer and the last Inception Module. Attention layer's output is then applied to the feature maps generated by the last Inception Module.

Loss function and training. In SGANet a novel curriculum loss strategy to address the issues caused by extremely dense regions was used. This is a strategy of model learning where easy examples are selected at the beginning of the training and more difficult ones are added to the training set gradually. A threshold is used for determining the difficulty score, where density map pixels with higher values than the threshold have higher difficulty scores, since such pixels are within the regions of denser crowds. The whole training set is used throughout the training process, however, the threshold is first set to a

low value and then gradually increased, which turns difficult pixels into easy ones so that they contribute more to the training.

The loss function is defined as a sum of two loss functions:

$$\mathcal{L} = \mathcal{L}^{\text{DEN}} + \lambda \mathcal{L}^{\text{SEG}}, \quad (16)$$

where λ is a hyper-parameter set to 20. The density map loss can be calculated as

$$\mathcal{L}^{\text{DEN}} = \frac{1}{2N} \sum_{i=1}^N \|\hat{M}_i^{\text{den}} - M_i^{\text{den}}\|_F^2 \quad (17)$$

and segmentation map loss is defined as the cross-entropy loss as

$$\begin{aligned} \mathcal{L}^{\text{SEG}} = & -\frac{1}{N} \sum_{i=1}^N \|M_i^{\text{seg}} \odot \log(\hat{M}_i^{\text{seg}}) \\ & + (1 - M_i^{\text{seg}}) \odot \log(1 - \hat{M}_i^{\text{seg}})\|_1, \end{aligned} \quad (18)$$

where $\|\cdot\|_1$ denotes the element-wise matrix norm, \odot denotes elementwise multiplication of two same-size matrices, and M^{seg} and M^{den} represent ground truth (without hat) and estimated (with hat) segmentation and density maps.

Ground truth. Ground truth density map M^{den} is obtained using a Gaussian kernel with fixed σ . Segmentation map ground truth is obtained similarly, but as

$$M^{\text{seg}}(x) = \sum_{i=1}^N \delta(x - x_i) * J_n(x), \quad (19)$$

where $J_n(x)$ is an all-one matrix of size $n \times n$ centered at the position x . [59] set $n = 25$.

Model uses the Adam optimizer for updating the parameters, where the initial learning rate is set to $1e - 4$ and reduced by a factor of 0.5 after every 50 epochs. The weights of the Inception layers are loaded from a pre-trained Inception-v3 model.

4. Experiments and Results

Here we describe our experiments, including data used, evaluation protocol, and present results and findings. We relied on implementations provided by the authors. All models were implemented in Pytorch and trained with provided default parameters.

4.1. Data

We test the described models on the three publicly available datasets, described below.



Fig. 2. Figure shows 4 randomly chosen images from the ShanghaiTech part B train set. We can see that the images contain relatively sparse crowds. The background often consists of buildings and vegetation, but can also include rivers as seen in the top left image



Fig. 3. Figure shows 4 randomly chosen images from the UCF-QNRF train set. We can see that the images are more realistic than the images from the ShanghaiTech part A, as they include not only crowds but also buildings, sky, and vegetation

ShanghaiTech Dataset. ShanghaiTech consists of two parts – part A and part B [68]. Part A contains 482 images downloaded from the internet, containing highly congested scenes. It contains a total of 241,667 annotated people, with a 501 average per image, and 3139 maximum. It comes split into a train and a test set, containing 300 and 182 images, respectively. Images in this dataset are challenging to count, as they contain extremely congested scenes, varied perspective, and unfixed resolution. Figure 1 shows some examples of ShanghaiTech part A train set images.

Part B contains 716 images that are taken from the busy streets of metropolitan areas of Shanghai. Images are of fixed size and contain total of 88,488 annotated people, with a 124 average per image, and 578 maximum. Same as the part A, it is already split into a train and a test set, containing 400 and 316 images, respectively. As images are captured in metropolitan areas, they contain relatively sparse crowds and include streets, buildings, vegetation, and sometimes rivers as well. In Figure 2 we show some examples of the ShanghaiTech part B train set images.

UCF-QNRF Dataset. UCF-QNRF is among the newest and the largest datasets for crowd counting problems [22]. It consists of 1525 images and contains a total of 1,251,642 annotated people, with a 815 average, and 12,865 maximum. It is split into a train and a test set, containing 1201 and 334 images, respectively. Dataset contains images with congested scenes with a diverse set of viewpoints, densities, and lighting variations. Different from the ShanghaiTech part A, which contains images with dense crowds that are cropped to contain crowds only, images from this set also contain buildings, vegetation, sky, and roads, as they are present in realistic scenarios captured in the wild, making the

dataset more realistic but also more difficult to count. Figure 3 shows some examples of UCF-QNRF train set images. We summarize the datasets in Table 1.

Table 1. A summary of used datasets. For each we show a number of images, average and maximum people count per image, and a total number of annotated people

Dataset	Images	Avg count	Max count	Annotations
ShanghaiTech A [68]	482	501.4	3, 139	241, 677
ShanghaiTech B [68]	716	123.6	578	88, 488
QNRF [22]	1, 535	815	12, 865	1, 251, 642

4.2. Evaluation metrics

We use Mean Absolute Error (MAE) and Mean Squared Error (MSE) for the evaluation and they are defined as follows

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |C_i - C_i^{GT}|, \quad (20)$$

$$\text{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n |C_i - C_i^{GT}|^2}, \quad (21)$$

where n is a number of images, C_i represents the inferred count, and C_i^{GT} represents the ground truth count.

4.3. Evaluation

We train and test the models on the three mentioned datasets – ShanghaiTech part A, part B, and UCF-QNRF. In addition to training and evaluating models separately for the three datasets, we also include the results of training the model with ShanghaiTech part A train set and evaluating it with UCF-QNRF test set, to inspect how well the models learn to generalize when trained on a similar dataset. We show the obtained MAE and MSE in Table 2 and also provide some qualitative evaluation of the results. Furthermore, in Table 3 we also compare the sizes (in millions of trainable parameters) of the evaluated models, showing that the performance of the proposed model is increased without increasing the training complexity.

Quantitative Analysis The best results in general are obtained on the ShanghaiTech part B (SHB) dataset, which is expected due to the low average count per image and relatively sparse crowds. We see that the best results are obtained by the SFA-Net (7.05 MAE) and SGA-Net (11.48 MSE), followed closely by the DM-Count (7.68 MAE). The worst performance is given by the CSRNet (11.27 MAE), which is outperformed by our

Table 2. In this table we show the evaluation of the models in terms of MAE and MSE on different datasets. The best results are marked in bold. We see that SFA-Net and DM-Count perform the best on ShanghaiTech part A (SHA), with the first giving the best performance on ShanghaiTech part B (SHB), and the latter giving the best performance also on the UCF-QNRF (QNRF). In terms of MSE, SGA-Net outperforms the SFA-Net on the SHB dataset. Bayesian Crowd Counting yields the best results when trained on SHA and evaluated on QNRF. We also show that our combination of Bayesian Crowd Counting model and CSRNet, Bayesian CSRNet, is in fact an improvement of the original CSRNet model. ”/” denotes situations where we could not execute the training due to our hardware limitations. However, in these cases, where possible, we report values from models’ papers – denoted by a *

Datasets	SHA		SHB		QNRF		QNRF on SHA	
	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
CSRNet	75.44	113.55	11.27	19.32	/	/	199.54	319.09
Bayesian CSRNet	69.46	111.73	8.48	13.55	103.94	186.22	139.83	260.59
Bayesian Crowd Counting	66.92	112.07	8.27	13.56	90.43	161.41	138.39	256.81
DM-Count	61.39	98.56	7.68	12.66	88.97	154.11	141.43	260.23
SFA-Net	59.58	99.43	7.05	12.18	<i>100.8*</i>	<i>174.5*</i>	170.29	365.59
SGA-Net	61.58	101.59	7.60	11.48	<i>89.1*</i>	<i>150.6*</i>	/	/

Table 3. A size comparison of models. For each model we show a number of trainable parameters (in table denoted by # of TP) in millions (M)

	CSRNet	Bay. CSRNet	Bay. Crowd Count.	DM-Count	SFA-Net	SGA-Net
# of TP [M]	16.3	16.3	21.5	21.5	17.0	18.1

improved model Bayesian CSRNet (8.48 MAE) and Bayesian Crowd Counting model (8.27 MAE).

The results on ShanghaiTech part A (SHA) are better than those on the UCF-QNRF due to the smaller dataset and smaller and less complicated images. The best results on SHA dataset are obtained by SFA-Net and DM-Count. While the first has a lower MAE (59.58), the second has lower a MSE (98.56). They are closely followed by SGA-Net (61.58 MAE). The worst performance is obtained by CSRNet (75.44 MAE), however, we show that our Bayesian CSRNet model is in fact an improvement of the original CSRNet, with MAE of 66.92.

Due to the bigger size of the images from the QNRF dataset and our hardware limitations, we were not able to train and evaluate all of the models. In cases like these modifying the models in order to be able to retrain them on our limited settings could result in falsely lower results. In order to avoid that, we either omit reporting results in these cases (denoted as ”/” in Table 2) or show results as reported in the models’ respective papers (written in italic and denoted by a star after the number in Table 2). However, since we

could not verify the procedure we do not consider them in the analysis. Nevertheless, we see that DM-Count once again performs the best (88.97 MAE), and is closely followed by Bayesian Crowd Counting (90.43 MAE). Out of the three, our improved Bayesian CSRNet performs the worst (103.94 MAE).

Due the problems with the QNRF dataset, we, in addition to the evaluation of the models on SHA, SHB, and QNRF, also show the results of training the model on SHA train set and evaluating it with QNRF test set, since they both contain relatively dense crowds. The idea behind this experiment is also to see how well the models can learn to generalize, when trained on similar, but slightly different images. We see that the overall results here are significantly worse due to the models being trained on images that are cropped to contain crowds only, not including buildings and vegetation in the background. As images in the test set include those objects in the backgrounds, models could misinterpret them and count them as a crowd. The best results are given by the Bayesian Crowd Counting model (138.39 MAE), followed relatively closely by DM-Count (141.43 MAE) and our Bayesian CSRNet (145.03 MAE). The worst performance is once again achieved by the CSRNet (199.54 MAE).

Note that some results differ from the results reported in the author's papers. We argue, that the primary reason for this is that some authors use different implementations in their papers (such as CSRNet, whose authors provide two official implementations – one in Pytorch and one in Caffe). Furthermore, we were unable to train some models due to the computational limitations (and our limited hardware) on the QNRF dataset. In cases like these modifying the models in order to be able to retrain them on our limited settings could result in falsely lower results. In order to avoid that, we either omit reporting results in these cases (denoted as “/” in Table 2) or show results as reported in the original papers (denoted with “*” after the number in Table 2).

Qualitative Analysis We show the results of our improved model in Figures 4 and 5. In the first figure we show the input images from the ShanghaiTech part A and part B test set, and predicted density maps and inferred counts on a model trained on ShanghaiTech datasets. In the second figure we show the input image from the UCF-QNRF test set and predicted density maps and inferred counts on models trained on UCF-QNRF and ShanghaiTech datasets.

5. Conclusion

We reviewed definitions and provided concise descriptions of 5 CNN based models – CSRNet, Bayesian Crowd Counting, DM-Count, SFA-Net and SGA-Net. In addition we trained and evaluated the models ourselves, contrary to many other related works who just provided evaluation results from author's papers. We evaluated the models on ShanghaiTech part A dataset, ShanghaiTech part B dataset, and UCF-QNRF dataset. Additionally, we wanted to see how good the results are when training the model on one dataset (ShanghaiTech part A) and evaluating it on another (UCF-QNRF). We saw that the best overall results are those obtained on ShanghaiTech part B dataset, as models work better on images that are less complicated or have less dense crowds. The best results in terms of MAE on the ShanghaiTech part A were obtained with the SFA-Net model, followed closely by the DM-Count model. The first also performed best on the ShanghaiTech part



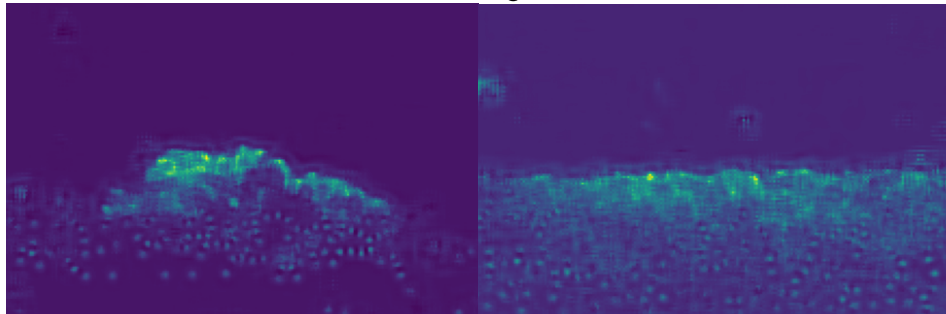
Fig. 4. Images in the left column represent input images from the ShanghaiTech part A (a – d) and ShanghaiTech part B (e – h) test sets, with 1156, 170, 106, and 92 annotated people, respectively. Images in the right column represent the predicted density maps obtained by our improved model Bayesian CSRNet. Estimated counts are 1116.56, 170.54, 107.42, and 89.33, respectively. We use the weights trained on the ShanghaiTech part A train images for the first two density maps (b and d) and weights trained on the ShanghaiTech part B train images for the bottom two density maps (f and h)



(a) Ground truth: 436.

(b) Ground truth: 479.

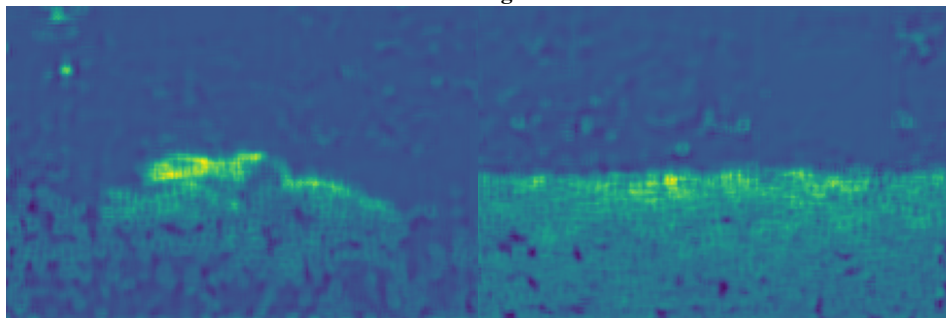
Trained on QNRF:



(c) Estimated: 437.30.

(d) Estimated: 518.70.

Trained on Shanghai Part A:



(e) Estimated: 444.24.

(f) Estimated: 466.95.

Fig. 5. The upper two images show input images from UCF-QNRF test set with 436 and 479 annotated people. The bottom 4 figures (c – f) show the predicted density maps obtained by our Bayesian CSRNet trained on UCF-QNRF (middle row) and on ShanghaiTech part A (bottom row) train sets. We see that the density maps in the middle row are clearer, as the model is trained on similar images that also contain buildings and streets, and it can better distinguish between them and the crowds. We also see that the inferred result is slightly better on the model trained on UCF-QNRF for the left input image, but the one trained on SHA performs slightly better for the input image from the right column

B, and the latter also performed best on the UCF-QNRF dataset. In terms of MSE, SGA-Net outperforms the SFA-Net on ShanghaiTech part B. The results of training the models on one dataset and evaluating them on the other were less good, however, that was expected due to the smaller train set with images that were cropped to contain crowds only, whereas the images from the test set also included buildings, sky, and vegetation.

In addition to the evaluation of the 5 mentioned models, we also suggested an improvement of the CSRNet. We implemented a new model based on the CSRNet and a Bayesian crowd counting loss function and pixel modeling. We showed that the new model is in fact an improvement of the original model.

Due to the computational limitations we were unable to train/evaluate some models on the QNRF dataset. For the future work we suggest the investigation of possible solutions. Since many datasets exist, we also suggest the evaluation of the models on other datasets (e.g., NWPU). SGA-Net also shows a possible investigation field, as it uses Inception-v3 model instead of VGG-16 or VGG-19, and yet still shows very promising results.

References

1. Aich, S., Stavness, I.: Leaf counting with deep convolutional and deconvolutional networks. In: Proceedings of the IEEE International Conference on Computer Vision Workshops. pp. 2080–2089 (2017)
2. Arteta, C., Lempitsky, V., Zisserman, A.: Counting in the wild. In: European conference on computer vision. pp. 483–498. Springer (2016)
3. Boominathan, L., Kruthiventi, S.S., Babu, R.V.: Crowdnet: A deep convolutional network for dense crowd counting. In: Proceedings of the 24th ACM international conference on Multimedia. pp. 640–644 (2016)
4. Chan, A.B., Liang, Z.S.J., Vasconcelos, N.: Privacy preserving crowd monitoring: Counting people without people models or tracking. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–7. IEEE (2008)
5. Chan, A.B., Vasconcelos, N.: Bayesian poisson regression for crowd counting. In: 2009 IEEE 12th international conference on computer vision. pp. 545–551. IEEE (2009)
6. Chapel, M.N., Bouwmans, T.: Moving objects detection with a moving camera: A comprehensive review. *Computer science review* 38, 100310 (2020)
7. Chen, K., Gong, S., Xiang, T., Change Loy, C.: Cumulative attribute space for age and crowd density estimation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2467–2474 (2013)
8. Chen, K., Loy, C.C., Gong, S., Xiang, T.: Feature mining for localised crowd counting. In: *Bmvc*. vol. 1, p. 3 (2012)
9. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). vol. 1, pp. 886–893. Ieee (2005)
10. Enzweiler, M., Gavrila, D.M.: Monocular pedestrian detection: Survey and experiments. *IEEE transactions on pattern analysis and machine intelligence* 31(12), 2179–2195 (2008)
11. Fan, Z., Zhang, H., Zhang, Z., Lu, G., Zhang, Y., Wang, Y.: A survey of crowd counting and density estimation based on convolutional neural network. *Neurocomputing* 472, 224–251 (2022), <https://www.sciencedirect.com/science/article/pii/S0925231221016179>
12. Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models. *IEEE transactions on pattern analysis and machine intelligence* 32(9), 1627–1645 (2009)

13. French, G., Fisher, M., Mackiewicz, M., Needle, C.: Convolutional neural networks for counting fish in fisheries surveillance video (2015)
14. Fu, M., Xu, P., Li, X., Liu, Q., Ye, M., Zhu, C.: Fast crowd density estimation with convolutional neural networks. *Engineering Applications of Artificial Intelligence* 43, 81–88 (2015)
15. Gall, J., Yao, A., Razavi, N., Van Gool, L., Lempitsky, V.: Hough forests for object detection, tracking, and action recognition. *IEEE transactions on pattern analysis and machine intelligence* 33(11), 2188–2202 (2011)
16. Gao, G., Gao, J., Liu, Q., Wang, Q., Wang, Y.: Cnn-based density estimation and crowd counting: A survey. *arXiv preprint arXiv:2003.12783* (2020)
17. Garcia-Garcia, B., Bouwmans, T., Silva, A.J.R.: Background subtraction in real applications: Challenges, current models and future directions. *Computer Science Review* 35, 100204 (2020)
18. Ge, W., Collins, R.T.: Marked point processes for crowd counting. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. pp. 2913–2920. IEEE (2009)
19. Girshick, R.: Fast r-cnn. In: *Proceedings of the IEEE international conference on computer vision*. pp. 1440–1448 (2015)
20. Haralick, R.M., Shanmugam, K., Dinstein, I.H.: Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics* (6), 610–621 (1973)
21. Idrees, H., Saleemi, I., Seibert, C., Shah, M.: Multi-source multi-scale counting in extremely dense crowd images. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 2547–2554 (2013)
22. Idrees, H., Tayyab, M., Athrey, K., Zhang, D., Al-Maadeed, S., Rajpoot, N., Shah, M.: Composition loss for counting, density map estimation and localization in dense crowds. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. pp. 532–546 (2018)
23. Jiao, L., Zhang, F., Liu, F., Yang, S., Li, L., Feng, Z., Qu, R.: A survey of deep learning-based object detection. *IEEE access* 7, 128837–128868 (2019)
24. Kang, D., Ma, Z., Chan, A.B.: Beyond counting: comparisons of density maps for crowd analysis tasks—counting, detection, and tracking. *IEEE Transactions on Circuits and Systems for Video Technology* 29(5), 1408–1422 (2018)
25. Kumagai, S., Hotta, K., Kurita, T.: Mixture of counting cnns: Adaptive integration of cnns specialized to specific appearance for crowd counting. *arXiv preprint arXiv:1703.09393* (2017)
26. Leibe, B., Seemann, E., Schiele, B.: Pedestrian detection in crowded scenes. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. vol. 1, pp. 878–885. IEEE (2005)
27. Lempitsky, V., Zisserman, A.: Learning to count objects in images. *Advances in neural information processing systems* 23, 1324–1332 (2010)
28. Li, M., Zhang, Z., Huang, K., Tan, T.: Estimating the number of people in crowded scenes by mid based foreground segmentation and head-shoulder detection. In: *2008 19th international conference on pattern recognition*. pp. 1–4. IEEE (2008)
29. Li, Y., Zhang, X., Chen, D.: Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 1091–1100 (2018)
30. Lin, S.F., Chen, J.Y., Chao, H.X.: Estimation of number of people in crowded scenes using perspective transformation. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 31(6), 645–654 (2001)
31. Lowe, D.G.: Object recognition from local scale-invariant features. In: *Proceedings of the seventh IEEE international conference on computer vision*. vol. 2, pp. 1150–1157. Ieee (1999)
32. Loy, C.C., Chen, K., Gong, S., Xiang, T.: Crowd counting and profiling: Methodology and evaluation. In: *Modeling, simulation and visual analysis of crowds*, pp. 347–382. Springer (2013)
33. Ma, Z., Wei, X., Hong, X., Gong, Y.: Bayesian loss for crowd count estimation with point supervision. In: *Proceedings of the IEEE International Conference on Computer Vision*. pp. 6142–6151 (2019)

34. Marana, A., Costa, L.d.F., Lotufo, R., Velastin, S.: On the efficacy of texture analysis for crowd monitoring. In: Proceedings SIBGRAP'98. International Symposium on Computer Graphics, Image Processing, and Vision (Cat. No. 98EX237). pp. 354–361. IEEE (1998)
35. Mundhenk, T.N., Konjevod, G., Sakla, W.A., Boakye, K.: A large contextual dataset for classification, detection and counting of cars with deep learning. In: European Conference on Computer Vision. pp. 785–800. Springer (2016)
36. Ojala, T., Pietikäinen, M., Mäenpää, T.: Gray scale and rotation invariant texture classification with local binary patterns. In: European Conference on Computer Vision. pp. 404–420. Springer (2000)
37. Onoro-Rubio, D., López-Sastre, R.J.: Towards perspective-free object counting with deep learning. In: European Conference on Computer Vision. pp. 615–629. Springer (2016)
38. Paragios, N., Ramesh, V.: A mrf-based approach for real-time subway monitoring. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. vol. 1, pp. I–I. IEEE (2001)
39. Pham, V.Q., Kozakaya, T., Yamaguchi, O., Okada, R.: Count forest: Co-voting uncertain number of targets using random forest for crowd density estimation. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 3253–3261 (2015)
40. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 779–788 (2016)
41. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems* 28, 91–99 (2015)
42. Ryan, D., Denman, S., Fookes, C., Sridharan, S.: Crowd counting using multiple local features. In: 2009 Digital Image Computing: Techniques and Applications. pp. 81–88. IEEE (2009)
43. Sabzmeydani, P., Mori, G.: Detecting pedestrians by learning shapelet features. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–8. IEEE (2007)
44. Saleh, S.A.M., Suandi, S.A., Ibrahim, H.: Recent survey on crowd density estimation and counting for visual surveillance. *Engineering Applications of Artificial Intelligence* 41, 103–114 (2015)
45. Shang, C., Ai, H., Bai, B.: End-to-end crowd counting via joint learning local and global count. In: 2016 IEEE International Conference on Image Processing (ICIP). pp. 1215–1219. IEEE (2016)
46. Sheng, B., Shen, C., Lin, G., Li, J., Yang, W., Sun, C.: Crowd counting via weighted vlad on a dense attribute feature map. *IEEE Transactions on Circuits and Systems for Video Technology* 28(8), 1788–1797 (2016)
47. Sindagi, V.A., Patel, V.M.: Cnn-based cascaded multi-task learning of high-level prior and density estimation for crowd counting. In: 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). pp. 1–6. IEEE (2017)
48. Sindagi, V.A., Patel, V.M.: A survey of recent advances in cnn-based single image crowd counting and density estimation. *Pattern Recognition Letters* 107, 3–16 (2018)
49. Sooraj, P., Kollerathu, V., Sudhakaran, V.: Real-time traffic counter using mobile devices. *Journal of Big Data Analytics in Transportation* 3(2), 109–118 (2021)
50. Tian, M., Guo, H., Chen, H., Wang, Q., Long, C., Ma, Y.: Automated pig counting using deep learning. *Computers and Electronics in Agriculture* 163, 104840 (2019)
51. Tian, Y., Sigal, L., Badino, H., De la Torre, F., Liu, Y.: Latent gaussian mixture regression for human pose estimation. In: Asian Conference on Computer Vision. pp. 679–690. Springer (2010)
52. Tseng, C.H., Kuo, Y.F.: Detecting and counting harvested fish and identifying fish types in electronic monitoring system videos using deep convolutional neural networks. *ICES Journal of Marine Science* 77(4), 1367–1378 (2020)
53. Tuzel, O., Porikli, F., Meer, P.: Pedestrian detection via classification on riemannian manifolds. *IEEE transactions on pattern analysis and machine intelligence* 30(10), 1713–1727 (2008)

54. Viola, P., Jones, M.J.: Robust real-time face detection. *International journal of computer vision* 57(2), 137–154 (2004)
55. Viola, P., Jones, M.J., Snow, D.: Detecting pedestrians using patterns of motion and appearance. *International Journal of Computer Vision* 63(2), 153–161 (2005)
56. Walach, E., Wolf, L.: Learning to count with cnn boosting. In: *European conference on computer vision*. pp. 660–676. Springer (2016)
57. Wang, B., Liu, H., Samaras, D., Nguyen, M.H.: Distribution matching for crowd counting. *Advances in Neural Information Processing Systems* 33 (2020)
58. Wang, C., Zhang, H., Yang, L., Liu, S., Cao, X.: Deep people counting in extremely dense crowds. In: *Proceedings of the 23rd ACM international conference on Multimedia*. pp. 1299–1302 (2015)
59. Wang, Q., Breckon, T.P.: Segmentation guided attention network for crowd counting via curriculum learning. *arXiv preprint arXiv:1911.07990* (2019)
60. Wang, Y., Ma, Z., Wei, X., Zheng, S., Wang, Y., Hong, X.: ECCNAS: Efficient crowd counting neural architecture search. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 18(1s), 1–19 (2022)
61. Wang, Y., Zou, Y.: Fast visual object counting via example-based density estimation. In: *2016 IEEE International Conference on Image Processing (ICIP)*. pp. 3653–3657. IEEE (2016)
62. Wu, B., Nevatia, R.: Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors. *International Journal of Computer Vision* 75(2), 247–266 (2007)
63. Wu, B., Nevatia, R.: Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors. In: *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*. vol. 1, pp. 90–97. IEEE (2005)
64. Xu, B., Qiu, G.: Crowd density estimation based on rich features and random projection forest. In: *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. pp. 1–8. IEEE (2016)
65. Zhan, B., Monekosso, D.N., Remagnino, P., Velastin, S.A., Xu, L.Q.: Crowd analysis: a survey. *Machine Vision and Applications* 19(5-6), 345–357 (2008)
66. Zhang, C., Li, H., Wang, X., Yang, X.: Cross-scene crowd counting via deep convolutional neural networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 833–841 (2015)
67. Zhang, S., Li, H., Kong, W.: A cross-modal fusion based approach with scale-aware deep representation for rgb-d crowd counting and density estimation. *Expert Systems with Applications* 180, 115071 (2021), <https://www.sciencedirect.com/science/article/pii/S0957417421005121>
68. Zhang, Y., Zhou, D., Chen, S., Gao, S., Ma, Y.: Single-image crowd counting via multi-column convolutional neural network. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 589–597 (2016)
69. Zhao, T., Nevatia, R., Wu, B.: Segmentation and tracking of multiple humans in crowded environments. *IEEE transactions on pattern analysis and machine intelligence* 30(7), 1198–1211 (2008)
70. Zhao, Z., Li, H., Zhao, R., Wang, X.: Crossing-line crowd counting with two-phase deep neural networks. In: *European Conference on Computer Vision*. pp. 712–726. Springer (2016)
71. Zhu, L., Zhao, Z., Lu, C., Lin, Y., Peng, Y., Yao, T.: Dual path multi-scale fusion networks with attention for crowd counting. *arXiv preprint arXiv:1902.01115* (2019)

Matija Terček holds a bachelor’s degree from Computer Science and Mathematics and is obtaining his master’s degree in Data Science. His main research areas are time series analysis, computer vision and lightweight convolutional neural networks.

Maša Kljun holds a bachelor's degree and master's degree from Computer Science and Mathematics. Her main research areas include predictive maintenance, time series analysis, and computer vision.

Peter Peer is a full professor at University of Ljubljana, Faculty of Computer and Information Science and holds PhD in computer and information science. As the head of the Laboratory of Computer Vision his latest research is focused mostly on biometrics with an emphasis on deep learning. He co-authored over 100 research papers in international conferences and journals.

Žiga Emeršič is a teaching assistant at University of Ljubljana, Faculty of Computer and Information Science and holds PhD in computer and information science. Within the Laboratory of Computer Vision, he is mostly dealing with biometrics with an emphasis on deep learning. He co-authored over 40 research papers in international conferences and journals.

Received: September 25, 2021; Accepted: June 05, 2022.

A novel Approach for sEMG Gesture Recognition using Resource-constrained Hardware Platforms

Matías J. Micheletto¹, Carlos I. Chesñevar², and Rodrigo M. Santos²

¹ *Golfo San Jorge* Research and Transfer Center
Ruta 1 KM 4 (9000) Chubut, Argentina
matias.micheletto@uns.edu.ar

² Institute for Computer Science and Engineering (CONICET-UNS)
Campus Palihue, Bahía Blanca (8000) Buenos Aires, Argentina
cic@cs.uns.edu.ar
ierms@criba.edu.ar

Abstract. Classifying human gestures using surface electromyographic sensors (sEMG) is a challenging task. Wearable sensors have proven to be extremely useful in this context, but their performance is limited by several factors (signal noise, computing resources, battery consumption, etc.). In particular, computing resources impose a limitation in many application scenarios, in which lightweight classification approaches are desirable. Recent research has shown that machine learning techniques are useful for human gesture classification once their salient features have been determined. This paper presents a novel approach for human gesture classification in which two different strategies are combined: a) a technique based on autoencoders is used to perform feature extraction; b) two alternative machine learning algorithms (namely J48 and K*) are then used for the classification stage. Empirical results are provided, showing that for limited computing power platforms our approach outperforms other alternative methodologies.

Keywords: sEMG, gesture recognition, autoencoder, decision trees, nearest neighbors.

1. Introduction

Wearable sensors and mobile devices have introduced a wide range of new applications to support daily life. These applications involve from simple pulse and oxygen sensors that measure the physical activity during training (e.g. [31, 28]) up to more complex e-health applications (e.g. [15, 32]). Gesture recognition from electromyographic (EMG) sensors is a relevant area in which wearable sensors play a major role, particularly surface EMG sensors (sEMG for short). Gesture recognition is intended to be used in different application settings to provide human-like communication through gestures in cases where speech recognition is not possible or appropriate (e.g. because of having a noisy environment or dealing with a user who is not able to type or interact with a touch screen).

In the last years there have been many different approaches for gesture recognition using EMG and sEMG signals [3]. However, most of these approaches focus on obtaining a high overall accuracy of the proposed model as the main goal to be attained. Even though accuracy is a very relevant issue in classification, the assessment of computational costs associated with the resulting models plays also a significant role in many situations (e.g.

a neural network model can have a high accuracy for classifying gestures but only at the expense of a high computational cost for building the model).

In this respect, our research is in line with the concept of *frugal innovation*, which refers to the development of technology for solving modern problems by adapting the requirements to the constrained accessibility of the local market [1, 11, 26, 12]. This aspect is crucial for many developing and emerging countries, where the deployment of new technologies (such as sEMG gesture recognition) is constrained by availability and prices of hardware resources. In this context, lightweight classification techniques are particularly relevant. To the best of our knowledge, the trade-off between accuracy vs. computational resources involved remains as an important issue to be discussed, particularly when considering the real-time performance of gesture recognition algorithms implemented in low cost platforms.

In this article we present a lightweight classification approach for gesture recognition from sEMG signals based on a combined strategy: first, *autoencoders* –a particular subclass of artificial neural networks (ANN)– are used to perform feature extraction. On the basis of the features that have been identified, two well-known classification algorithms (namely J48 and K*) are then used for gesture recognition. From our experiments we can conclude that the proposed approach enables to correctly identify nine different gestures with a high accuracy (87% for a generic user, and up to 96% when training the algorithms with data from a single user).

The rest of this article is structured as follows. Section 2 discusses related work for gesture recognition with an emphasis on sEMG usage, summarizing some relevant advances in the state of the art on the topic. Section 3 presents our approach for gesture recognition, integrating commercial low-cost sEMG sensors and a lightweight model for gesture classification. This model is based on autoencoders for selecting salient features along with the algorithms J48 and K* (corresponding to decision tree learning and instance-based learning using an entropic distance measure, respectively). Sections 5 and 6 show how the lightweight model can be effectively built and how to assess the resulting computer hardware cost. Section 7 discusses the obtained results, contrasting them with other more expensive alternatives found in the literature. Finally, Section 8 concludes.

2. Related work

The process of gesture recognition from acquired signals using sEMG sensors involves usually three stages, namely *filtering*, *feature extraction* and *classification*. First, in the filtering step, signals must be conditioned to eliminate possible interference or noise and to extract the time window that contains the dataset necessary to perform gesture classification. In some cases, the signal envelope is extracted directly from the raw data acquired through the sensors. Next, as a second stage, a feature extraction method is usually employed, providing the basis for performing the gesture recognition process by means of an appropriate classification algorithm, which will take place in the third and final stage.

Next we summarize some of the most relevant papers related to gesture classification, grouping them according to the proposed method for each processing stage, contrasting as well the computing platforms used for validating their performance.

2.1. First stage – data acquisition and filtering

An in-depth analysis of this first stage is usually ignored in the literature. Some research work makes use of public datasets and focuses almost exclusively on the resulting algorithm performance [17, 5]. In other cases, commercial sEMG sensors are used with embedded filtering and denoising capabilities [2, 30, 7, 14, 13]. Less frequently, some articles include a specific section to detail the underlying features for acquisition and filtering [21].

2.2. Second stage – feature extraction

In the literature there are many alternatives for feature extraction for gesture recognition. The most simple consists on using statistical descriptors of the raw signal (e.g. mean average value (MAV), minimum and maximum values, standard deviation (STD), slope sign changes (SSL), zero crossings (ZC), root mean square value (RMS), waveform length (WL), autoregression coefficients (AR), among many others [2, 21, 30].

Another common technique consists on using the discrete wavelet transform (DWT) [8, 6, 24], the Fourier transform (FFT) [19] or the short time Fourier transform (STFT) [17]. These approaches imply higher computational cost, and for the case of embedded devices they are usually implemented using dedicated processing units for computing the transforms in real time.

Fractal dimension (FD) is a less explored method that takes advantage on the fact that EMG signals may show self similarity traces (as proposed in [5]). Another widely used technique applied to dimensionality reduction (detecting features which are deemed as non-relevant for classifying EMG data) is principal component analysis (PCA) and independent component analysis (ICA) [22]. The use of autoencoders (as proposed in this article) as a method for feature extraction has also proven to be useful, being its complexity as shallow neural networks relatively low [7, 18, 13, 22].

2.3. Third stage – gesture classification

For the third stage, the spectrum of possible approaches is wide, as each classifier may have many particular variations. Artificial neural networks (ANN) in their many forms are one of the most frequent methods, as they have gained relevance in the last years, growing along with the increase of available computing power. Thus, we find ANNs ranging from simple perceptron architectures up to deep and convolutional neural networks being used in the classification stage for EMG gesture recognition [13, 5, 18, 29]. In general, the main drawback of ANN approaches derives from the high computational cost required for both the training and inference stages.

The linear discriminant analysis (LDA) method basically consists on finding a linear combination of the features of the model in order to build a space where the projection of the data points corresponding to each class are as distant as possible from each other, allowing to define gesture classification rules. This technique was applied in [18]. The major drawback of this method is the requirement that data should be normally distributed, being thus prone to negative effects from outliers and requiring a carefully data cleaning in the model building phase.

Support vector machines (SVM) share some aspects with LDA and are also applied in gesture classification scenarios [2, 7]. SVM consist on computing the set of hyperplanes that best separates the different classes. However, this method allows to define a more complex decision boundary than LDA, allowing to focus on the correct classification of outliers. The computational cost and accuracy trade-off of this technique can also be manageable.

As we will detail in the next section, our proposal is based on two lightweight classification algorithms (namely decision trees using J48 and K-nearest neighbors using K*). These algorithms were used for gesture classification in previous research work (e.g. [8, 24, 21, 30, 18, 17]) with two main differences: a) they were part of more complex ensemble models such as gradient boosting (GBDT) and bagging ; b) the focus of their usage was on accuracy rather than on finding a suitable approach for the trade-off between accuracy and high computational cost, as proposed in this article.

3. Data acquisition and preprocessing stage

To perform the data acquisition, commercial EMG sensors were used (in our case MyoWare Muscle Sensors, manufactured by Advancer Technologies³). These sensors are characterized by being small sized, low cost and having a good market availability.



Fig. 1. Surface EMG sensors placement

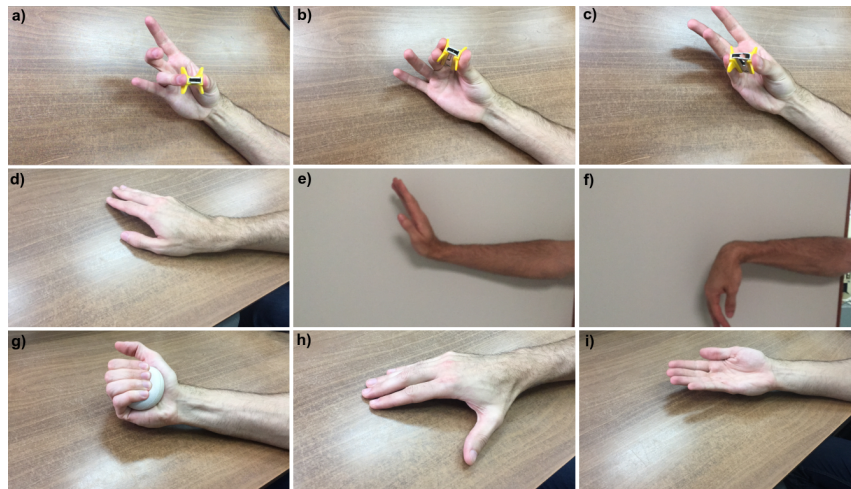


Fig. 2. Selected gestures: a) anu, b) may, c) men, d) adu, e) ext, f) fle, g) pel, h) pul, i) sup

3.1. Sensor placement

The sensor electrodes measure an electric potential difference generated by muscle activity. The resulting signal is amplified, processed through a 400Hz low-pass filter and rectified in order to obtain the signal envelope as a final output. The sensor has a third reference electrode that must be located on an area with little muscle activity to be taken as a comparison point for voltage reference. In general, skin areas close to joints or bones have little electrical activity, so they are appropriate for the placement of the reference electrode.

Figure 1 shows the adhesive electrode pads placement in the forearm of a sample subject. Following the guidelines provided by [23], after contrasting several tests the three sensor channels were located in the following positions:

Channel 1: Placed on the extensor carpi ulnaris and extensor digitorum muscles (involved in the extension of the fingers, wrist and adduction of the hand).

Channel 2: Placed on the flexor carpi ulnaris muscle, which intervenes in the flexion of the wrist.

Channel 3: Located on the brachioradial muscle, which is the main actuator of the supination movement.

Once the three sensors were placed on the subject, their output data was measured and recorded using a Hioki 8861-50 oscilloscope (with 16 channels of 16-bit each). Sampling frequency was set at 200 Hz.

3.2. Gestures selection

Different gestures were selected, made with the right or left hand, so that the signals observed in the oscilloscope resulted in a good amplitude and differences observable to the

³ <http://www.advancertechnologies.com>

naked eye. The nine gestures shown in Figure 2 were selected and identified with different acronyms: Ring finger (anu), middle finger (may), pinky finger (men), wrist adduction (adu), wrist extension (ext), wrist flexion (fle), ball squeeze (pel), thumb extension (pul) and wrist supination (sup).

The executor was asked to perform the movements of the selected gestures repeatedly every certain interval of time, spaced enough to identify the time window with the main variations of the signals associated with the gesture execution. The acquired data was exported in plain text format to be later analyzed.

4. Feature extraction stage

An autoencoder was used as feature extraction stage. Both the input and output layers were configured using the sigmoid activation function. The three sEMG signals were sampled using a window of 200 16-bit integer values each, so that the size of the input layer of the autoencoder comprised 600 values.

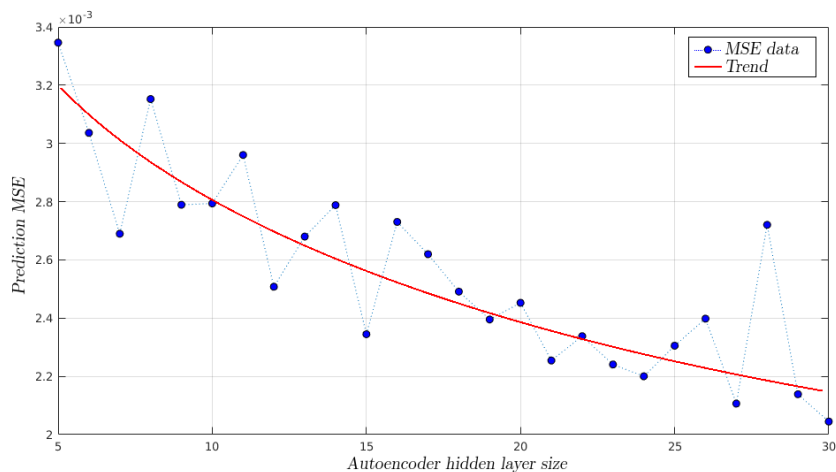


Fig. 3. Mean square error of an autoencoder: prediction MSE vs. hidden layer size

The hidden layer size of an autoencoder is the main attribute that limits the performance of the model as a lossy compression method and is also related to the classifier performance (when used together with a classification algorithm). To avoid the tight coupling between the feature extraction and classification stages, the hidden layer size of the autoencoder was determined measuring its performance when reconstructing the compressed signal and not when classifying the gestures. In order to do this, several tests were performed between 5 to 30 parameters, calculating the mean square error (MSE) in the reconstruction of the validation data (which consisted on random selected holdout subsets with a proportion of 25% of the total dataset).

The learning process was limited to 500 epochs. Fig. 3 shows the MSE values when predicting validation data as the number of neurons in the hidden layer increases. Fig.

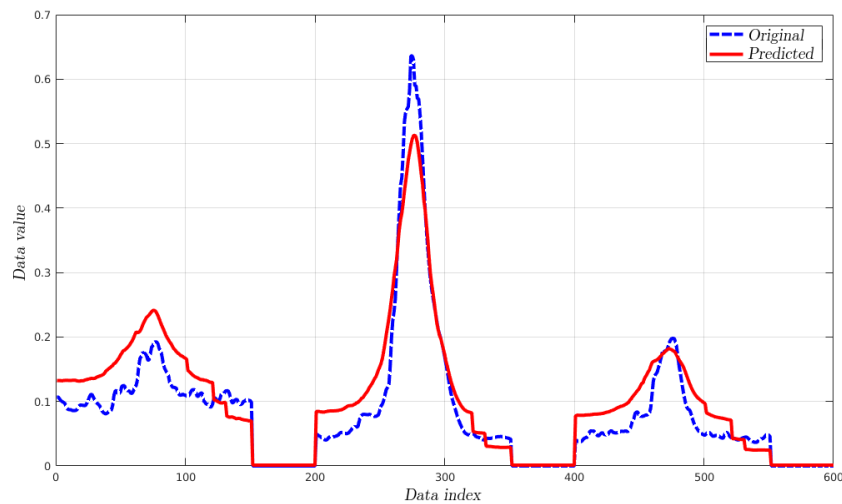


Fig. 4. Input (original) and output (predicted) signals of the trained autoencoder

4 shows the comparison between a signal from the validation data set and its respective reconstruction obtained with the trained autoencoder. It can be observed that the generated signal is not as noisy as the original one, preserving the overall associated shape.

For the classification stage, a hidden layer size of 15 neurons was chosen (as a trade-off between error rate and number of features). Since the holdout subset was randomly selected, three repetitions were performed when training the autoencoders, generating three different classification datasets (one for each executor and a fourth one that combines the gestures from all executors). This resulted in a total of 12 different datasets with 15 features and the labels for the corresponding gestures.

5. Classification stage

Two alternative classification models were built using the open-source software Weka [27, 10]. Two classification schemes were used: *decision trees*, implemented via the C.45 algorithm (available in Weka as J48 [20]) and a particular variation of *k nearest neighbors* (using entropy as a distance function via the so-called K* algorithm [4]). Two validation methods were used: a) holdout with 3/4 train/validation proportion, and b) 10 fold cross-validation. In general, very similar results were obtained between both validation techniques and between both classification methods.

The classification process was repeated for each of the three iterations in order to verify that the results remained constant in different cases. The resulting sets were named with the letter that identifies the executor and the index that indicates the repetition number (e.g. *M3* stands for the data set obtained from the 3rd iteration when measuring results from executor *M*).

It must be noted that the generated decision trees do not use all the features at the same time. This suggests that the number of features could still be reduced, which would

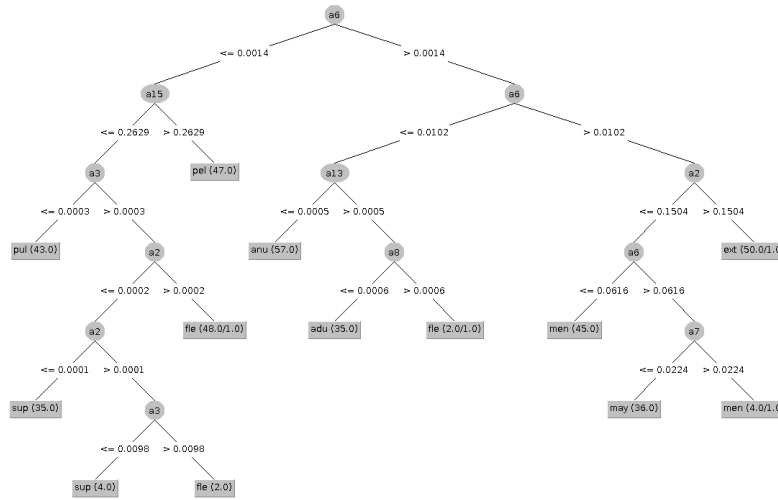


Fig. 5. Decision Tree model for the gesture recognition of the *M2* dataset

imply removing neurons from the trained autoencoder model. To give an example of the decision tree model complexity, Fig. 5 shows the generated model by means of the C4.5 method for the gesture classification trained with the *M2* data set.

6. Hardware requirements analysis

We analyzed the computational complexity and resource requirements of the proposed techniques following similar criteria as those provided by [16, 9, 25]. As an upper bound, we considered a 15-neuron fully connected encoder for 600 16-bit input values and a decision tree of 185 nodes (the latter corresponds to the most complex decision tree from the ones calculated in our experiments).

The data encoder stage requires memory space to store the weight values of 601 parameters (600 weights plus one bias) per neuron and to perform 1,201 floating point operations (FLOPs). The activation function will be evaluated one time per hidden layer neuron, which means the evaluation of a quantized sigmoid function, resulting in a total of 4 FLOPs. Thus, for an encoder with 15 neurons in its hidden layer, the total number of FLOPs required to extract the features of a 600 data values window sums up to 18,075 FLOPs.

The 185-node decision tree requires 185 FLOPs which may result negligible compared to the feature extraction stage. Consequently, using a rough estimation, the total number of FLOPs required is approx. 20,000.

To simplify the real-time analysis, we will assume that a window of 600 data values is processed at the signal sampling frequency (i.e. 200Hz). This is because the classifier was not trained with time-shift invariance considerations. Then, the processor speed should allow to perform all the computations in less than 5 ms. For the case of 16 bit architectures

where a floating point operation could be performed on each clock tick, a processor of at least 4 MHz is required to perform 20,000 operations in 5 ms. For 8-bit processors, the clock frequency needs to be doubled, being thus at least 8MHz. This approximation does not take into account the time required to perform the program control, but still by doubling the minimum frequency we would obtain 16Mhz (which is a nominal frequency for low resource microcontrollers).

The memory usage on the other hand, takes into account the storage for all the model parameters and the program itself, which includes all the control instructions. The set of 16-bit 9,015 weights of the encoder itself takes up 17.61Kb of flash memory usage. Most embedded devices allow to expand their capacity by adding external memory, however this implies a non-negligible reading time and should be taken into account on the overall computing time.

The complexity of the classification stage of the instance based learner is harder to analyze as it depends on the training stage results. In order to make an estimate of the required computing resources, we will assume that an instance based classifier uses all the 15 features and 1,000 instances, from the total of 1,142 rows. Assuming two bytes of memory per feature plus a byte per class label, this represents a total memory usage of 31,000 bytes or 30.27Kb (only to store the database with examples).

Regarding the FLOPs number, we will assume that the classification stage of this algorithm compares the test instance to all the given instances to determine the probability that the test instance belong to the class of each compared example. We will also assume that the computation of this probability requires two FLOPs per instance in the database. Note that the entropy function is defined by equation 1,

$$H(x) = \sum_{i=0}^{N-1} -p_i(x) \cdot \log(p_i(x)). \quad (1)$$

Where $N = 9$ is the number of classes. The logarithm implementation for the range (0.0, 1.0) using a four-degree Taylor series would involve up to eight FLOPs, and adding another FLOP for the multiplication in equation 1 results in 9 FLOPs. Then, adding the operations used for the computation of the probability $p_i(x)$, the entropy function $H(x)$ would require 2,009 FLOPs, and should be performed nine times (i.e., one time per class), which results in a total of 18,081 FLOPs. This number falls inside the estimated range as for the case of the decision tree model.

In summary, the requirements to storage the complete classification model and compute real-time results would be 16MHz as operating frequency and around 32Kb of required memory storage (matching the specs of a low cost microcontroller, such as the AVR ATmega 328p⁴).

7. Results and discussion

Table 1 shows the accuracy of each classification method using different validation techniques. All values are expressed in percentages of classification efficiency and each cell

⁴ <https://en.wikipedia.org/wiki/ATmega328>

corresponds to the average results between the three repetitions. The *E*, *L* and *M* sets indicate the different gesture executors and the *E+L+M* set contains the data from all three executors. Average values across columns and rows are shown in bold.

Table 1. Comparative average accuracy values by validation method

Set	J48			K*		
	Holdout	Crossvalidation	Average	Holdout	Crossvalidation	Average
E	96	94.2	95.1	93.9	95.6	94.7
M	88.4	88.4	88.4	91.8	94	92.6
L	97.2	95.6	96.4	98.3	98.4	98.4
Average	93.87	92.73	93.3	94.67	96	95.23
E+M+L	87.1	86.9	87	92	92.8	92.4

It can be observed that the K* method obtained slightly better accuracy percentages than J48. However, the decision tree classification is computed about ten times faster than K*. The Table 2 shows the sizes and number of characterization parameters used by each decision tree (numbered columns correspond to each of the three repetitions). This result shows that when the decision tree models are trained with data corresponding to a single executor many parameters are left aside by the classifier. This allows to reduce the computational cost of the model by removing neurons from the hidden layer of the already trained encoder while maintaining the same classification accuracy level. However, when using data from multiple executors, the decision tree model complexity grows notably, and almost all 15 parameters are used.

Another experiment was performed in order to test to what extent the classification accuracy is affected by reducing the number of characterization parameters, i.e. the autoencoder's hidden layer size. In table 3, the accuracy values are shown when repeating the training process for the classifiers and using only eight parameters to characterize the signals of the three data sets combined. Here we can observe that when reducing the parameter number from 15 to 8, the accuracy of the classification stage is reduced around a 6% for the decision tree and 7.9% for the K* model. We can conclude that a substantial improvement is achieved in terms of computational cost, considering that the autoencoder size was reduced to almost the half.

Table 2. Number of used features and the resulting tree size for the classification decision tree using J48

Set	Number of used features						Size of tree (node count)					
	Holdout			Crossvalidation			Holdout			Crossvalidation		
	1	2	3	1	2	3	1	2	3	1	2	3
E	12	11	8	8	10	9	31	37	25	25	31	27
M	12	12	13	13	10	10	39	49	35	39	33	35
L	6	8	7	8	10	11	19	21	25	25	23	27
E+M+L	15	15	15	15	14	15	149	165	137	185	133	165

Table 3. Classification results using eight characterization parameters

Set	J48		K*	
	Holdout	Crossvalidation	Holdout	Crossvalidation
E+M+L	87.9	86.6	86.9	87.7

Table 4. Confusion matrix for the decision tree trained with ELM3 set

Class	anu	may	men	adu	ext	fle	pel	pul	sup
anu	40	0	0	0	0	4	0	1	0
may	0	46	1	0	0	1	1	1	0
men	0	1	43	1	3	1	0	0	0
adu	0	0	0	27	4	1	0	5	0
ext	0	1	1	4	34	0	0	0	0
fle	2	0	0	0	0	32	1	0	0
pel	0	0	1	0	0	1	43	1	1
pul	0	1	2	3	0	2	0	34	1
sup	0	0	0	0	0	0	2	1	32

Table 4 shows the resulting confusion matrix for the decision tree generated with Weka's J48 algorithm applied to the complete data set of 1142 instances and validated via the holdout method. The rows correspond to the classes of the set and the columns correspond to the classification results for each instance. From 381 instances evaluated, 335 were correctly classified (87.93%).

8. Conclusion

We have presented a lightweight classification approach to sEMG gesture recognition. The approach includes a feature extraction stage that uses autoencoders of 15-neuron hidden layer and a classification stage involving two alternatives: a decision tree (via the J48 algorithm) and an instance based learner (via the K* algorithm).

Based on previous work and by a rough estimate of minimal resource requirements, we have shown that the proposed prediction models can be implemented on low resource architectures. An example of a hardware platform that matches with the minimum requirements is the AVR ATmega328p microcontroller (a widely off-the-shelf used platform with a low price and good market availability, used for the Arduino UNO prototyping platform). The resulting decision tree based classifier was able to correctly identify nine gestures with an accuracy of 87% (for a generic user) and up to 93.3% when calibrating the algorithms using data from a single user. The instance based learning classifier reaches 92.4% accuracy when trained with multiple user data and 95.23% when calibrated by a single user data.

The results presented on this work opens the possibility of implementing sEMG based gesture classification systems on low-resource platforms with high market availability,

paving the way for developing several low cost and accessible products, from commercial small wearable devices to educational prototyping tools.

References

1. Agarwal, N., Brem, A.: Frugal innovation-past, present, and future. *IEEE Engineering Management Review* 45(3), 37–41 (2017)
2. Akhmadeev, K., Rampone, E., Yu, T., Aoustin, Y., Carpentier, E.L.: A testing system for a real-time gesture classification using surface emg. *IFAC-PapersOnLine* 50(1), 11498 – 11503 (2017), 20th IFAC World Congress
3. Buongiorno, D., Cascarano, G.D., De Feudis, I., Brunetti, A., Carnimeo, L., Dimauro, G., Bevilacqua, V.: Deep learning for processing electromyographic signals: A taxonomy-based survey. *Neurocomputing* 452, 549–565 (2021)
4. Cleary, J.G., Trigg, L.E.: K*: An instance-based learner using an entropic distance measure. In: Prieditis, A., Russell, S. (eds.) *Machine Learning Proceedings 1995*, pp. 108–114. Morgan Kaufmann, San Francisco (CA) (1995)
5. Coelho, A.L., Lima, C.A.: Assessing fractal dimension methods as feature extractors for emg signal classification. *Engineering Applications of Artificial Intelligence* 36, 81 – 98 (2014)
6. David Orjuela-Cañón, A., Ruíz-Olaya, A.F., Forero, L.: Deep neural network for emg signal classification of wrist position: Preliminary results. In: *2017 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*. pp. 1–5 (Nov 2017)
7. Farouk Ibrahim Ibrahim, M., Ali Al-Jumaily, A.: Auto-encoder based deep learning for surface electromyography signal processing. *Advances in Science, Technology and Engineering Systems Journal* 3, 94–102 (01 2018)
8. Gokgoz, E., Subasi, A.: Comparison of decision tree algorithms for emg signal classification using dwt. *Biomedical Signal Processing and Control* 18, 138 – 144 (2015)
9. Gordon, A., Eban, E., Nachum, O., Chen, B., Wu, H., Yang, T., Choi, E.: Morphnet: Fast and simple resource-constrained structure learning of deep networks (2018)
10. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: *The weka data mining software: An update* (2000)
11. Hossain, M.: Frugal innovation: Conception, development, diffusion, and outcome. *Journal of Cleaner Production* 262, 121456 (2020)
12. Hossain, M.: Frugal innovation and sustainable business models. *Technology in Society* 64, 101508 (2021)
13. Huang, Y., Chen, K., Zhang, X., Wang, K., Ota, J.: Joint torque estimation for the human arm from semg using backpropagation neural networks and autoencoders. *Biomedical Signal Processing and Control* 62, 102051 (2020)
14. Jiang, Y., Chen, C., Zhang, X., Chen, C., Zhou, Y., Ni, G., Muh, S., Lemos, S.: Shoulder muscle activation pattern recognition based on semg and machine learning algorithms. *Computer Methods and Programs in Biomedicine* 197, 105721 (2020)
15. Lowe, S., ÓLaighin, G.: Monitoring human health behaviour in one’s living environment: A technological review. *Medical Engineering & Physics* 36(2), 147–168 (2014)
16. Mitra, S., Chattopadhyay, P.: Challenges in implementation of ann in embedded system. In: *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*. pp. 1794–1798 (2016)
17. Rabin, N., Kahlon, M., Malayev, S., Ratnovsky, A.: Classification of human hand movements based on emg signals using nonlinear dimensionality reduction and data fusion techniques. *Expert Systems with Applications* 149, 113281 (2020)
18. Zia ur Rehman, M., Gilani, S.O., Waris, A., Niazi, I.K., Slabaugh, G., Farina, D., Kamavuako, E.N.: Stacked sparse autoencoders for emg-based classification of hand motions: A comparative multi day analyses between surface and intramuscular emg. *Applied Sciences* 8(7) (2018)

19. Sadikoglu, F., Kavalcioglu, C., Dagman, B.: Electromyogram (emg) signal detection, classification of emg signals and diagnosis of neuropathy muscle disease. *Procedia Computer Science* 120, 422 – 429 (2017), 9th International Conference on Theory and Application of Soft Computing, Computing with Words and Perception, ICSCCW 2017, 22-23 August 2017, Budapest, Hungary
20. Salzberg, S.L.: C4.5: Programs for machine learning by j. ross quinlan. morgan kaufmann publishers, inc., 1993. *Machine Learning* 16, 235–240 (1994)
21. Song, W., Han, Q., Lin, Z., Yan, N., Luo, D., Liao, Y., Zhang, M., Wang, Z., Xie, X., Wang, A., Chen, Y., Bai, S.: Design of a flexible wearable smart semg recorder integrated gradient boosting decision tree based hand gesture recognition. *IEEE Transactions on Biomedical Circuits and Systems* 13(6), 1563–1574 (2019)
22. Spuler, M., Irastorza Landa, N., Sarasola Sanz, A., Ramos-Murguialday, A.: Extracting muscle synergy patterns from emg data using autoencoders. In: Villa, A.E., Masulli, P., Pons Rivero, A.J. (eds.) *Artificial Neural Networks and Machine Learning - ICANN 2016*. pp. 47–54. Springer International Publishing, Cham (2016)
23. Stegeman, D., Hermens, H.: Standards for surface electromyography: The european project surface emg for non-invasive assessment of muscles (seniam). *SENIAM Project* 1, 352–360 (01 2007)
24. Subasi, A., Yaman, E., Somaily, Y., A. Alynabawi, H., Alobaidi, F., Altheibani, S.: Automated emg signal classification for diagnosis of neuromuscular disorders using dwt and bagging. *Procedia Computer Science* 140, 230–237 (01 2018)
25. Tang, R., Adhikari, A., Lin, J.: Flops as a direct optimization objective for learning sparse neural networks. In: *NIPS 2018 Workshop on Compact Deep Neural Networks with Industrial Applications (CDNNRIA)*. pp. 1–4 (11 2018)
26. Winkler, T., Ulz, A., Knobl, W., Lercher, H.: Frugal innovation in developed markets - adaption of a criteria-based evaluation model. *Journal of Innovation and Knowledge* 5(4), 251–259 (2020)
27. Witten, I.H., Frank, E., Hall, M.A., Pal, C.J.: *Data Mining, Fourth Edition: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 4th edn. (2016)
28. Xiao, N., Yu, W., Han, X.: Wearable heart rate monitoring intelligent sports bracelet based on internet of things. *Measurement* 164, 108102 (2020)
29. Yu, Y., Chen, C., Sheng, X., Zhu, X.: Multi-dof continuous estimation for wrist torques using stacked autoencoder. *Biomedical Signal Processing and Control* 57, 101733 (2020)
30. Cedeño Z., C., Cordova-Garcia, J., Asanza A., V., Ponguillo, R., Muñoz M., L.: k-nn-based emg recognition for gestures communication with limited hardware resources. In: *2019 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computing, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. pp. 812–817 (2019)
31. Zhang, L., Yang, L., Wang, Z., Yan, D.: Sports wearable device design and health data monitoring based on wireless internet of things. *Microprocessors and Microsystems* p. 103423 (2020)
32. Zhou, X.: Wearable health monitoring system based on human motion state recognition. *Computer Communications* 150, 62–71 (2020)

Matías Micheletto received the Electrical Engineering degree and the Doctorate (Ph.D.) in Engineering from the Universidad Nacional del Sur (UNS) Bahía Blanca, Argentina in 2016 and 2020 respectively. He is currently a posdoctorate CONICET scholar at "Golfo San Jorge" Research and Transfer Center, in Comodoro Rivadavia, Argentina. His research interest is in the field of numerical modeling and optimization, evolutionary computing and embedded systems engineering.

Carlos Chesñevar is principal researcher and full professor at the Universidad Nacional del Sur in Bahía Blanca, Argentina. From 2015 he is the director of the Institute of Computer Science and Engineering (ICIC CONICET UNS). His research is oriented towards different applications of AI technologies. He has participated as PC member in most major AI conferences (IJCAI, AAMAS, ECSQARU, etc.) and has published more than 40 journal articles, 10 book chapters and more than 100 international conference papers. He has led international projects in Artificial Intelligence funded by DAAD (Deutscher Akademischer Austauschdienst) and Microsoft Research Latinamerica.

Rodrigo Santos received the Electrical Engineering degree and the Doctorate (Ph.D.) in Engineering from the Universidad Nacional del Sur (UNS) in Bahía Blanca, Argentina in 1997 and 2001, respectively. He is currently CONICET Adjunct Researcher and Full Professor at UNS. His research interests are scheduling embedded real-time systems, data ad-hoc networks and collaborative systems. He has been a visiting Professor at Universidad Nacional de San Agustín, Arequipa, Peru, Scuola Superiore Sant'Anna, Universidad Nacional de la Patagonia San Juan Bosco and Universidad Argentina de la Empresa.

Received: February 28, 2022; Accepted: June 25, 2022.

Fabric-GC: A Blockchain-based Gantt Chart System for Cross-organizational Project Management

Dun Li^{1,*}, Dezhi Han^{1,*}, Benhui Xia^{1,*}, Tien-Hsiung Weng^{2,**}, Arcangelo Castiglione³, and Kuan-Ching Li^{2,**}

¹ College of Information Engineering, Shanghai Maritime University
201306, Shanghai, China
lidunshmtu@outlook.com

² Dept. of Computer Science and Information Engr. (CSIE), Providence University
43301, Taichung, Taiwan
kuancli@pu.edu.tw

³ Department of Computer Science University of Salerno, Fisciano
84084, SA, Italy
arcastiglione@unisa.it

Abstract. Large-scale production is always associated with more and more development and interaction among peers, and many fields achieve higher economic benefits through project cooperation. However, project managers in the traditional centralized approach cannot rearrange their activities to cross-organizational project management. Thanks to its characteristics, the Blockchain can represent a valid solution to the problems mentioned above. In this article, we propose Fabric-GC, a Blockchain-based Gantt chart system. Fabric-GC enables to realize secure and effective cross-organizational cooperation for project management, providing access control to multiple parties for project visualization. Compared with other solutions, the proposed system is versatile, as it can be applied to project management in different fields and achieve effective and agile scheduling. Experimental results show that Fabric-GC achieves stable performance in large-scale request and processing distributed environments, where the data synchronization speed of the consortium chain reached four times faster than a public chain, achieving faster data consistency.

Keywords: Cross-organizational secure cooperation, Blockchain, Gantt chart, Project management, Hyperledger fabric, Data sharing

1. Introduction

Project Management (PM) is an activity carried out by project managers who plan, organize, direct, coordinate, control, and evaluate projects through scientific and management activities to reach the project objectives. Due to limited available resources (i.e., time, budget, labor), there are many constraints on the activities and development that affect the overall schedule of PM. Due to the reasons mentioned above, project management can coordinate and guide project implementation under constraints and limitations, reducing complexity and operational costs and improving the efficiency of project implementation.

* Authors contribute equally to this work

** Corresponding authors

Traditional project management systems typically rely on the *Storage-Business-Interface* triad. *Storage* provides permanent storage and ready access to data, *Business* includes all functional modules of the management system, and *Interface* converts data into meaningful management models through visualization [12]. To date, the *Gantt* chart is one of the most commonly used project management tools, as it divides the entire project into smaller portions of tasks sequenced under specific rules (e.g., time). In this way, project managers can track the execution status of each task under a given schedule and monitor the completion level of the entire project. This tracking enables the evaluation of the entire project resource budget, optimizing the completion time schedule to make adjustments to the execution plan, and most importantly, making the right decisions.

With the further advancement of science and technology and the development of productivity, multi-party project cooperation is a standard practice, widely used in scientific R & D, industrial production, software development, supply chain [8] among several other fields. Indeed, the collaboration between organizations and individuals with different technologies enlarges the rate for the success of complex projects [50]. Nevertheless, cross-organizational projects pose difficulties for project managers in managing task scheduling and progress feedback that relies on timely information sharing [57]. The independence and heterogeneity among participating organizations may turn data sharing difficult. Besides, traditional data sharing relies on third-party organizations (e.g., cloud, specialized service provider, transcription services, call center services, consulting), and therefore, the privacy and security of data cannot be guaranteed [29, 33, 55, 56]. In fact, although in general, the sharing of information is of great benefit and provides several advantages for all the entities involved, however, these entities may not trust each other, or even worse, they may compete with each other. Consequently, in the context of cross-organizational project management, safety is a crucial factor, which must be guaranteed for the entire life cycle of project management. For example, in the field of cross-organizational collaborative decision-making, there is a great deal of private information that companies are reluctant to leak, even when such information is needed for collaborative data analysis. This issue is emphasized on the one hand by the lack of adequate mechanisms for protecting privacy in cross-organizational collaborative decision-making processes and on the other by the ever-increasing use of big data [59]. Similarly, the same issues described above apply to workflow management, which is crucial for improving business productivity. Indeed, many workflow systems go outside the organizational boundaries and often require organizations to interact with each other. Each organization has its own private business processes and can operate autonomously, but at some point, all the organizations involved need to be synchronized to complete certain tasks. It is easy to imagine how such organizations are unwilling to share business details with others [38]. Another non-negligible problem in this context is that while some organizations may be allied for a project, the same organizations may be competitors for other projects [54]. Furthermore, ever-increasing security issues are emerging regarding cross-organizational cooperation in ubiquitous computing environments, mainly due to the interoperability problems deriving from the different security mechanisms and policies put in place by each organization [21]. Very often, the implementation of cross-organizational business processes requires systems that allow federated identity management. Indeed, in such processes, there are administrative domains of different partner organizations that need to interact with each other, and all this, in some way, requires that the partners trust each other [52].

Blockchain [48] is a data storage technology that originated from Bitcoin, a peer-to-peer cryptocurrency [46] that realizes block synchronization through peer-to-peer transmission technology and consensus algorithm, ensuring the data consistency of each member node in the network. The tampering resistance of the data registered in the Blockchain network against external attacks has been proven to be efficient [40]. The data state is read or changed through transactions assembled and packaged into blocks under a specific structure in a Blockchain network. Each block keeps the previous block's hash value, so if any block's hash value is changed, the entire chain will be invalidated. Depending on the level of trust between nodes, Blockchains can be divided into a *public chain*, *consortium chain*, and *private chain*. The nodes of a private chain all belong to the same organization and are fully trusted. The nodes of a consortium chain belong to different organizations that trust each other, and lastly, all nodes of a public chain do not trust each other. Hyperledger Fabric is an open-source Blockchain platform that can be used to implement consortium chain networks [5], and realizes all characteristics of Blockchain, including decentralization, irreversibility, consensus, identity authentication, smart contract, and others. Compared with other Blockchain platforms, Hyperledger Fabric provides higher throughput, a more effective consensus mechanism, a channel isolation mechanism, a multi-chain mechanism, and flexible expansion capability. Several studies use Blockchain as the underlying data platform to solve information-sharing problems among project participating organizations [34–36]. In particular, Liao et al. have proposed a Blockchain-based cross-organizational integrated platform, called *BCOIP* [37], which enables to issue and redeem of reward points. Lu et al. use Blockchain technology to store users' access control lists. In this way, thanks to its tamper-proof and decentralized features, Blockchain technology allows the creation of cross-organizational authentication systems where organizations can share data and resources between them [44]. Again, Fridgen et al. show how Blockchain can be a viable solution to achieve secure cross-organizational workflow management [13]. In particular, Blockchain in business process management allows improving the auditability and automation of manual processes through a decentralized system. Furthermore, it is essential to underline that the development and deployment of Blockchain-based systems for cross-organizational workflows management cannot ignore the legal regulations regarding data processing, such as the *General Data Protection Regulation (GDPR)* in force in Europe [15]. However, most of the studies proposed in the literature are based on domain-specific implementations and do not provide a generic management tool that project managers can reuse.

In this paper, we propose a general-purpose project management system, referred to as *Fabric-GC*, realized using Blockchain as a data-sharing platform. More precisely, the proposed system uses the Gantt chart model to manage the entire project allocation and execution progress, besides visually providing such relevant information to project managers. Again, *Fabric-GC* applies Blockchain technology so that project data can be shared safely and efficiently among multiple organizations, facilitating cross-organizational project collaboration. In detail, *Fabric-GC*, which represents the first Gantt chart management system for cross-organizational project management, is based on hyperledger fabric. The consortium chain is selected as the underlying storage model for the system proposed. The main contributions of this article are as follows:

- 1) Blockchain and Gantt chart are the building blocks of *Fabric-GC*. The proposed solution enables the migration of the traditional Gantt chart model from a centralized to

a distributed architecture to provide visual expression. Besides, the Blockchain is also tackled to deal with the secure storage and sharing of data, where smart contracts define the structure and operation of data in a project.

2) The proposed solution referred to as *Fabric-GC* aims at dividing the entire project into multiple chunks of small tasks. The project manager defines the project plan and assigns such chunks to different organizations in task schedules; then, it uses smart contracts to specify the read and write operations on the project plan. The proposed solution effectively improves the flexibility of project cooperation and guarantees versatile project management.

3) The proposed solution enables the visualization of task schedules as a Gantt chart, besides providing a progress feedback mechanism that assists project managers in grasping the project completion status and making real-time adjustments to the project plan.

4) Experimental results show that *Fabric-GC* has stable performance and high production efficiency under different consensus mechanisms.

The remaining of this article is organized as follows. Section 2 introduces the work related to this proposed research, Section 3 presents some preliminary concepts, including the data storage mechanism of Hyperledger fabric and structure of the Gantt chart. Section 4 introduces the system architecture, data structure, smart contract design, and workflow. Section 5 discusses the operation steps of the *Fabric-GC* system and shows the system's stability under different consensus mechanisms through comparative experiments. Finally, Section 6 summarizes our contributions and brings items as future work.

2. Related Work

Widely speaking, Blockchain technology enables the realization of decentralized, immutable, and incorruptible public ledgers [33]. Due to its ability to create smart contracts, Blockchain is perfectly suitable for project management, which phases include project creation, project allocation, project execution, and project acceptance. As known, the entire project cycle requires information sharing and oversight from multiple parties. In this context, the ability to access electronic data securely and efficiently enhances the ability to perform quality assurance-type projects. Therefore, the applicability of Blockchain in project management has been investigated by many researchers, as shown in Table 1, which summarizes these studies.

Table 1. Comparison with related work

Research	Application	Model	Generality	Visualization
[57], [19]	Construction Engineering	N	N	N
[6], [45]	Scientific Research Project Management	N	N	N
[20]	Supply Chain	N	N	Y
[42]	industrial Production	N	N	N
[24], [14]	Government Project Management	N	N	N
This paper	General Project Management System	Y	Y	Y

To address the issues of poor communication, weak file sharing privacy, and low-quality submission efficiency in projects construction, Yang et al. [57] analyzed the business processes of the public and private Blockchain in the construction industry and presented the challenges faced by the construction industry after applying Blockchain, aimed at improving the efficiency and productivity of construction projects. In addition, Hargaden et al. [19] proposed to apply Blockchain to sizeable structural engineering projects. They concluded that incorporating Blockchain improves efficiency, trust, transparency, and regulation in the construction industry effectively. However, the above works proposed in the literature do not introduce a specific system model to address the multi-party project management problem.

Scientific and engineering project works also need strict regulation and monitoring to reduce human communication and supervision costs [23]. Bai et al. [6] proposed a Blockchain-based *scientific research project management system (SRPMS)* and analyzed the five functional modules of the proposed model. Meng et al. [45] used consortium Blockchain and *IPFS (InterPlanetary File System)* technology to realize a reliable and efficient scientific research project management system that overcomes the limitations on the breach of contract and confidentiality in project management, also reducing the time and labor cost for the project implementation. Helo et al. [20] applied Blockchain in the supply chain to solve the delivery problem of multi-supplier participation by ensuring real-time tracking, control of data, and real-time visibility of all the processes in the project production process under the control of a project manager. Liu et al. [42] used Blockchain to manage the life cycle of products in the industrial production process, enabling the coordinating production information across departments and partners, quickly and accurately tracking the production and sales process, improving interoperability and collaboration among stakeholders in the product chain. To cope with several government-supported projects, Lee et al. [24] proposed a generic project sharing platform that achieves project information sharing while ensuring the platform's anti-forgery with the help of *POA (Proof-of-Authority)* consensus algorithm. Lastly, Green [14] showed that the adoption of Blockchain in the digital management of government projects could significantly improve workload and productivity, besides improving the strategic decision-making of the government.

The abovementioned issues show that Blockchain technology can be applied to the project management process to improve many aspects such as collaboration capability, information security, and real-time tracking functions of project implementation in a multi-organizational cooperation mode, to enhance the project completion efficiency. Several studies indicate that decomposing large projects into multiple small task schedules and sequencing the execution of task sets in a time series can achieve rational resource planning, besides saving time and labor costs [7, 22, 25, 43, 47, 49]. Similarly, Blockchain is applied in ensuring secure data storage in areas such as online education, finance, Internet of Thing (IoT), healthcare, and Vehicular ad-hoc network (VANET) [10, 11, 17, 18, 26–28, 30–32, 39, 41, 51, 58]. However, the current strategies have not saved project costs from the details of rational planning projects, nor providing sufficient simulation experiments to demonstrate performance sustainable performance under large-scale and multi-harmonic tasks.

Thus, in this paper, we proposed a generic distributed management tool for project managers to adequately handle the management and coordination of decentralised, complex or large projects.

3. Preliminaries

This section presents the background and related methods for the system's design and implementation to give further details of the proposed system. The notation used in this article is outlined in Table 2.

Table 2. The descriptions of notations

Notations	Description
u_i	External user i
p	Represent a project
t_j	j -th task scheduling of p
P	Project list
T^n	A set of n tasks
bPT	Start time of the project
ePT	End time of the project
bT	Start time of the task
eT	End time of the task
cT	Completed time of the task
uN	User name
tN	Task name
pN	Project name
PI	Index of user and projects

3.1. Hyperledger Fabric

Hyperledger Fabric is used to build enterprise-level consortium Blockchain and realize data sharing among multiple organizations to collaborate to form Blockchain networks. As an open-source project, Hyperledger Fabric has been started by the Linux Foundation and maintained by several corporate organizations. Basically, Hyperledger Fabric is characterized by a modular design concept. It has a sophisticated tiered policy structure, where each fabric component is extensible and mainly includes identity authentication, consensus module, intelligent contract, data storage. Each component is a container, so it is high the flexibility to build the fabric network. The entire system runs in the docker container. The container separates the running environment from the hardware environment as a sandbox environment to achieve total data confidentiality and security. The protocol used for the secure channel is TLS (Transport Layer Security). TLS/SSL is a specification for an encrypted channel that uses symmetric encryption, public-private key asymmetric encryption. Finally, all nodes in the fabric network need authentication and authorization. These requirements enable the meeting of the characteristics of mutual trust among members of the consortium Blockchain.

Main Components There are three most important types of nodes in Hyperledger Fabric: CA, Orderer, and Peer.

CA: In fabric networks, the identity certificate is required for communication. Without loss of generality, we assume that external users intend to communicate with one of the nodes. The CA acts as a trusted entity and holds the public keys of all users, but the algorithm for generating public and private key pairs for user registration is executed locally. In that case, it needs to be registered by the administrator at the CA node to generate a unique digital certificate and key for data transmission.

Orderer: It is mainly responsible for data consensus, and all validated transactions are submitted to the Orderer node for sorting. Next, the Orderer node packages the transactions into blocks according to the predefined rules (block out time, block size, the maximum number of transactions, etc.) and then sends them to the Peer node. A consensus algorithm maintains the consistency of data, in which consensus mechanisms provided by the fabric are *Solo*, *Kafka*, and *Raft*.

Peer: It is a data storage node, either for Blockchain state or block data. In addition, it has the function of validating transactions by simulating the execution of chaincodes to verify the legitimacy of transactions, i.e., endorsement. Only legitimate transactions are submitted to the Orderer node waiting to be packaged into blocks. Lastly, their modifications on the state are written to the Blockchain.

Chaincode. Chaincode is the smart contract of fabric and is implemented mainly using the Go programming language. It is the interface of Blockchain to the external environment. By calling the methods defined by chaincode, the external environment can execute operations such as data storage, indexing, or modification to the Blockchain, which functionally is similar to SQL language in relational database [9]. Developers writing different chaincode programs can achieve different application functions.

Ledger. There are two types of data on fabric, the world state and block. As shown in Fig 1, external data d_i is packaged as a transaction operation Tx_i by calling the SDK [1–4]. Then, the transaction writes d_i to the world state by calling the method f_i in the smart contract and is stored to the state database in the form of $\langle k - v \rangle$.

The legal Tx_i will be submitted to the Orderer node, waiting to be packaged into blocks and stored permanently by the Peer node.

3.2. Gantt Chart

Gantt chart is a management tool for planning and project arrangement proposed by Henry Gantt [7], widely used in many fields, such as educational activities, software development [49], technology transfer [25], production plant scheduling [22], and several others. Gantt chart shows graphically the project plan, which can be handy to track the task scheduling in each period.

The horizontal axis of a Gantt chart represents time, and the vertical axis represents task scheduling. For a project p , it can be divided into n small task schedules based on time, resources, manpower, etc., and $p = \{t_1, t_2, \dots, t_n\}$. If these tasks are scheduled to be executed only in time order, the total execution time of a project can be characterized as follows.

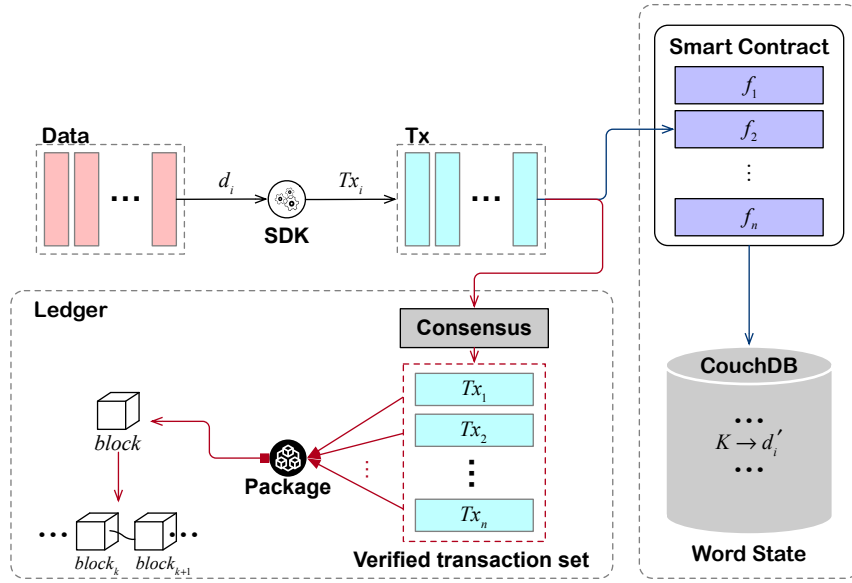


Fig. 1. Structure of the ledger in Hyperledger Fabric

$$\begin{aligned}
 \Delta t &= p.ePT - p.bPT \\
 &= (t_1.eT - t_1.bT) + (t_2.eT - t_2.bT) + \\
 &\quad \dots + (t_n.eT - t_n.bT) \\
 &= \sum_{i=1}^n t_i.eT - t_i.bT
 \end{aligned} \tag{1}$$

In the case we analyze the key execution order $\{t_{s1}, t_{s2}, \dots, t_{sk}\}$ in the task set [25], where $s1 \leq s2 \leq \dots \leq sk$ and $\{s1, s2, \dots, sk\} \subset [1, n]$, then, the overall project execution time is given as follows.

$$\begin{aligned}
 \Delta t_{theory} &= (t_{s1}.eT - t_{s1}.bT) + (t_{s2}.eT - t_{s2}.bT) + \\
 &\quad \dots + (t_{sk}.eT - t_{sk}.bT) \\
 &= \sum_{i=1}^k t_{si}.eT - t_{si}.bT, \quad 1 \leq k \leq n
 \end{aligned} \tag{2}$$

Gantt chart can visualize the execution relationship between each task schedule. Due to such, project managers utilize the Gantt chart to plan and adjust the project execution. In this way, they can potentially more accessible estimate the project cost, evaluate the project deadline, and achieve or approach the theoretical time cost Δt_{theory} .

Although the Gantt chart achieves excellent performance in project planning, most of the current Gantt chart systems in the market utilize a centralized model, as in Fig 2. Therefore, when multiple organizations are involved in a project, problems such as lag-

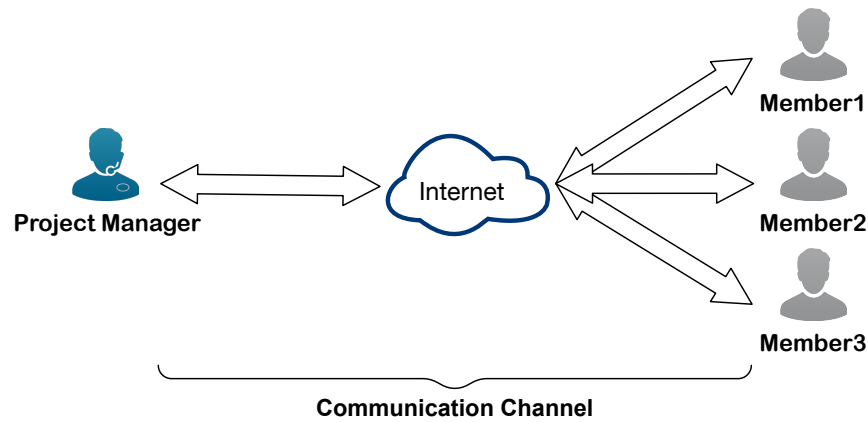


Fig. 2. The centralization mode of information interaction

ging news and untimely feedback inevitably occur. The unsynchronized information may cause management moil and bring severe negative impact to projects.

4. System Architecture and Design

It is introduced in this section the architecture of Fabric-GC. The overall design architecture is discussed first, followed by the data structure defined in Fabric-GC. Next, the design method of the smart contract, and lastly, the workflow of Fabric-GC.

4.1. System Architecture

The Blockchain-based Gantt chart system proposed in this work consists of three parts: *consortium Blockchain*, *server layer*, and *user layer*. The functions of Fabric-GC include permission to participants from different organizations to join the system, project plan sharing in the form of Gantt charts, and feedback from project members on the project execution progress.

Consortium Blockchain: It is a distributed network composed of nodes representing different organizations for global data synchronization and storage. The nodes in the consortium are mutually trusted. More precisely, they realize identity verification through digital certificates to ensure the security and integrity of data in the system. The smart contract running on it regulates the various steps in project management and stores the project data in Ledger for permanent storage.

Server layer: It contains servers maintained by each organization, interacts with a Blockchain system and smart contracts, and provides an endpoint interface to project members of the same organization. Project data is removed from the Blockchain and converted into meaningful project plans at that layer and visualized as Gantt charts provided to the project manager.

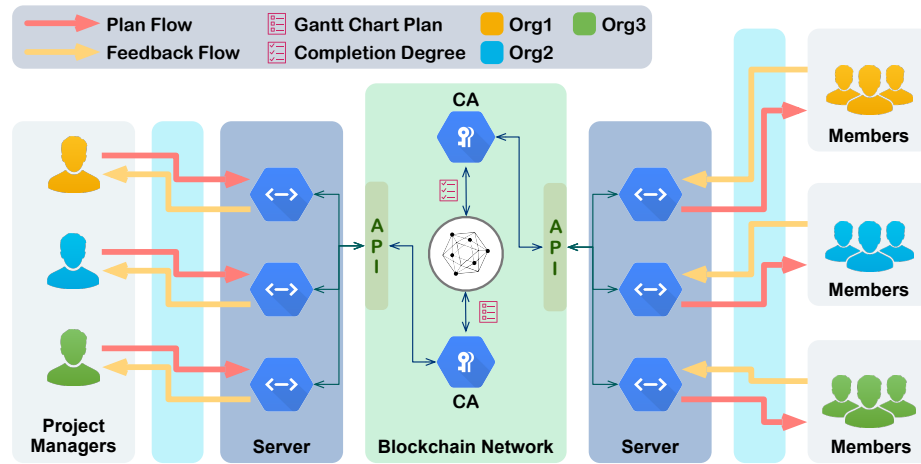


Fig. 3. Architecture of fabric-GC

User layer: It contains all project members and includes project managers and project participants. Project managers are responsible for the planning, procurement, and execution of a project, in any undertaking that has a defined scope. Project participants follow the project arrangement and feedback on the project progress. However, in Fabric-GC, the user identity is not distinguished, and thus, the design eliminates the organizational differences and realizes the conversion of logical identity.

As shown in Fig 3, there are two different data streams in Fabric-GC: *Plan Flow* and *Feedback Flow*. According to the existing project resources, the project manager seeks to achieve defined goals by using plans, schedules the project execution, and then draws the Gantt chart. This chart is then submitted to the Blockchain system through Plan Flow, so other project participants can obtain specific project plans from the Blockchain and complete the assigned tasks of the project according to the project arrangement and schedule. Any modification in the project execution processing will notify project members in time. When the project members conduct the project, they submit the completed progress through Feedback Flow, also feedback it to the project manager through the Blockchain system. Upon receiving such relevant updated information, the project manager makes appropriate adjustments to the plan according to the progress. In this way, closed-loop data exchange is formed to realize the dynamic management of projects across organizations.

4.2. Data Structure

The world state of Blockchain is similar to table data in a relational database. The data structure of the state is analogous to the table structure [9]. Five data structures are defined in this article, listed as *Project*, *Task*, *ProjectIndex*, *TaskIndex*, and *User*. As shown in Fig 4, there is an index relationship between the data structures, and depicted in Eq. (3). We remark that defining the data structures in this way facilitates uniform data access operations and reduces the system's complexity.

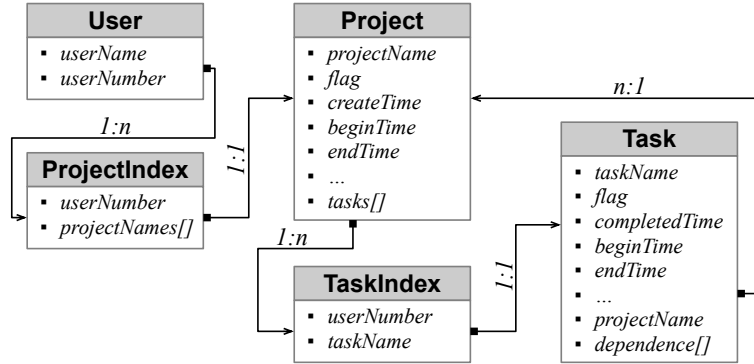


Fig. 4. Relationship between data structures

$$\begin{cases} User \rightarrow ProjectIndex & = 1 : n \\ ProjectIndex \rightarrow Project & = 1 : 1 \\ Project \rightarrow TaskIndex & = 1 : n \\ TaskIndex \rightarrow Task & = 1 : 1 \\ Task \rightarrow Project & = n : 1 \end{cases} \quad (3)$$

User This structure defines a project member in the Blockchain system, and only project members added to the Blockchain can be authorized to participate in relevant projects. The User structure contains two fields, $u_i = \{userName, userNumber\}$, where $userName$ represents the user name used by all participants to identify member i , and $userNumber$ denotes the unique identification of member i when data is stored. Also, to enhance the uniqueness and randomness of the $userNumber$, we introduce a timestamp so that the $userNumber$ can be calculated as calculated by Eq. (4).

$$userNumber = Hex[Hash(pubKey_i, Timestamp)] \quad (4)$$

Project This structure holds the properties of the project, obtained from pN . The essential attributes of this structure are as follows.

i) $flag$: records the status of the current project. $processing$ indicates that the project is in progress, and $done$ indicates that the project is completed. The default value is $processing$.

$$flag = \begin{cases} processing \\ done \end{cases} \quad (5)$$

ii) $beginTime(bPT)$: Project start time used to limit the left interval of task scheduling.

iii) $endTime(ePT)$: Project end time, used to limit the right range of task scheduling.

iv) $tasks$: The collection of all tasks for this project.

Task. This structure represents the data structure of task scheduling, which stores the attributes of task scheduling and is indexed by tN . The task set of p is expressed as T^n , then p expressed according to Eq. (6).

$$p = T^n = \{t_1, t_2, \dots, t_n\} \quad (6)$$

Additionally, the essential attributes of the task structure are as follows.

i) *flag*: Record the status of this task schedule. From Eq. (6), $p.flag = done$ is equivalent to

$$\forall t_j \in T^n, t_j.flag = done \quad (7)$$

ii) *beginTime(bT)*: The start time of the task schedule. The value range of this attribute is $bT \in [bPT, ePT]$.

iii) *endTime(eT)*: The end time of the task schedule. The value range of this attribute is $eT \in (bPT, ePT]$ and $bT < eT$. At the same time, it must satisfy the following equation:

$$\begin{aligned} t_n.eT - t_1.bT &\leq p.ePT - p.bPT \\ t &\in T^n \end{aligned} \quad (8)$$

iv) *completedTime(cT)*: The completion degree of the current task schedule. It is noted that, if and only if $cT = eT$, $flag = done$ holds.

v) *dependence*: In actual task scheduling, the start of a task may require completing other tasks, so this property saves the set of dependent tasks for the current task.

ProjectIndex. Fabric-GC allows multiple projects to co-exist in the system. A member can participate in multiple projects, so the structure defines two properties: *userNumber* and *projectNames*. The former identifies a member, and the latter records the name of each project the member participates in.

TaskIndex. The structure is saved in the *tasks* of the Project. Two attributes are defined, where *userNumber* represents the member responsible for scheduling the task, and *taskName* records the name of the task schedule and can index the entire task scheduling data.

4.3. Smart Contract Design

The contract part mainly defines data access operations and relies on the conventional MVC (Model-view-controller) software design pattern [53]. This part avoids using excessive business processing logic to reduce functional redundancy and improve the system's scalability. In this research, the following methods are defined to access the data corresponding to members, projects, and task scheduling, respectively.

Member Data Access. There are two ways to register members in the contract: *createUser()* and *queryUser()*. Again, to ensure the uniqueness of the identity of participants belonging to multiple organizations, a member is stored in the state database of the Blockchain. Meanwhile, before creating and indexing a member, the system checks whether the currently created member has already been stored in the Blockchain system.

There are four methods related to project data in the contract. listed as:

1. *createProject()* writes the data of the project to the state database. Before creating the project, we must check whether the project already exists. After the new project is successfully created, a ProjectIndex should be established between the creator and the project.
2. *queryProIndex()* takes the member as the input to obtain the project name set participated by the member.
3. *queryProject()* indexes the data of the project through the project name.
4. *changeProject()* is used to index the project data and modify the data of the current created project to realize the flexibility of data access.

Task Data Access. There are three contract methods related to task scheduling data, listed as:

1. *assignTask()* allocates task scheduling for the specified project. After successful creation, the TaskIndex needs to be saved to the *tasks* of the project. Then, the ProjectIndex is established for the members responsible for the task.
2. *queryTask()* method obtains the specific task scheduling data through the task name.
3. *changeTask()* can modify the specified task scheduling information, as shown in Algorithm 1. The modification of scheduling can be divided into two categories: when the value of *target* is "changeInfo", it indicates that only the data of the current task scheduling needs to be modified. When the value of *target* is "changeManager", the current project leader needs to be modified. At this point, the information of the task scheduling and the project attribute and project index related to the task must be modified.

4.4. Workflow

The system operation is divided into three logical parts to enable the Blockchain-based Gantt chart (i.e., Fabric-GC). Aimed to realize the cross-organizational project management function, it consists of participant login, project creation, and task scheduling allocation. In the following, we provide the details of the three parts abovementioned.

The complete participant login processing can be divided into three parts: administrator registration, project member registration, and project member login. This process can realize the storage of personnel information among different organizations on the Blockchain and solve relatively weak information exchange among members of different organizations. In detail, each organization has a Fabric CA node used to store each member's ID, private key, certificate, and other information. Before registering a member, we need to register an administrator user to connect to the Fabric CA node. The steps to carry out this operation are as follows.

Algorithm 1: changeTask

Input: tN , $target$, $taskData$
Output:

```

1 if  $target == "changeInfo"$  then
2   DelState( $tN$ );
3   PutState( $tN$ , $taskData$ );
4   return;
5 else
6   if  $target == "changeManager"$  then
7     oldTData = GetState( $tN$ );
8     project = GetState(oldData.projectName);
9     while  $t$  in  $project.tasks$  do
10      if  $t.taskName == tN$  then
11        t.userNumber =  $taskData.manager$ ;
12    PutState(oldData.projectName, $project$ );
13    createProjectIndex( $taskData.manager$ , oldData.projectName);
14    PutState( $tN$ , $taskData$ );
15    return;

```

1. The server creates a *Wallet* locally,
2. sets the administrator's name and password and sends it to the fabric CA node of the organization through the SDK,
3. Fabric CA node generates unique $signID$, $privKey_{admin}$, $pubKey_{admin}$ and certificate for the administrator,
4. sends these to the server and saves them in the created *Wallet* as permanent storage.

Notably, Administrators do not need to be stored on the Blockchain to distinguish different organization members and reduce unnecessary data conflicts. Algorithms 2 and 3 show how to register and log u_i into Fabric-GC. When the members are registered, u_i needs to provide the uN and the organization's name to which it belongs. Then, according to the organization name, the server selects the corresponding organization's SDK to call. Before a member is registered, it is necessary to check whether it has been registered and whether uN already exists in the CA node. If the member is not registered, the server logs into the Fabric CA node through the administrator account and password, then the CA node registers u_i and generates the field $\{signID, pubkey_i, privKey_i, certificate\}$, and then sent to the server. Finally, the server calculates the $usernumber$ by using the Eq. (4) and executes the $createUser()$ method of the smart contract to invoke the member u_i into the Blockchain state database, completing the registration process of the project members.

During the member login process, the uN and the organization name are also provided by u_i . The server will first determine whether the member exists. If the result provided by the query locates such a member, u_i can connect to the Blockchain network through the public key $pubKey_i$ and private key $privKey_i$, and then query the member data by calling the $queryUser()$ method of the smart contract. If successful, the data will be sent to the server, and the server responds with the login results to u_i .

Algorithm 2: Register u_i in the system.

Input: $u_i.userName, u_i.Org$ **Output:**

```

1 // Register the  $u_i$ .
2 Send  $u_i.userName, u_i.Org$  to the server;
3 Call the SDK specified by  $u_i.Org$ ;
4 if  $!isExists(u_i.userName)$  then
5   | ca  $\leftarrow$  connect({admin, adminpw});
6   | CA generates  $\{signID, pubKey_i, privKey_i, certificate\}$ ;
7   | CA sends  $\{signID, pubKey_i, privKey_i, certificate\}$  to the server;
8   | The server save them to the wallet;
9   |  $u_i.userNumber \leftarrow$  Hex[md5(pubKey_i)];
10  | SDK.createUser( $u_i.userName, u_i.userNumber$ );
11  | return;
12 else
13  | return 'An identity for  $u_i$  already exists in the wallet.';

```

Algorithm 3: Log in u_i in the system

Input: $u_i.userName, u_i.Org$ **Output:**

```

1 // Log in  $u_i$ .
2 Send  $u_i.userName, u_i.Org$  to the server;
3 Call the SDK specified by  $u_i.Org$ ;
4 if  $isExists(u_i.userName)$  then
5   | Get  $\{privKey_i, pubKey_i\}$  from the wallet by  $\{u_i.userName, admin\}$ ;
6   | Connect to fabric network by  $\{privKey_i, pubKey_i\}$ .
7   |  $\{u_i.userName, u_i.userNumber\} \leftarrow$  SDK.queryUser( $u_i.userName$ );
8   | return;
9 else
10  | return 'An identity for  $u_i$  does not exists in the wallet.';

```

Project Creation. Solely when the project is created in the Fabric-GC, and relevant attributes of the project are specified, the project manager can assign task scheduling for the project and draw a Gantt chart. The project creation process of the system is outlined in Fig 5.

Step 1: Member u_i accesses the server and sends $\{userName_i, Org_i\}$ to the server's 'process_login' module to requests system login. After the server responds to the successful login, u_i sends the project creation request to the "create_project" module and submits the project-specific attribute values.

Step 2: The server generates Project data structure p . Next, it sets $p.tasks = null$ and checks whether the value of $p.flag$ is *processing*. If the data format meets the requirements, the chaincode *createProject()* is called through the SDK specified by Org_i , and the parameter $\{userNumber_i, JSON(p)\}$ is passed in. Besides, if *isExists(p)* is false, it

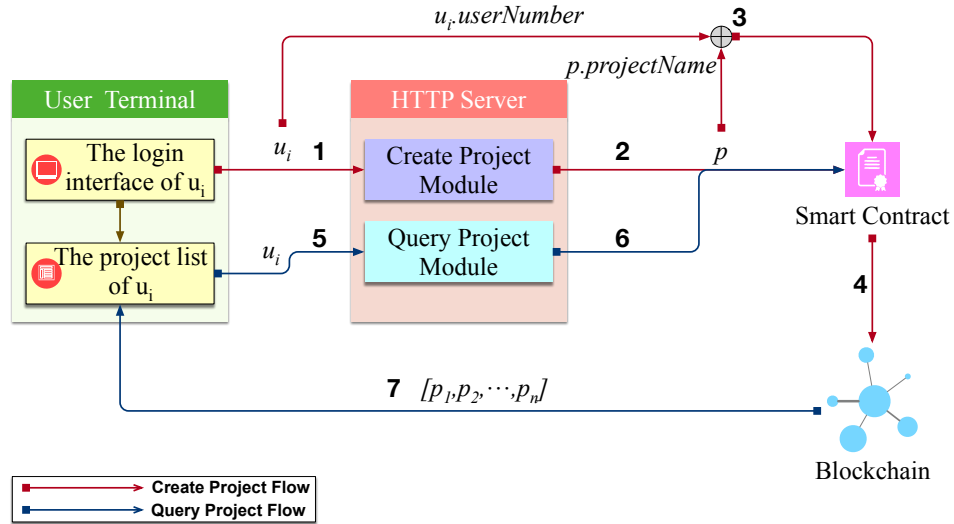


Fig. 5. The description of project creation process.

means that the project has not been created before and $\text{PutState}(\{p.\text{projectName}, p\})$ is called to create the project.

Step 3: After the project is created successfully, the index PI of $userNumber_i$ and $p.\text{projectName}$ is created by $\text{createProIndex}()$.

Step 4: Project p and index PI are stored in the state database of Blockchain.

Step 5: Member u_i sends a request to the server to query the list of projects that participates in.

Step 6: After receiving the request, the server executes the "query_project" module and indexes the data based on Eq. (9).

$$\begin{aligned}
 u_i &\xrightarrow{uN_i} \{u_i, [pN_1, pN_2, \dots, pN_n]\} \\
 &\xrightarrow{pN_j} [p_1, p_2, \dots, p_n], \quad j = 1, 2, \dots, n
 \end{aligned} \tag{9}$$

The chaincode $\text{queryProIndex}()$ is called by SDK specified by Org_i , and the list of project names $PN = [pN_1, pN_2, \dots, pN_n]$ is obtained. After the query is successful, the server iteration PN , and the chaincode $\text{queryProject}(queryProject(pN_j))$, $j = 1, 2, \dots, n$ is called to obtain all the list of projects $P = [p_1, p_2, \dots, p_n]$.

Step 7: The project list P is sent to the client terminal for visualization by member u_i .

Task Schedule Allocation. The task schedule allocation is the most critical part of the Fabric-GC proposed system. The main characteristic of this part is that it enables to achieve sharing and dynamicity. Besides, the underlying Blockchain also ensures the integrity and synchronization of data in the project management. As Fig 6 shows, task

schedule allocation is divided into four modules: data request, task scheduling allocation, completion modification, and Gantt chart visualization. A detailed description of these four modules is provided next.

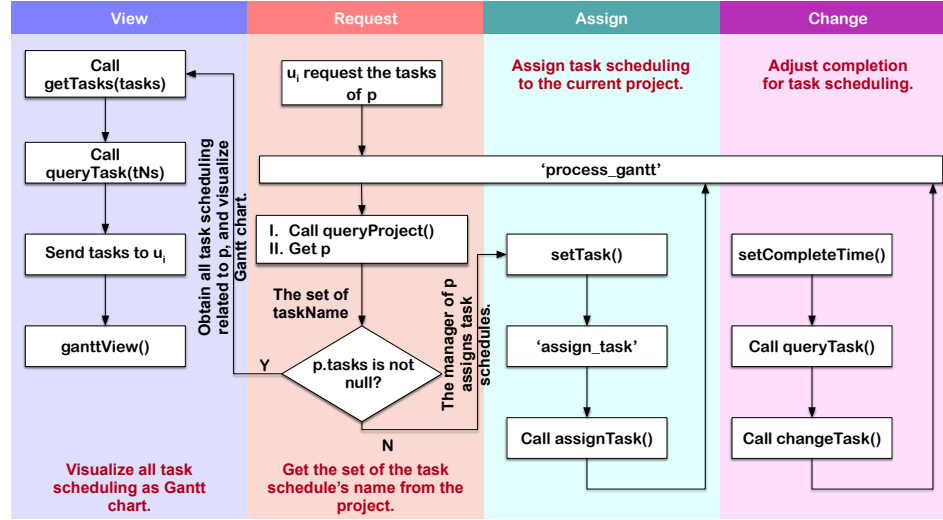


Fig. 6. The flow of task schedule allocation process

Request: As soon as the member u_i retrieves the list of projects it participates in, a project p is selected and requests next the tasks from the server. Afterward, the server runs the "process_gantt" module by calling the `queryProject()` method of the chaincode through the SDK to obtain the task name set $p.tasks$ of p . If $p.tasks$ is *null*, it means that the current project p has no tasks assigned yet, and the project leader needs to make a reasonable task scheduling assignment.

Assign: The task scheduling process is assigned by the person in charge of the project p . As shown in Algorithm 4, manager u_i sends the task data to the server, and the server runs the "assign_task" module. The assign task scheduling sets the value of *completedTime* as *null* and attribute *projectName* as the name of project p to construct task data structure t_j . Then, it retrieves t_j 's dependent task set, T' , through SDK and attribute *dependence*. When checking the legitimacy of t_j , the assign task scheduling first checks whether the *flag* is *processing*, then checks whether the start and end times of t_j are within the scope of p , and finally checks whether t_j conforms to the dependency set T' . If t_j is legal, the `assignTask()` method of the chaincode is called to write t_j to the state database and return the execution result to the client. Finally, if the execution is successful, the client requests the server to execute the "process_gantt" module again.

Change: The system proposed Fabric-GC not only provides project participants with shared task scheduling data for the entire project but also provides shared completion of the project during execution. Algorithm 5 outlines the process of adjusting the completion of a specified task scheduling. The client member u_i specifies the

Algorithm 4: Assign task scheduling to p

Input: tN , manager, bT , eT , $flag$, info, dependence
Output:

- 1 The server runs the 'assign_task' module;
- 2 $cT = \text{null}$; // The task does not begin.
- 3 $pN = p.\text{projectName}$;
- 4 $T' \leftarrow \text{Call SDK.queryTask}(\text{dependence})$;
- 5 **if** $flag \neq \text{'processing'}$ **then**
- 6 **return** err ;
- 7 $t_j \leftarrow \{tN, \text{manager}, bT, eT, flag, \text{info}, \text{dependence}, cT, pN\}$;
- 8 **if** $t_j \notin T_i^n$ — t_j does not depend on T' **then**
- 9 **return** err ;
- 10 $res \leftarrow \text{Call SDK.assignTask}(t_j)$;
- 11 **if** res is successful **then**
- 12 Send the sign of success to the client, and the client requests the 'process_gantt' module again.

$\{taskName, completedTime\}$ to send to the server, which runs the 'setCompletedTime' module. Next, the SDK obtains the task scheduling t_j and the project p to which t_j belongs. If u_i is not the person in charge of t_j and does not belong to the manager of p , this means that u_i does not have the permission to modify t_j and the system returns a permission error. Before any modification, the system ensures that the completion degree $completedTime$ falls within the starting and ending range of t_j . After setting the completion degree of t_j , if $completedTime == t_j.endTime$, the system sets $t_j.done = done$ and checks whether $\forall t. flag = done, t \in p$ holds and the entire project has been completed. After the successful execution, the server will return the execution result to the client, and the client will request the server to execute the "process_gantt" module again.

View: This module visualizes all task scheduling sets as a Gantt chart. Besides, after getting all the task name sets $tasks$ of p , this module calls the $getTasks()$ method provided by the server. Through iterating $tasks$, the chaincode $queryTask()$ method is called to retrieve the entire task collection. Finally, this module calls the $gantView()$ to visualize the task set as a Gantt chart.

5. Experiment and Comparison

This section describes the experimental process and the results achieved by evaluating the functions and performance of the proposed Fabric-GC system. The creation process of the system and the implementation of cross-organizational project management based on the system are depicted and followed by comparative experiments and performance results and analysis.

5.1. Design, Implementation, and Evaluation of fabric-GC

The experiments are divided into two parts. The former describes the network structure and project structure of Fabric-GC. At the same time, the latter introduces the operation

Algorithm 5: Adjust completion for t_j

Input: tN, cT, u_i
Output:

- 1 The server runs the 'setCompletedTime' module;
- 2 $t_j \leftarrow$ Call SDK.queryTask(tN);
- 3 $p \leftarrow$ Call SDK.queryProject(t_j .projectName);
- 4 **if** t_j .manager \neq u_i — p .manager \neq u_i **then**
- 5 | **return** *err*;
- 6 **if** $cT \notin (t_j$.beginTime, t_j .endTime] **then**
- 7 | **return** *err*;
- 8 t_j .completedTime \leftarrow cT ;
- 9 **if** $cT == t_j$.endTime **then**
- 10 | t_j .flag = 'done';
- 11 | **if** t .flag == done, $\forall t \in p$ **then**
- 12 | | p .flag = 'done';
- 13 $res \leftarrow$ Call SDK.changeTask(tN , 'changeInfo', t_j);
- 14 **if** res is successful **then**
- 15 | Send the sign of success to the client, and the client requests the 'process_gantt'
 | module again.

steps of Fabric-GC, including how to create projects in a multi-organization environment, task scheduling, and Gantt chart visualization.

There are six docker containers in Fabric-GC runtime that constitute the Blockchain network, as shown in Table 3.

Table 3. Fabirc-GC nodes

Node name	Description	Number
fabric-couchdb	database node	4
fabric-ca	CA node	2
fabric-peer	peer node	4
fabric-orderer	orderer node	1
fabric-tools	cli node	1
fabric-gantt/chaincode	chaincode node	2

The complete project structure consists of four parts.

1. **bin:** Binary tool directory. It is mainly used to generate certificates, block configuration, channel configuration, and other files.
2. **chaincode:** The directory where the chaincode is stored.
3. **client:** The main directory of the project. Save the network startup script, chaincode installation script, server-side source code.
4. **network:** Network configuration file directory. It includes a docker container configuration file, block configuration file, and certificate generation file.

The project initialization steps are as follows.

Step 1: Run the script code, *start.sh*. This script first removes the volume nodes and data that have been started. Next, it generates the certificate file, block configuration, channel configuration, and other files required by the startup container. The container is started, the channel initialized, and each peer node is added to the channel. Lastly, the chaincode is installed and initialized.

Step 2: Create administrator accounts for *Org1* and *Org2*, and account information not stored in the Blockchain.

Step 3: Start the server process to receive client requests.

Evaluation of Fabric-GC in the multi-organizational environment. This experiment assumes that the project manager is *user1* and belongs to organization *Org1*. The project *project1* is divided into six tasks assigned to different participants, where *user2* and *user3* belong to organization *Org1*, while *user4*, *user5*, *user6* and *user7* belong to organization *Org2*.

Table 4. The task set of *project1*.

Task Name	Principal	Organization	beginTime	endTime
task1	user2	Org1	2020.11.15	2020.11.28
task2	user3	Org1	2020.11.29	2020.12.05
task3	user4	Org2	2020.12.06	2020.12.10
task4	user5	Org2	2020.12.11	2020.12.15
task5	user6	Org2	2020.11.29	2020.12.10
task6	user7	Org2	2020.12.16	2020.12.31

Each project member is registered in the system through its terminal. *user1* logs into the system and creates project *project1* at first hand, and then assign task scheduling for members according to Table 4. After that, each member logs into Fabric-GC. After successful login, the project list of all projects that the member participates in is displayed. Project *project1* is queried by *user4*, indicating that data sharing is successful.

The Gantt chart represented by *project1* is shown in Fig 7, where the blue bar chart represents the planned duration, while the gray bar chart represents the completed duration. The status of a project implementation can be seen from the graph. That is, the Blockchain system ensures that all members participating in the project can obtain the latest status information.

To assign task scheduling to *project1*, as shown in Fig 8, *user1* can select the 'assign' button to pop up the dialog box and fill in the information of the task to be assigned. When *user3* needs to feedback the completion progress of *task2*, click the 'completed' button, as shown in Fig 9, and specify the task name and completed time. After *user1* receives the feedback information, he can adjust the project in real-time according to the completion status and repeat the above process so that the project Manager *user1* can always grasp the project's overall implementation to achieve the goal of the project saving resource cost and improving execution efficiency.

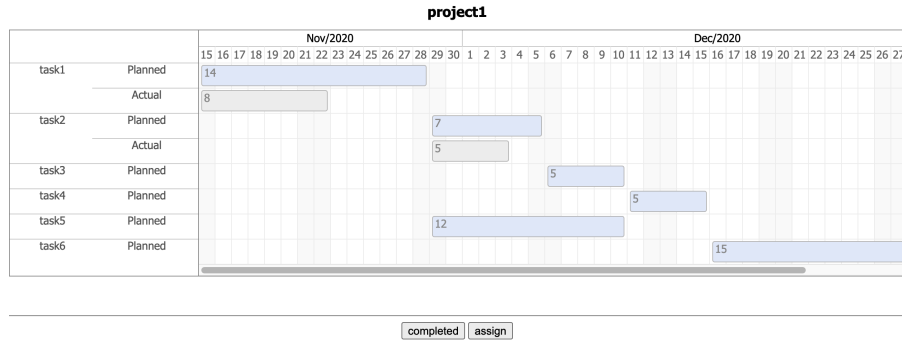


Fig. 7. The interface of Gantt chart

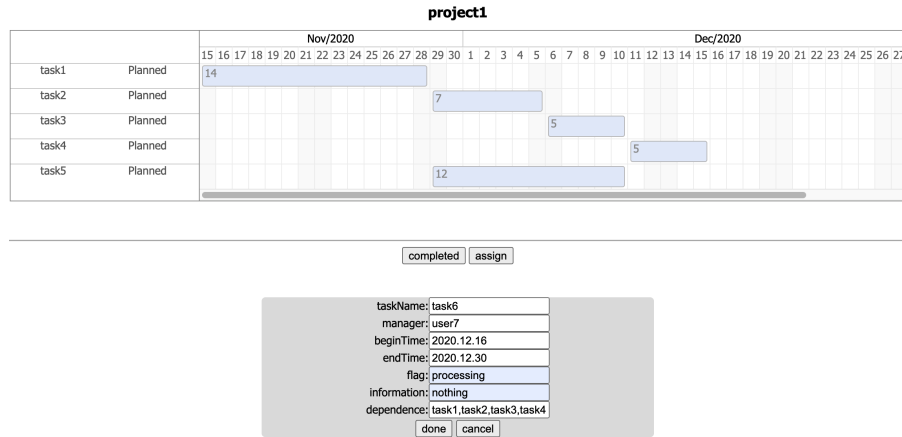


Fig. 8. Assign a task schedule

5.2. Result and Comparison

To test the performance of the Fabric-GC, we compare the execution time when the project data is stored and choose *tape* [16] for the TPS (Transaction Per Second) throughput testing of the chaincode. We remark that the performance bottleneck of Blockchain is mainly in the consensus mechanism, and different consensus mechanisms impact the data synchronization rate between nodes. Since the nodes of the public chain do not trust each other, the *PoW (Proof-of-Work)* consensus algorithm is required to achieve data synchronization. In this case, the nodes' arithmetic power is used to mine for packing rights, and its execution is inefficient. As shown in Fig 10, the execution time of PoW impacts more than 10 seconds on the write operations *createUser()*, *createProject()*, and *assignTask()*. On the other hand, if the nodes of the consortium chain trust each other, the data synchronization time is about 2.5 seconds under the consensus algorithm (*Solo, Kafka, and Raft*).

Several methods with high request in Fabric-GC use read operations, e.g., *queryUser()*, *queryProject()*, *queryTask()* and *changeTask()*. Fig 11(a)-11(d) show TPS under

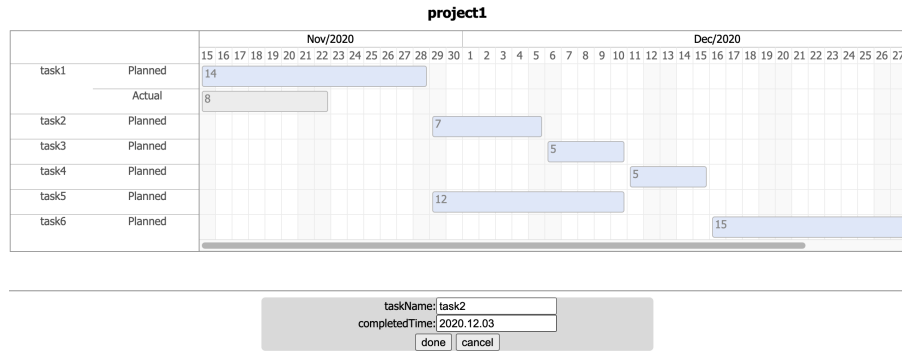


Fig. 9. The process of completion setting

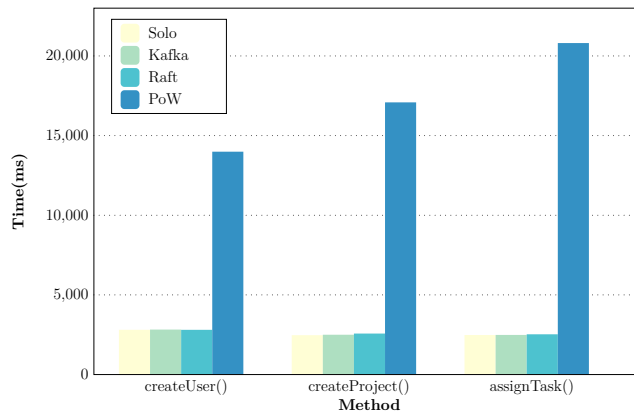


Fig. 10. Execution time of write operations under Solo, Kafka, Raft and PoW

different *Transactions*. Under the same number of transactions, the TPS of the four methods has little difference. This aspect shows that Fabric processes various transactions in the same process. In addition, we can observe that, when *Transactions* = 100, and the number of requests is more than 100, the throughput of the system can reach ranging 400 to 500. Besides, when *Transactions* = 5, the throughput of the system is around 50. The smaller the number of transactions, the more blocks generated in the same time period. The system consumes too many resources for the packaging and verification of blocks.

Fig 12 shows the throughput curve of *queryProject()* method under different consensus algorithms when the block size is set to 100MB. More precisely, with the Solo and Kafka consensus mechanism, the difference between their TPS is not significant. When the number of requests exceeds 400, the system throughput can be kept from 400 and 500. The throughput of Raft is lower than the other two. Therefore, Kafka's consensus should be selected in the production environment as far as possible to ensure high performance and high fault tolerance.

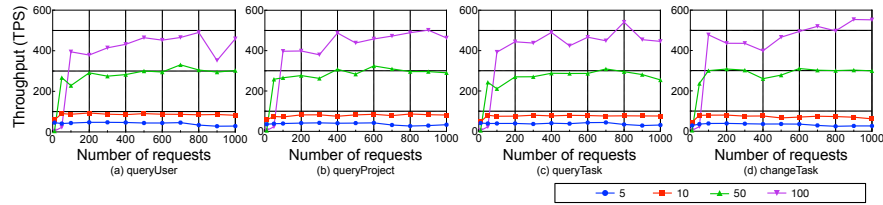


Fig. 11. Throughput under different *Transactions*. *Transactions* represents the number of transactions in a block

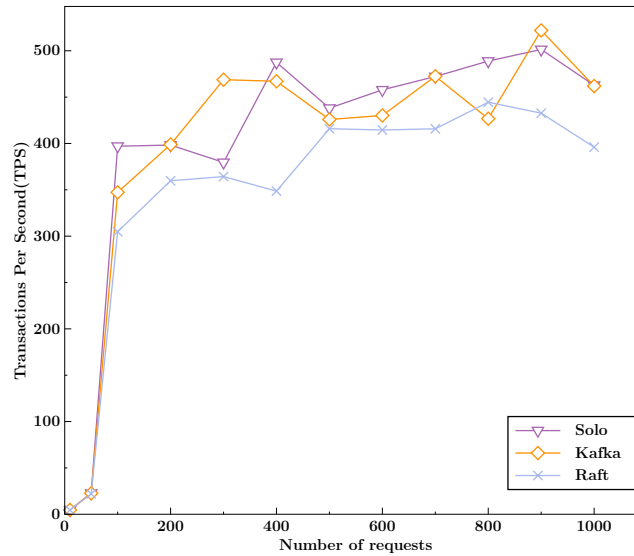


Fig. 12. Throughput of *queryProject()* under different consensus algorithms when *Transactions* = 100

The comparative experiments show that fabric-GC can maintain high throughput under large-scale requests and adapt to Blockchain networks under different consensus algorithms with relatively stable performance output.

6. Conclusions and Future Work

Multi-party project cooperation is a standard practice, widely used in scientific R & D, industrial production, software development, supply chain [8] among several other fields. Indeed, the collaboration between organizations and individuals with different technologies improves the rate of the success of complex projects [50]. Nevertheless, cross-organizational projects pose difficulties for project managers in managing task scheduling and progress feedback that relies on timely information sharing [57]. The independence and heterogeneity among participating organizations can make data sharing dif-

difficult. Moreover, since traditional data sharing relies on third-party organizations (e.g., cloud, specialized service provider, transcription services, call center services, consulting), the privacy and security of data cannot be guaranteed [56].

In this article, we propose a Blockchain-based Gantt chart system, named *Fabric-GC*. The proposed system mitigates the difficulties arising from human resource management and information transfer in multi-organizational project cooperation scenarios. In this proposed research, the Blockchain eliminates the heterogeneity between different partners, enabling them to maintain and manage the same project jointly. In detail, with the support of smart contracts, the project manager can communicate the Gantt chart schedule to the participants across the organization through *Fabric-GC*. Therefore, the participants can achieve real-time feedback on the project progress, and the project manager can make timely adjustments to the project schedule. Experimental results show that the proposed system can deal with large-scale data request scenarios while maintaining stable performance under different consensus mechanisms.

As future research directions, we intend to work on the following list of items:

1. The performance testing and evaluation of *Fabric-GC* have been conducted in a single machine environment. As future work, we plan to consider a distributed environment containing multiple nodes for testing and verifying the performance of our proposal;
2. In project management, not only time cost needs to be considered, but also resource allocation, among other issues and costs. Additional features in this regard are under investigation and will be included in future extensions of *Fabric-GC*;
3. *Fabric-GC* cannot store a large amount of data. Interconnecting *Fabric-GC* with distributed storage systems such as the *Interplanetary File System (IPFS)* will be considered to overcome this limitation. In this way, *Fabric-GC* will also enable the sharing of project-related resources, including files, video, audio, and other resources.
4. The management of participants will be enhanced to realize the evaluation of participants' capabilities. In this way, project managers can make more effective and reasonable project planning and execution.

Acknowledgments. This research is supported by the National Natural Science Foundation of China under Grant 61873160, Grant 61672338, and the Natural Science Foundation of Shanghai under Grant 21ZR1426500.

Declaration of interests. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. fabric-sdk-go (2020), <https://github.com/hyperledger/fabric-sdk-go>
2. fabric-sdk-java (2020), <https://github.com/hyperledger/fabric-sdk-java>
3. fabric-sdk-node (2020), <https://github.com/hyperledger/fabric-sdk-node>
4. fabric-sdk-py (2020), <https://github.com/hyperledger/fabric-sdk-py>
5. Androulaki, E., Barger, A., Bortnikov, V., etc.: Hyperledger fabric: A distributed operating system for permissioned blockchains. In: Proceedings of the Thirteenth EuroSys Conference. EuroSys '18, Association for Computing Machinery, New York, NY, USA (2018)

6. Bai, Y., Li, Z., Wu, K., Yang, J., Liang, S., Ouyang, B., Chen, Z., Wang, J.: Researchchain: Union blockchain based scientific research project management system. In: 2018 Chinese Automation Congress (CAC). pp. 4206–4209 (2018)
7. Bednjanec, A., Tretnjak, M.F.: Application of gantt charts in the educational process. In: 2013 36th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). pp. 631–635 (2013)
8. Chang, S.E., Chen, Y.: When blockchain meets supply chain: A systematic literature review on current development and potential applications. *IEEE Access* 8, 62478–62494 (2020)
9. Chitti, P., Murkin, J., Chitchyan, R.: Data management: Relational vs blockchain databases. In: Proper, H.A., Stirna, J. (eds.) *Advanced Information Systems Engineering Workshops*. pp. 189–200. Springer International Publishing, Cham (2019)
10. Cui, M., Han, D., Wang, J.: An efficient and safe road condition monitoring authentication scheme based on fog computing. *IEEE Internet of Things Journal* 6(5), 9076–9084 (2019)
11. Cui, M., Han, D., Wang, J., Li, K.C., Chang, C.C.: Arfv: an efficient shared data auditing scheme supporting revocation for fog-assisted vehicular ad-hoc networks. *IEEE Transactions on Vehicular Technology* 69(12), 15815–15827 (2020)
12. Fang Hua, Chen Tian, Xie Ying, Sun Yu: Order planning and scheduling of rod and wire production based on gantt chart. In: *Proceeding of the 11th World Congress on Intelligent Control and Automation*. pp. 3417–3421 (2014)
13. Fridgen, G., Radszuwill, S., Urbach, N., Utz, L.: Cross-organizational workflow management using blockchain technology: Towards applicability, auditability, and automation. In: *51st Annual Hawaii International Conference on System Sciences (HICSS)* (2018)
14. Green, S.: A digital start-up project – carm tool as an innovative approach to digital government transformation. *Computer Systems Science and Engineering* 35(4), 257–269 (2020)
15. Guggenmos, F., Lockl, J., Rieger, A., Wenninger, A., Fridgen, G.: How to develop a gdpr-compliant blockchain solution for cross-organizational workflow management: Evidence from the german asylum procedure. In: *Proceedings of the 53rd Hawaii International Conference on System Sciences* (2020)
16. Guo, J.: A simple traffic generator for hyperledger fabric (2020), <https://github.com/guoger/tape>
17. Han, D., Pan, N., Li, K.C.: A traceable and revocable ciphertext-policy attribute-based encryption scheme based on privacy protection. *IEEE Transactions on Dependable and Secure Computing* (2020)
18. Han, D., Zhu, Y., Li, D., Liang, W., Souri, A., Li, K.C.: A blockchain-based auditable access control system for private data in service-centric iot environments. *IEEE Transactions on Industrial Informatics* (2021)
19. Hargaden, V., Papakostas, N., Newell, A., Khavia, A., Scanlon, A.: The role of blockchain technologies in construction engineering project management. In: *2019 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*. pp. 1–6 (2019)
20. Helo, P., Shamsuzzoha, A.: Real-time supply chain—a blockchain architecture for project deliveries. *Robotics and Computer-Integrated Manufacturing* 63, 101909 (2020)
21. Hilia, M., Chibani, A., Amirat, Y., Djouani, K.: Cross-organizational cooperation framework for security management in ubiquitous computing environment. In: *2011 IEEE 23rd International Conference on Tools with Artificial Intelligence*. pp. 464–471. IEEE (2011)
22. Jia, H., Fuh, J., Nee, A., Zhang, Y.: Integration of genetic algorithm and gantt chart for job shop scheduling in distributed manufacturing systems. *Computers & Industrial Engineering* 53(2), 313–320 (2007)
23. Jiang, Y., Liang, W., Tang, J., Zhou, H., Li, K., Gaudiot, J.: A novel data representation framework based on nonnegative manifold regularisation. *Connection Science* 33(2), 136–152 (2021)
24. Lee, E., Yoon, Y., Lee, G., Um, T.: Blockchain-based perfect sharing project platform based on the proof of atomicity consensus algorithm. *Tehnicki Vjesnik*

25. Lee, S., Shvetsova, O.A.: Optimization of the technology transfer process using gantt charts and critical path analysis flow diagrams: Case study of the korean automobile industry. *Processes* 7(12) (2019)
26. Li, D., Han, D., Crespi, N., Minerva, R., Li, K.C.: A blockchain-based secure storage and access control scheme for supply chain finance. *The Journal of Supercomputing* pp. 1–30 (2022)
27. Li, D., Han, D., Crespi, N., Minerva, R., Sun, Z.: Fabric-scf: A blockchain-based secure storage and access control scheme for supply chain finance. *arXiv preprint arXiv:2111.13538* (2021)
28. Li, D., Han, D., Liu, H.: Fabric-chain & chain: A blockchain-based electronic document system for supply chain finance. In: *International Conference on Blockchain and Trustworthy Systems*. pp. 601–608. Springer, Singapore (2020)
29. Li, D., Han, D., Weng, T.H., Zheng, Z., Li, H., Liu, H., Castiglione, A., Li, K.C.: Blockchain for federated learning toward secure distributed machine learning systems: a systemic survey. *Soft Computing* pp. 1–18 (2021)
30. Li, D., Han, D., Zheng, Z., Weng, T.H., Li, H., Liu, H., Castiglione, A., Li, K.C.: Moocschain: A blockchain-based secure storage and sharing scheme for moocs learning. *Computer Standards & Interfaces* 81, 103597 (2022)
31. Li, H., Han, D., Tang, M.: A privacy-preserving charging scheme for electric vehicles using blockchain and fog computing. *IEEE Systems Journal* 15(3), 3189–3200 (2020)
32. Li, H., Han, D., Tang, M.: A privacy-preserving storage scheme for logistics data with assistance of blockchain. *IEEE Internet of Things Journal* (2021)
33. Liang, W., Fan, Y., Li, K., Zhang, D., Gaudiot, J.: Secure data storage and recovery in industrial blockchain network environments. *IEEE Transactions on Industrial Informatics* 16(10), 6543–6552 (2020)
34. Liang, W., Tang, M., Long, J., Peng, X., Xu, J., K.Li: A secure fabric blockchain-based data transmission technique for industrial internet-of-things. *IEEE Transactions on Industrial Informatics* 15(6), 3582–3592 (2019)
35. Liang, W., Xiao, L., Zhang, K., Tang, M., He, D., Li, K.: Data fusion approach for collaborative anomaly intrusion detection in blockchain-based systems. *IEEE Internet of Things Journal*
36. Liang, W., Zhang, D., Xia, L., Tang, M., Li, K., Zomaya, A.: Circuit copyright blockchain: Blockchain-based homomorphic encryption for ip circuit protection. *IEEE Transactions on Emerging Topics in Computing*
37. Liao, C.H., Teng, Y.W., Yuan, S.M.: Blockchain-based cross-organizational integrated platform for issuing and redeeming reward points. In: *Proceedings of the Tenth International Symposium on Information and Communication Technology*. pp. 407–411 (2019)
38. Liu, C., Zeng, Q., Cheng, L., Duan, H., Zhou, M., Cheng, J.: Privacy-preserving behavioral correctness verification of cross-organizational workflow with task synchronization patterns. *IEEE Transactions on Automation Science and Engineering* (2020)
39. Liu, H., Han, D., Li, D.: Blockchain based trust management in vehicular networks. In: *International Conference on Blockchain and Trustworthy Systems*. pp. 333–346. Springer (2020)
40. Liu, H., Han, D., Li, D.: Fabric-iot: A blockchain-based access control system in iot. *IEEE Access* 8, 18207–18218 (2020)
41. Liu, H., Han, D., Li, D.: Behavior analysis and blockchain based trust management in vanets. *Journal of Parallel and Distributed Computing* 151, 61–69 (2021)
42. Liu, X., Wang, W., Guo, H., Barenji, A.V., Li, Z., Huang, G.Q.: Industrial blockchain based framework for product lifecycle management in industry 4.0. *Robotics and Computer-Integrated Manufacturing* 63, 101897 (2020)
43. Liu, Y.C., Gao, H.M., Yang, S.M., Chuang, C.Y.: Application of genetic algorithm and fuzzy gantt chart to project scheduling with resource constraints. In: Huang, D.S., Jo, K.H., Wang, L. (eds.) *Intelligent Computing Methodologies*. pp. 241–252. Springer International Publishing, Cham (2014)

44. Lu, P.J., Yeh, L.Y., Huang, J.L.: An privacy-preserving cross-organizational authentication/authorization/accounting system using blockchain technology. In: 2018 IEEE International Conference on Communications (ICC). pp. 1–6. IEEE (2018)
45. Meng, Q., Sun, R.: Towards secure and efficient scientific research project management using consortium blockchain. *Journal of Signal Processing Systems* (Apr 2020)
46. Nakamoto, S.: Bitcoin: A peer-to-peer electronic cash system (2008), <http://www.bitcoin.org/bitcoin.pdf>
47. Nurre, S.G., Weir, J.D.: Interactive excel-based gantt chart schedule builder. *INFORMS Transactions on Education* 17(2), 49–57 (2017)
48. Satoshi, N.: Bitcoin—open source p2p money (Nov 2019), <https://bitcoin.org/en/>
49. Seniv, M., Sambir, A., Seniv, M.: Working hours controls methods and increasing its efficiency in the it company. In: 2016 XII International Conference on Perspective Technologies and Methods in MEMS Design (MEMSTECH). pp. 235–238 (2016)
50. Skowron, P., Rzacca, K., Datta, A.: Cooperation and competition when bidding for complex projects: Centralized and decentralized perspectives. *IEEE Intelligent Systems* 32(1), 17–23 (2017)
51. Sun, Z., Han, D., Li, D., Wang, X., Chang, C.C., Wu, Z.: A blockchain-based secure storage scheme for medical information. *EURASIP Journal on Wireless Communications and Networking* 2022(1), 1–25 (2022)
52. Thurm, B., Hu, J.: Automated creation and realization of security federation for cross-organizational business processes. In: 2008 IEEE Symposium on Advanced Management of Information for Globalized Enterprises (AMIGE). pp. 1–5. IEEE (2008)
53. Voorhees, D.P.: Model–View–Controller: TUI Versus GUI, pp. 297–304. Springer International Publishing, Cham (2020)
54. Warnier, M., Lukosch, S., Heutelbeck, D.: Intellectual property management in cross-organizational collaboration. In: Workshop on Security and Privacy in Collaborative Working, Cardiff, Sept. 13-16, 2010, authors version (2010)
55. Xiao, T., Han, D., He, J., Li, K., de Mello, R.: Multi-keyword ranked search based on mapping set matching in cloud ciphertext storage system. *Connection Science* 33(1), 95–112 (2021)
56. Yang, P., Xiong, N., Ren, J.: Data security and privacy protection for cloud storage: A survey. *IEEE Access* 8, 131723–131740 (2020)
57. Yang, R., Wakefield, R., Lyu, S., Jayasuriya, S., Han, F., Yi, X., Yang, X., Amarasinghe, G., Chen, S.: Public and private blockchain in construction business process and information integration. *Automation in Construction* 118, 103276 (2020)
58. Zhijie, S., Han, D., Li, D., Wang, X., Chang, C.C., Wu, Z.: A blockchain-based secure storage scheme for medical information (2022)
59. Zhu, H., Liu, H., Ou, C.X., Davison, R.M., Yang, Z.: Privacy preserving mechanisms for optimizing cross-organizational collaborative decisions based on the karmarkar algorithm. *Information Systems* 72, 205–217 (2017)

Dun Li received the B.S. degree in Human Resource Management from the Huaqiao University, Quanzhou, China, in 2013, and the M.S. degree in Finance from the Macau University of Science and Technology, Macau, China, in 2015. He is currently doing his Ph.D. degree in Information Management and Information Systems at Shanghai Maritime University. His research interests mainly include smart finance, big data, machine learning, IoT, and blockchain.

Dezhi Han received the BS degree from Hefei University of Technology, Hefei, China, the MS degree and PhD degree from Huazhong University of Science and Technology,

Wuhan, China. He is currently a professor of computer science and engineering at Shanghai Maritime University. His specific interests include storage architecture, blockchain technology, cloud computing security and cloud storage security technology.

Benhui Xia received the B.S. degree from China University of Mining and Technology, where he is currently pursuing the M.S. degree with Shanghai Maritime University. His main research interests include network security, cloud computing, distributed computing and blockchain.

Tien-Hsiung Weng is currently a professor at the Department of Computer Science and Information Engineering at Providence University, Taichung, Taiwan. He received a Ph.D. in Computer Science from the University of Houston, USA. His research interests include parallel programming models, performance measurement, and compiler analysis for code improvement.

Arcangelo Castiglione received a Ph.D. degree in Computer Science from the University of Salerno, Italy. He is a tenure-track assistant professor at the Department of Computer Science, University of Salerno (Italy). His research mainly focuses on cryptography, network security, data protection, digital watermarking, and automotive security. He is an Associate Editor for several Scopus-Indexed journals, and he has been Guest Editor for several Special Issues and Volume Editor for Lecture Notes in Computer Science (Springer). He has been involved in several organizational roles (steering committee member, program chair, publicity chair, etc.) for many international conferences. He has been a reviewer for several top-ranked scientific journals and conferences. He has been appointed as a member of the IEEE Technical Committee on Secure and Dependable Measurement. He is a founding member of the IEEE TEMS Technical Committee (TC) on blockchain and Distributed Ledger Technologies.

Kuan-Ching Li is currently appointed as a professor in the Dept. of Computer Science and Information Engineering (CSIE) at Providence University, Taiwan, where he also serves as the Director of the High-Performance Computing and Networking Center. Besides the publication of articles in renowned journals and conferences, he is co-author or co-editor of more than 40 books published by Taylor & Francis, Springer, IGI Global and McGraw-Hill. He is a Fellow of IET and a senior member of the IEEE. Professor Li's research interests include parallel and distributed computing, Big Data, and emerging technologies.

Received: November 05, 2021; Accepted: June 25, 2022.

Efficient Generative Transfer Learning Framework for the Detection of COVID-19

J. Bhuvana, T. T. Mirnalinee, B. Bharathi, and Infant Sneha

Sri Sivasubramaniya Nadar College of Engineering, Chennai, India
{bhuvanaj, mirnalineett, bharathib}@ssn.edu.in
infantsneha17059@cse.ssn.edu.in

Abstract. Deep learning plays a major role in detecting the presence of Coronavirus 2019 (COVID-19) and demands huge data. Availability of annotated data is a hurdle in using Deep learning technique. To enhance the accuracy of detection Deep Convolutional Generative Adversarial Network (DCGAN) is used to generate synthetic data. Densenet-201 is identified as the deep learning framework to detect COVID-19 from X-ray images. In this research, to validate the effectiveness of the Densenet-201, we explored conventional machine learning approaches such as SVM, Random Forest and Convolutional Neural Network (CNN). The feature map for training the machine learning approaches are extracted using Densenet-201 as feature extractor. The results show that Densenet-201 as feature representation with SVM is performing well in detecting COVID-19 with high accuracy. Moreover we experimented the proposed methodology without using DCGAN as well. DenseNet-201 based approach is capable of detecting the presence of COVID-19 with high accuracy. Experiments demonstrated that the proposed transfer learning approach based on DenseNet-201 along with DCGAN based augmentation outperforms the State of the art approaches like ResNet50, CNN, and VGG-16.

Keywords: COVID-19, Densenet-201, DCGAN, Disease Classification, Data Augmentation, Deep learning.

1. Introduction

The first case of novel Coronavirus disease (COVID-19) was reported in Wuhan, China at the end of December 2019. COVID-19 became an epidemic all over the World [11]. This is a respiratory disease caused by a severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The common symptoms of COVID-19 are fever, cough, short breathing, sore throat, headache, and diarrhea [27]. Vanishing of taste, tiredness, aches, loss of smell, and nasal blockade are the additional symptoms that also have been observed in patients.

Initially, Real-Time Reverse Transcription Polymerase Chain Reaction (RT-PCR) is the only technique to detect the COVID-19 from respiratory samplings [33]. RT-PCR is an effective method for the diagnosis of SARS-CoV-2. The main drawbacks of RT-PCR are time-consuming and error-prone results [24]. This method needs a laboratory kit, the provision of which is difficult or even impossible for many countries during crises and epidemics. Like all diagnostic and laboratory methods in healthcare systems, this method isn't error-free and is biased. It requires an expert laboratory technician to sample the nasal and throat mucosa which may be a painful method, and this is often why many

of us refuse to undergo nasal swab sampling. Due to RT-PCR's limited availability and drawbacks, it has posed challenges to prevent the dissemination of coronavirus infection.

In contrast to this, radiological imaging techniques are used for the diagnosis of SARS-CoV-2 by coalescing with the infected person's clinical symptoms, travel history, and laboratory findings. Radiological imaging such as chest X-rays and chest CT scans can be helpful to isolate the infected persons timely and control this epidemic situation. The first choice of radiologists is chest X-ray as most of the hospitals are equipped with X-ray machines [6].

Deep learning (DL) techniques are widely utilized in the automated analysis of radiological images. These algorithms can enhance the accuracy of classifying different types of knowledge [10]. One of the most common uses of DL in radiology was the detection of tissue-skeletal anomalies and, as a result, disease classification. The convolutional neural network has proven to be one of the foremost important DL algorithms and therefore the best technique in detecting abnormalities and pathologies in chest radiographs [18]. DL techniques can train the weights of networks on large datasets as well as fine-tuning the weights of pre-trained networks on small datasets. Hence, the main aim of this project is to use the pre-trained deep learning architectures as an automated tool to detect and diagnose COVID-19 in chest x-ray images.

Our contributions towards COVID-19 detection are :

1. Proposed an automated system that detects COVID-19 from chest X-ray images.
2. The accuracy of detecting COVID 19 is enhanced by using Deep learning generative model, DCGAN.
3. By simulation study identified a best computational model to detect COVID -19 with an F1 score of 96.99%.
4. The statistical significance of the model is evaluated using paired t-test.

Review of literature is presented in section 2, followed by the proposed methodology in section 3. Experiments covering the implementation details of the various deep learning models, discussion on their performance and comparison of the best performing model with existing techniques are given in section 4. Section 5 concludes the work with future directions.

2. Literature survey

Image classification refers to the task of classifying images into various categories. Image classification can be done by applying both machine learning [4] and deep learning algorithms. With the invention of deep learning, image classification has become more widespread. The deep learning model has a powerful learning ability [7],[21] [5], [15] which integrates the feature extraction and classification process. A pre-trained model is a deep learning model that was trained on a benchmark dataset to solve a problem similar to the one that we want to solve and one can re-use the pretrained model in many ways. Transfer learning based pre-trained models are used for image classification in variety of domains [13]. Authors have substantiated the importance of transfer learning and how the pretrained models can be customized for the custom image classification [8].

A large number of research work has been done to detect the COVID-19 from radiological imaging. Due to the ample availability of X-ray machines, disease diagnosis using

X-ray images are widely used by healthcare experts. In case of any suspect of COVID-19, instead of using test kits, an alternate way to detect pneumonia from the X-ray images is required, so that further investigation can be narrowed down for COVID-19 identification. And with the deep learning techniques, we can identify the COVID-19 patient effectively.

Jianpeng Zhang et al. [34] used the CAAD model which is composed of an anomaly detection network and confidence prediction network. They used the X-VIRAL dataset which consists of 5977 anomaly or positive (viral and non viral pneumonia) and 37,393 negative (healthy) cases. Narinder Singh Punn, Sonali Agarwal [25] used Random Over-sampling and weighted approach for data preprocessing and NasNet Large model for classification. They used the dataset of COVID-19 open dataset collection which contains 153 COVID and normal chest X-rays and Radiological Society of North America (RSNA) which contains 923 other pneumonia and normal chest X-rays and National Library of Medicine (NLM) with 138 Other pneumonia and normal chest X-rays. This model performs better in binary classification than multi-class classification. Mizuho Nishio et al. [22] used a combination of data augmentation (conventional and mixup) techniques and VGG-16 (Transfer learning) model for classification. They used the dataset of Github-COVID-19 chest X-rays and RSNA-other pneumonia and normal chest X-rays. In this model, the combination of two types of data augmentation methods was more effective than single type or no data augmentation methods. Marko Arsenovic et al. [28] used traditional data augmentation methods and ResNet (Transfer Learning-pre trained with grayscale images of ImageNet dataset). They used COVID-19 Chest X-ray (publicly available) dataset with 434 COVID-19 chest x-ray images and ChestXray2017 dataset with 2200 normal and other pneumonia chest x-ray images. Though this model is trained in a small dataset, it solved the problem of over fitting due to the dense blocks in ResNet architecture and gave good accuracy.

Terry Gao(2020) [9] used VGG-19 deep learning model for COVID-19 detection. Author have used 1600 publicly available chest x-rays (400 normal,800 other pneumonia, 400 COVID-19). Abbas et al. [2] used DeTraC-Class decomposition,Transfer learning (VGG-19), class composition method for COVID-19detection. They used the dataset of Japanese Society of Radiological Technology (JSRT) with 80 normal chest x-ray images and 105 COVID-19 chest xray images. This model has the ability to cope with data irregularity and the limited number of training images too.

Wang et al. [31] used CNN for COVID-19 detection. They used COVID-19 image data collection and RSNA with normal and other pneumonia chest x-ray images. Halgurd S. Maghdid et al. [20] used CNN for COVID-19 detection and used British Society of Thoracic Imaging (BSTI) and Github dataset with 85 COVID-19 chest x-ray images and Kaggle and Radiopedia dataset with 85 normal chest x-ray images. Sabbir Ahmed et al. [3] used CNN with a multi-level CNN based preprocessor for COVID-19 detection. They used Covidx dataset (Github) dataset and CheXpert (Github) and Kaggle dataset with other pneumonia and normal images. The elegance of the modular network is the preprocessor block that dynamically filters the input images for enhancing signs of COVID-19 infection, thereby making the task easier for the feature extraction and classification block placed in cascade with it. ReCoNet has identified further unwanted structures that were not useful for its learning and/or enhanced the regions that were important for COVID-19 detection. By incorporating different transformations, ReCoNet improved its performance accuracy. Ahmed Sedik et al. [26] used GAN with CNN for COVID-19 detection. They

used publicly available dataset. This model solves the problem of inadequate image data by using CGAN (Convolutional Generative Adversarial Network). This enhanced dataset was subsequently used to improve the learnability of the proposed deep learning models. Jaiswal A, Gianchandani N, Singh D, Kumar V, Kaur M. [14] used DenseNet-201 based deep transfer learning for COVID-19 detection, used dataset available in Kaggle. The authors conclude that DenseNet-201 based CNN performs significantly better as compared to some well-known deep transfer learning models.

Similar work is done to detect the COVID-19 images with three optimization algorithms to tune the hyper-parameters [30]. Grey Wolf Optimizer variation (GWO-TL), Differential Evolution variation (DE-TL) and Genetic Algorithm variation (GA-TL) were used to optimize the transfer learning hyper-parameters, among which the DE-TL has outperformed the other model proposed and also passed the Friedman test for statistical significance.

From the literature it is observed that deep learning algorithms are widely used for image classification task. Among the deep learning algorithms, CNN based custom defined networks and pretrained transfer learning networks are extensively applied for the classification. Our proposed system is designed with the novel combination of different algorithms towards the classification of COVID-19. The class imbalance problem is observed in the dataset, from the literature, it was inferred that Densenet-201 is the one of the best pretrained transfer learning approaches. The class imbalance problem is handled using DCGAN, followed by the feature extractor Densenet-201 and conventional SVM classifier.

Following this context, this work proposes to contribute for early diagnosis of COVID-19 using one of the state-of-the-art deep learning architectures like DenseNet-201 architecture and class imbalance problem is handled using DCGAN.

3. Proposed Methodology

Deep learning plays a major role in disease diagnosis from medical images. Due to the necessity of predicting the presence of COVID-19 from images accurately, we explored various deep learning techniques. The problem of detecting COVID-19 is formulated as a classification problem with three classes COVID-19, Pneumonia, Normal. In this proposed method an efficient learning strategy that combines generative models with deep learning for classification to compensate the lack of annotated data. Thus this proposal supports knowledge sharing by enabling high accuracy in classification.

When the dataset is analysed, it was observed that the samples in each class are not balanced. The class imbalance problem of the dataset has been a contributing factor for not providing a better performance. This motivated to perform data augmentation using a Generative Adversarial Network (GAN) which is Deep Convolutional Generative Adversarial Network (DCGAN) in specific. The augmented images added to the existing dataset and used for the further classification in this proposed work.

The deep learning algorithm chosen for classification is Densenet, which has four different variants such as DenseNet-121, DenseNet-169, DenseNet-201 and DenseNet-264. Among these Densenet-201 is chosen to detect the presence of COVID-19 in this work, since it is found to be the better classifier on medical images [32].

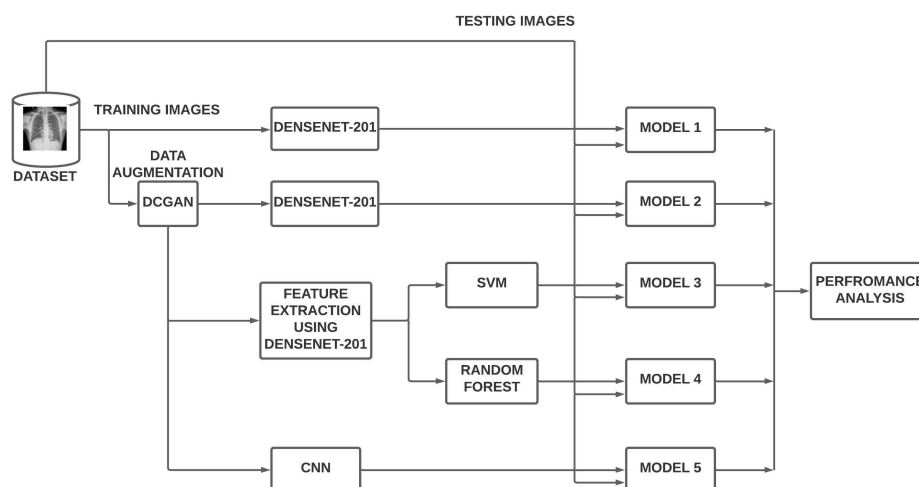


Fig. 1. Proposed architecture

To evaluate the effectiveness of DCGAN, the classifier DenseNet-201 is trained with the dataset without augmentation and termed as Model 1. To address the challenge of feature representation in conventional machine learning, we employed DenseNet-201 for feature extraction. SVM and RF classifiers are trained over the features extracted using DenseNet and the models are represented as Model 3, Model 4. On top of these we trained the pretrained deep learning models DenseNet-201 and CNN with the COVID-19 dataset augmented with DCGAN, whose models are termed as Model 2 and Model 5. The overall system architecture is illustrated in Fig.1.

The effectiveness of all the models are evaluated using the performance metrics and identified Model 3 with DenseNet-201 as feature extractor and SVM as classifier is the best model for detecting the presence of COVID-19 effectively. Following two subsections describes about the two deep learning building blocks used in our proposed work.

3.1. DCGAN

Conventionally, a GAN model consists of two stages: Generator and Discriminator. The generator network generates feature maps from the input images, while the discriminator network discriminates between the real and generated images using a classification layer. The GAN architecture is shown in Fig. 2.

However, GAN will have problems like instability during the training process. Compared with the simple GAN, DCGAN up-samples the images by using transposed convolution layer. Moreover, the leaky ReLU activation function is employed within the discriminator to stop gradient sparseness and no maxpooling is employed here in both generator and discriminator.

The generator phase of DCGAN consists of 5 convolutional transpose layers (Conv2D Transpose) - ConvTranspose1, ConvTranspose2, ConvTranspose3, ConvTranspose4, each consisting of 128 filters. The input images are first enrolled into a denoising dense layer

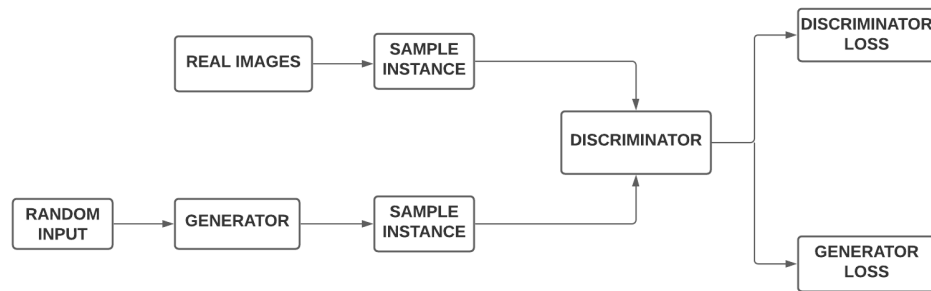


Fig. 2. Working of GAN [1]

that is primed at a size of $128 \times 8 \times 8$, following a sequence of Conv2D Transpose layers and Leaky RELU layers and in turn is followed by a Conv2D layer with tanh activation are applied to generate a feature map of the input images.

The discriminator phase of DCGAN consists of five convolutional layers (Conv2D) Conv1, Conv2, Conv3, Conv4, and Conv5 each consisting of 128, 64, 32, 16 and 8 filters respectively, and a sequence of LeakyRELU layers. Finally, the Dense layer in the discriminator is used to classify the real and fake data.

After data augmentation using DCGAN, the augmented dataset contains 4000 images in each class of the training set which solves the problem of imbalanced data.

3.2. DenseNet-201

DenseNet, which is a short form of Dense Convolutional Network, needs fewer number of parameters than a conventional CNN as it does not learn redundant feature maps. The layers in DenseNet are very narrow i.e., 12 filters, which add a lesser set of latest feature maps. A 5 layered dense block representing direct connections between layers is shown in Fig. 3. Each layer in DenseNet has direct access to the original input image and gradients from the loss function. Each layer receives collective knowledge from the previous set of layers. Therefore, the computational cost is significantly reduced, which makes DenseNet a far better choice for image classification.

For each Dense Block, Pre-Activation Batch Normalisation (BN), and ReLU, then 3×3 Convolutional layers are used. To reduce the model complexity and size, BN-ReLU- 1×1 Conv is completed before BN-ReLU- 3×3 Conv. 1×1 Conv followed by 2×2 average pooling is used because of the transition layers between two contiguous dense blocks. Feature map sizes are equivalent within the dense block in order that they will be concatenated together easily. At the top of the last dense block, a global average pooling is performed followed by classifier with a softmax activation.

Densenet-201 with and without data augmentation DenseNet-201, is trained with the dataset, the features extracted are given to one or more fully connected layers with a softmax activation in final layer. The extracted features are given to a Average Pooling layer followed by flattening, Dense layer and dropout layer. The softmax activation outputs the probability distribution over each possible class label. Here the Densenet-201 is used as

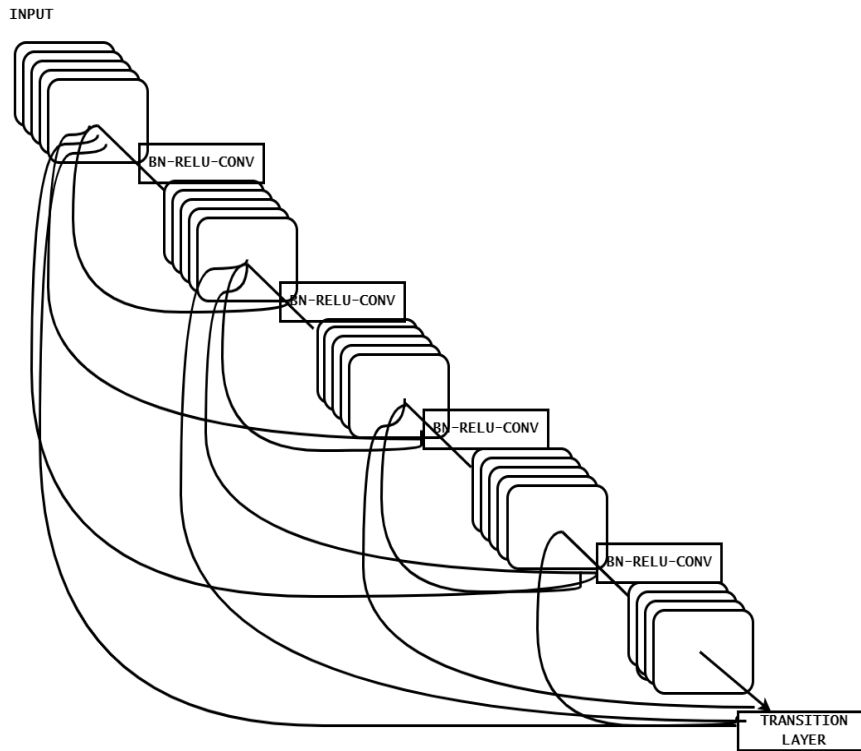


Fig. 3. A 5 layered dense block representing direct connections between layers [12]

the classifier trained with the images with and without augmentation in order to study the class imbalance problem among the classes of images in the dataset.

Support Vector Machine (SVM) SVM is used as a classifier that classifies the feature maps from DenseNet into COVID-19, Pneumonia, and normal. Tang (2013) [29] claims that training an SVM classifier on the features generated by the convolutional base can enhance classification accuracy. SVM has deployed linear kernel.

Random Forest (RF) Random forest is a supervised machine learning algorithm, ensemble by sequence of decision trees. Random forest classifier selects the best feature from the set of features which enhances the accuracy in prediction. Individual decision trees are trained in parallel and aggregates the decisions with good generalization without over fitting issue. Our architecture has used 100 trees with gini as the criteria with default minimum split as 2 and the maximum depth of tree is defined until all leaves are pure or until all leaves contain less than 2. The maximum number of features is taken as the default square root of the n_{features} and the there is no limit to the number of leaf nodes. A node in this architecture will split when it triggers a decrease in impurity below the default threshold 0.



(a) Covid 19 image of a patient [23]

(b) Pneumonia Chest image [23]

(c) Normal Chest image of a patient [23]

Fig. 4. Chest X- ray images

3.3. Simple CNN

CNN has been widely used as image classifier with a stack of convolutional layers, the network learns and extracts features used to discriminate images into COVID-19, Pneumonia and normal. In our proposed approach, CNN is applied after the images are augmented. CNN consists of an input Conv layer of 32 filters and 3 blocks of Conv layer with 3x3 kernel size, max pooling layer with 2x2 kernel size, and dropout layer which is followed by a flattening layer. The flattened feature vector is fed to a dense layer with 64 filters and a dropout layer which is connected to a final dense layer with a sigmoid activation function.

4. Experiments

4.1. Dataset

Chest X-ray dataset from Kaggle [23] having chest x-ray images of COVID-19, Pneumonia and normal patients is used for evaluating our system. It has 6,432 images in total where 80% is used for training and 20% is used for testing. In the training set, it has 460 COVID-19, 1266 normal and 3418 pneumonia images. In the testing set, it has 116 COVID-19, 317 normal and 855 pneumonia images. The sample images of COVID-19, Pneumonia and normal samples from the actual dataset before augmentation are shown in Fig. 4.

4.2. Implementation

The implementation was done in Python using Nvidia GPU system and the code is available in github.¹

¹ <https://github.com/infantsneha-s/COVID-19-Detection-using-Deep-Learning.git>

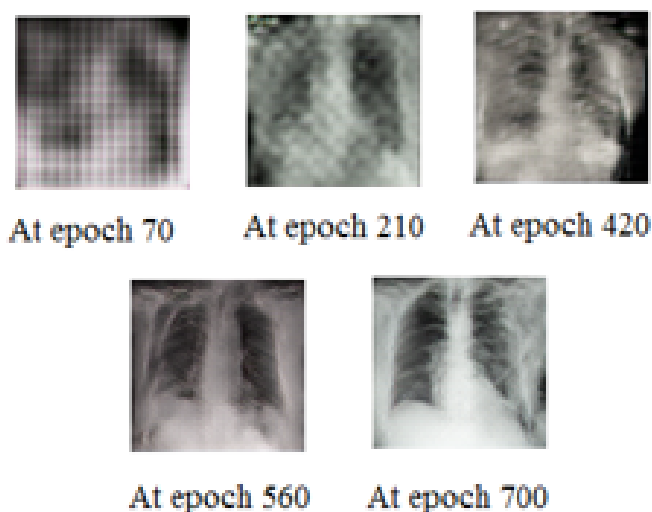


Fig. 5. Generated images of COVID-19 during the training process of DCGAN

Model 1: DenseNet-201 without augmentation Input dataset without augmentation is directly fed into the DenseNet-201 model for training which extracts features from the chest images. The features are then fed into its fully connected layer which classifies the images into COVID-19, pneumonia and normal. Model 1 is trained for 150 epochs and is evaluated. It has achieved a training accuracy of 97.31%.

As the dataset lacks in number of samples, generative models are used to enhance the dataset. DenseNet-201 framework is modified by enhancing the dataset with DCGAN based augmented images for the three classes. The images generated in different epochs is illustrated in Fig. 5, where the quality of the generated images are observed to be enhanced at epoch 700. DCGAN executed for different epochs with goal of reducing the loss, sample shown in Figure 6. In this plot the loss is saturated towards 0 across generations with the loss of real and fake samples of discriminator and generator shown during training. DCGAN is experimented with different hyper parameters and are listed in Table 1.

Now, the trained DCGAN model is used to generate the augmented images for each of the three classes and then added with the original dataset. The augmented dataset has around 4000 images in each class which solved the problem of imbalanced data. A sample set of augmented images are shown in Fig. 8.

Model 2: DenseNet-201 with DCGAN To improve the performance of the model 1, DCGAN is used to overcome the problem of an imbalanced dataset. Input dataset is fed into the DCGAN to augment the dataset. Augmented dataset is then fed into DenseNet-201 model which extracts features from the chest images. **The features are then help to discriminate** the images into COVID-19, other pneumonia and normal. **This model has achieved training accuracy of 99.33% after 150 epochs.** The data augmentation using DCGAN has shown an increase in training performance from 97.31 % to 99.33%

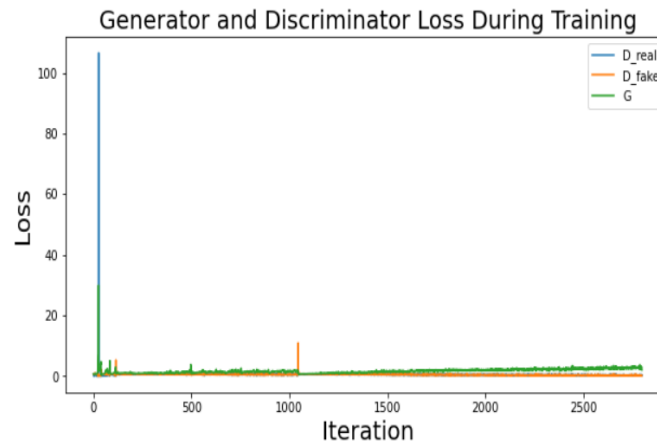


Fig. 6. DCGAN loss across epochs

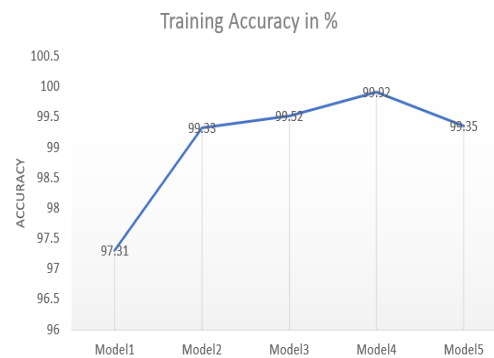


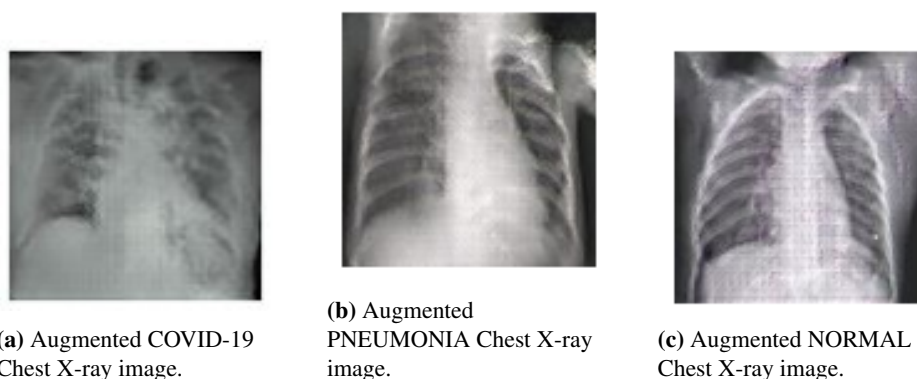
Fig. 7. Accuracy obtained during training

Model 3: SVM Using DenseNet-201 with DCGAN Model 3 uses the DenseNet-201 as feature extractor, that work upon the DCGAN augmented dataset. The features are then fed into the Support Vector Machine (SVM) classifier which classifies the images into COVID-19, other pneumonia and normal. SVM uses linear kernel, l2 penalty, with C set 1 whose stopping criterion tolerance is $1e^{-3}$, and the maximum iterations is not limited. This model has yielded a training accuracy of 99.52%.

Model 4: Random Forest Using DenseNet-201 with DCGAN Similarly, Densenet-201 is used as feature extractor from DCGAN augmented images and given to random forest classifier that classifies the features into COVID-19, other pneumonia and normal with a training accuracy of 99.92%. Random forest takes the default parameters namely the estimators are set to 100 with gini impurity, minimum number of samples at the leaf node set to 1 and the maximum number of feature will be taken as the square-root of the number of features with unlimited number of leaf nodes.

Table 1. DCGAN hyper parameters

Parameter	value
number of epochs	700
batch size	128
Size of z latent vector	100
Image size	128, 128
Number of channels	3
Learning rate for optimizers	0.0002
Activation Function	LeakyReLU (for all layers except last) and Sigmoid
Loss	binary crossentropy
Optimizer	Adam

**Fig. 8.** Augmented Chest X- ray images

Model 5: DCGAN with CNN To analyse and compare our proposed Densenet-201 model with other models, we used a simple CNN model. After augmentation with DCGAN, the augmented dataset is then fed into CNN model which extracts features from the chest images and classifies the images into COVID-19, other pneumonia and normal. This CNN model is trained for around 150 epochs and obtained a training accuracy of 99.35%.

The training accuracy of all the five models are shown in Figure 7. From this figure, it can be observed that the model 4 has shown good accuracy followed by Model 3. But when tested the Model 3 out shown the other models, discussed elaborately about its performance in the next section.

4.3. Results and Discussion

The proposed deep learning based system for COVID-19 detection is evaluated with the testing dataset and compared with the state of the art techniques.

Performance Metrics Performance of the model is evaluated using the metrics namely, Accuracy, Sensitivity, Specificity, Precision and F1-Score.

Table 2. Performance analysis of all models; Model 1: DenseNet-201; Model 2: DCGAN + DenseNet-201; Model 3: DCGAN + DenseNet-201 + SVM; Model 4: DCGAN + DenseNet-201 + RF; Model 5: DCGAN + CNN

Model	Accuracy in%	Sensitivity	Specificity	Precision	F1 Score
Model 1	94.87	0.9289	0.9570	0.9367	0.9413
Model 2	95.18	0.9316	0.9621	0.9560	0.9433
Model 3	95.49	0.9378	0.9665	0.9561	0.9699
Model 4	95.03	0.9394	0.9670	0.9476	0.9668
Model 5	94.79	0.9255	0.9672	0.9462	0.9653

The primary building blocks of the metrics are True Positive, True Negative, False Positive and False Negative. True Positive (TP) is the actual value was positive and the model predicted a positive value. True Negative (TN) is the actual value was positive and the model predicted a negative value. False Positive (FP) is the actual value was negative but the model predicted a positive value. This measure contributes to the calculation of Type 1 error or False positive rate (FPR). False Negative (FN) is the actual value was negative but the model predicted a negative value. FN helps to compute False Negative Rate which is also known as the Type 2 error.

Accuracy is a metric that generally describes how the model performs across all classes. It is the ratio between the number of correct predictions to the total number of predictions.

Sensitivity is a measure of the proportion of actual positive cases that got predicted as positive (or true positive rate). It can be interpreted as capacity of a classifier to predict the positive samples.

Specificity is defined as the proportion of actual negatives, which got predicted as the negative (or true negative rate). It can be understood as the capacity of a classifier to predict negative samples.

F1-score is the harmonic mean of Precision and Recall and gives a better measure of the incorrectly classified cases than the accuracy metric.

A Confusion matrix is an $N \times N$ matrix used for evaluating the performance of a classification model, where N is the number of target classes. The matrix compares the actual target values with those predicted by the trained machine learning model. Cohen's kappa is another metric computed for all the classification models in this work along with the misclassification rate. Kappa coefficient can assess the performance of a classification model by capturing intrinsic nature of the data.

Table 2 depicts the metrics to compare the performance of various deep learning framework. It is inferred that DenseNet-201 as feature representation with SVM classifier outperforms rest of the DenseNet based architecture.

On feeding our dataset directly to the DenseNet-201 model gives an accuracy of 94.87%. To improve the accuracy of the model, DCGAN is used which solved the problem of imbalanced data and gave an accuracy of 95.18%. To further improve the accuracy of the model, DenseNet-201 is added with Support Vector Machine (SVM) which improved the accuracy by 0.31% and gave 95.49% accuracy. This model performs well than simple CNN by an accuracy difference of 0.70%.

From the confusion matrix of Model 1 (DenseNet-201), it has been observed that the true positives are 110 for COVID-19, 271 as normal, and 841 as pneumonia images. True negatives are 1169 COVID-19, 956 normal, 385 pneumonia images. False positives are 3 COVID-19, 15 normal, 48 pneumonia images. False negatives are 6 COVID-19, 46 normal, 14 pneumonia images. The misclassification error is computed as 5.13% which is the second highest among the models under consideration here. The kappa coefficient is 89% which is the lowest of all the five classification models. This model has classified 13% normal images as Pneumonia and around 1.5% of pneumonia images as normal. Overall observation is more number of images are classified as Pneumonia, the classifier did not extract enough discriminating features to classify the COVID-19, normal and pneumonia samples. This model has got least sensitivity measure among all that is observed as the classifier is not good at detecting the positive samples.

From the confusion matrix of Model 2 (DCGAN + DenseNet-201), it is analysed that the true positives are 107 COVID-19, 284 normal, and 835 pneumonia images. True negatives are 1170 COVID-19, 950 normal, 394 pneumonia images. False positives are 2 COVID-19, 21 normal, 39 pneumonia images. False negatives are 9 COVID-19, 33 normal, 20 pneumonia images. This model has shown significant improvement than the classifier without DCGAN in terms of F1 score, but slightly lesser than the Model 3 which has SVM as classifier. It is observed that DenseNet-201's fully connected layers are not discriminating the features generated the convolutional layers of the same as compared with Model 3. This is the second classifier where more number of normal images are predicted as Pneumonia. This behaviour adds to misclassification accuracy, which is 4.82% , but still better than the model without DCGAN, along with the kappa coefficient as well.

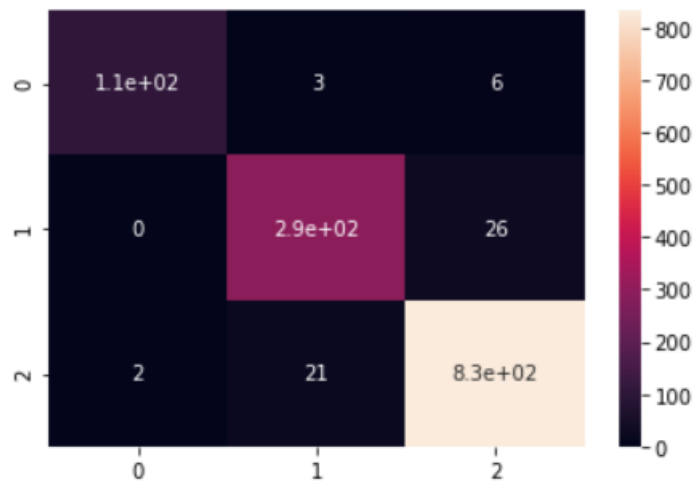


Fig. 9. Confusion matrix for Densenet-201 with SVM

From the confusion matrix (shown in Fig. 9) of Model 3 (DCGAN + DenseNet-201 + SVM), it is analysed that the true positives are 107 COVID-19, 287 normal, and 834

pneumonia images. True negatives are 1170 COVID-19, 949 normal, 397 pneumonia images. False positives are 2 COVID-19, 22 normal, 36 pneumonia images. False negatives are 9 COVID-19, 30 normal, 21 pneumonia images. The misclassification error for this model is 4.5%, the smallest of all the 5 models evaluated, with Kappa coefficient 0.91. These observations have shown that the combination network with DCGAN + DenseNet-201 + SVM has predicted more number of actuals, which was also supported by the F1 score of 96.99%. This infers that, scoring a high F1 score is due to the fact that number of False Positives and False Negatives must be low among all the classifiers. Also adds to the fact that Model 3 has generated more number of True Positives than any other model considered in this work.

Table 4 shows the metrics for each of the three classes for best performing Model 3 (DCGAN + DenseNet201 + SVM). From the results it has been observed that the more images are predicted as Pneumonia. That has also reflected here in the metrics of Table 4 where Pneumonia class of images have higher F1 score of about 96.80 %, followed by the COVID-19 class with 95.11%. This can be interpreted as more number of True positives are predicted by the model. This framework has generated second highest sensitivity among all, which is interpreted as the classifier being good at predicting the positive samples.

The actual training set before applying DCGAN for COVID-19 class has 460 images which were augmented to 4000 by DCGAN. COVID-19 class has the least set of original images among all. The results show that the application of DCGAN for augmentation has performed well and has not degraded the performance of the Model 3 by any means in the prediction.

From the confusion matrix of Model 4 (DCGAN + DenseNet-201 + RF), it is analysed that the true positives are 107 COVID-19, 295 normal, and 824 pneumonia images. True negatives are 1170 COVID-19, 939 normal, 405 pneumonia images. False positives are 2 COVID-19, 32 normal, 28 pneumonia images. False negatives are 9 COVID-19, 22 normal, 31 pneumonia images. This classifier has given a second best performance among all with a F1 score of 96.68% with a misclassification rate of about 4.97%. Also this combination has given the highest sensitivity that shows that the classifier is good at predicting the positive samples. With respect to specificity this model has shown a moderate behaviour in detecting the negative samples.

From the confusion matrix of Model 5 (DCGAN + CNN), it is analysed that the true positives are 100 covid, 304 normal, and 817 pneumonia images. True negatives are 1171 COVID-19, 928 normal, 410 pneumonia images. False positives are 1 COVID-19, 43 normal, 23 pneumonia images. False negatives are 16 COVID-19, 13 normal, 38 pneumonia images. This model with image augmentation and CNN to generate features and classify is better than model without augmentation in terms of F1 score but significantly lesser than models with DenseNet-201 as feature extractor. This shows that the dense blocks with strengthened feature propagation in the DenseNet-201 have outperform the simple convolutional layers of CNN. The other characteristic exhibited by this framework is having highest specificity that is this classifier is good at detecting negative samples.

To show the proposed approach (DCGAN+DenseNet-201+ SVM) is statistically significant than the other four approaches, Paired t-test with equal variance was applied between the proposed model and the other four models on training accuracy across the dif-

Table 3. Misclassification error and Kappa Coefficient

Model	Misclassification Rate	Kappa Coefficient
Model 1	5.13	0.893
Model 2	4.82	0.90
Model 3	4.51	0.91
Model 4	4.97	0.899
Model 5	5.21	0.895

Table 4. Class-wise metrics for Model 3 (DCGAN + DenseNet201 + SVM)

Class	Per Class Precision	Per class Recall	Per class F1
COVID-19	92.24	98.17	95.11
Normal	91.80	93.38	92.58
Pneumonia	97.31	96.3	96.80

ferent epochs. With 95% confidence interval, the two tailed p value has shown significant difference between the proposed and the four other approaches to detect the COVID-19.

Analysis

1. DenseNet-201 is observed to be a better network as feature extractor than simple sequential CNN layers.
2. Models with traditional machine learning classifiers namely SVM, Random Forest have out performed than the neural network based classifiers. Thus from the research performed, the DenseNet-201 model with SVM outperformed any of the other combinations of models constructed and passed the statistical t-test.
3. From the results shown in Tables 2, 3 and 4 it can be inferred that data augmentation is helping the system to learn better.
4. Most of the Normal samples are misclassified as Pneumonia, since the feature pattern of Pneumonia appear to be similar to Normal samples. Neither the CNN nor the DenseNet-201 has the ability to identify this to reduce the number of False positives.

4.4. Comparison with existing works

The literature [17], [16], [19] show that the effectiveness of deep learning on COVID-19 detection. As reported in [19] simple CNN has obtained an accuracy of 89.75%. It is inferred that CNN is not performing well because of the data imbalance problem and hence enough samples are not available for training leads to this performance. DCGAN in the proposed framework has handled that problem and has given improvement in performance and obtained an accuracy of 94.79%. This improvement is significant due to the presence of DCGAN. Comparison of other existing methods with our model that have used the same dataset is shown in Table 5.

VGG-16 is a 16 layer architecture with a pair of convolution layers, a pooling layer and at the end a fully connected layer. It features a plethora of weight parameters, the

Table 5. Comparison of other existing methods with our proposed DCGAN + DenseNet-201 + SVM model

Model	Accuracy in %
VGG-16 [17]	66.3
Simple CNN [19]	89.75
ResNet50 [16]	95.10
Proposed DCGAN + CNN	94.79
Proposed DCGAN + DenseNet-201+ SVM	95.49

models are very heavy, which also means a long inference time. It also has a vanishing gradient problem. Due to these, results show the poor performing VGG-16 than any other model taken up for comparison here.

The results show minor difference between ResNet50 [16] and proposed DCGAN +DenseNet-201+ SVM, but the suitable network for the detection of COVID-19 is proven to be the proposed one. The reasons are as follows: ResNet was proposed to beat the issues of VGG styled CNNs. ResNets need less memory, have a faster inference time, and allow for the training of deeper networks. It also solved the problem of vanishing gradient by skipping the connections. Densenet adds shortcuts among layers and has thinner network and having fewer number of channels is better than the identity mapping feature of ResNet. Also, in contrast to ResNet, a dense layer gets all outputs from preceding layers and concatenates them in the depth dimension. Densenet also uses much fewer parameters than ResNet with fewer redundant layers as well. Fewer redundant layers mean more parameter efficiency and less computation and hence DCGAN +DenseNet-201+ SVM proven to be a better framework for detection of COVID-19.

Analysis

1. Compared to CNN and VGG-16, DenseNet-201 is performing well as shown in the Tables 5 and 2, since the Model 1 which is without augmentation has given an accuracy of 94.87%.
2. The presence of skip connections and the handling the vanishing gradient problem have made both ResNet and DenseNet-201 better than VGG-16 and CNN
3. Between ResNet and DenseNet-201, the proposed work with DenseNet-201 (Model 3) has exhibited a better performance, because of the augmentation of images, less complexity in terms of layers, parameters and hence computational time.

5. Conclusion and Future work

In this paper, a novel deep learning model is designed for COVID-19 disease detection with the help of Deep Convolutional Generative Adversarial Network (DCGAN) and DenseNet-201 with SVM classifier. The proposed model is able to detect COVID -19 in chest images with the accuracy of 95.49% and with F1 score of about 96.99%. Comparative analyses revealed that the DenseNet-201, a transfer learning based deep learning

framework with SVM classifier performs significantly better as compared to other approaches. Therefore, the proposed model can act as an lead for further research process.

In future we will explore the approaches for evolving and newer data patterns to provide early alerts in addition to detecting the severity of the infection. Also the limitation observed in this work is, Normal images are more often classified as Pneumonia. The features generated are enough but not sufficient to discriminate them with less false positives.

References

1. Overview of GAN Structure). https://developers.google.com/machine-learning/gan/gan_structure (April 22, 2019)
2. Abbas, A., Abdelsamea, M.M., Gaber, M.M.: Classification of covid-19 in chest x-ray images using detrac deep convolutional neural network. *Applied Intelligence* 51(2), 854–864 (2021)
3. Ahmed, S., Yap, M.H., Tan, M., Hasan, M.K.: Reconet: Multi-level preprocessing of chest x-rays for covid-19 detection using convolutional neural networks. *medRxiv* (2020)
4. Anand, A., Anandan, K.R., Jayaraman, B., Thai, M.T.N.T.: Simple Neural Network based TB Classification. *Proceedings of the Working Notes of CLEF 2021* 2936, 1145–1150 (2021)
5. Balwal, U., Yeragudipati, S.A., Bhuvana, J., Mirnalinee, T.T.: Deep learning based tb severity prediction. In: *CLEF (Working Notes)* (2020)
6. Basavegowda, H.S., Dagnev, G.: Deep learning approach for microarray cancer data classification. *CAAI Trans. Intell. Technol.* 5(1), 22–33 (2020)
7. Bhuvana, J., Mirnalinee, T.T.: An approach to plant disease detection using deep learning techniques. *ITECKNE* 18(2), 1–14 (2021)
8. Dąbrowski, M., Michalik, T.: How effective is transfer learning method for image classification. In: *Proceedings of the Position Papers of the 2017 Federated Conference on Computer Science and Information Systems*. vol. 12, pp. 3–9 (2017)
9. Gao, T.: Chest x-ray image analysis and classification for covid-19 pneumonia detection using deep cnn. *medRxiv* (2020)
10. Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. *science* 313(5786), 504–507 (2006)
11. Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., et al.: Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. *The lancet* 395(10223), 497–506 (2020)
12. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 4700–4708 (2017)
13. Jaisakthi, M.S., Mirunalini, P., Thenmozhi, D., Muthukumar, V.: Fish species recognition using transfer learning techniques. *International Journal of Advances in Intelligent Informatics* 7(2), 188–197 (2021)
14. Jaiswal, A., Gianchandani, N., Singh, D., Kumar, V., Kaur, M.: Classification of the covid-19 infected patients using densenet201 based deep transfer learning. *Journal of Biomolecular Structure and Dynamics* pp. 1–8 (2020)
15. Kavitha, S., Poornima, S., Sitara, N.S., Sarada Devi, A.: Classification of lung tuberculosis using non parametric and deep neural network techniques. In: *2020 4th International Conference on Computer, Communication and Signal Processing (ICCCSP)*. pp. 1–5 (2020)
16. Korkmaz, A.: Prediction from x-ray images (resnet50). <https://www.kaggle.com/ahmetkorkmaz/prediction-from-xray-images-resnet50> (2021, January 25)
17. Korkmaz, A.: Prediction from x-ray images(vgg16). <https://www.kaggle.com/ahmetkorkmaz/prediction-from-xray-images-vgg16> (2021, January 25)

18. Lu, M.T., Ivanov, A., Mayrhofer, T., Hosny, A., Aerts, H.J., Hoffmann, U.: Deep learning to assess long-term mortality from chest radiographs. *JAMA network open* 2(7), e197416–e197416 (2019)
19. Luis, M.: Convolutional neural networks to detect lung disease in Chest X-ray images. <https://www.kaggle.com/marcelor/cnn-chestxray-87-f1-score/notebook> (2020, November 17)
20. Maghdid, H.S., Asaad, A.T., Ghafoor, K.Z., Sadiq, A.S., Mirjalili, S., Khan, M.K.: Diagnosing covid-19 pneumonia from x-ray and ct images using deep learning and transfer learning algorithms. In: *Multimodal Image Exploitation and Learning 2021*. vol. 11734, p. 117340E. International Society for Optics and Photonics (2021)
21. Marimuthu S, Bhuvana, J., Mirlalinee, T.T.: Disease detection in tomato plants using deep learning. *Intelligent Systems and Computer Technology, Advances in Parallel Computing* 37, 190–195 (2020)
22. Nishio, M., Noguchi, S., Matsuo, H., Murakami, T.: Automatic classification between covid-19 pneumonia, non-covid-19 pneumonia, and the healthy on chest x-ray image: combination of data augmentation methods. *Scientific reports* 10(1), 1–6 (2020)
23. Patel, P.: Chest X-ray (Covid-19 & Pneumonia). <https://www.kaggle.com/prashant268/chest-xray-covid19-pneumonia> (2020, September 17)
24. Pathak, Y., Shukla, P.K., Tiwari, A., Stalin, S., Singh, S.: Deep transfer learning based classification model for covid-19 disease. *Irbm* (2020)
25. Punn, N.S., Agarwal, S.: Automated diagnosis of covid-19 with limited posteroanterior chest x-ray images using fine-tuned deep neural networks. *Applied Intelligence* 51(5), 2689–2702 (2021)
26. Sedik, A., Ilyasu, A.M., El-Rahiem, A., Abdel Samea, M.E., Abdel-Raheem, A., Hammad, M., Peng, J., El-Samie, A., Fathi, E., El-Latif, A., et al.: Deploying machine and deep learning models for efficient data-augmented detection of covid-19 infections. *Viruses* 12(7), 769 (2020)
27. Singhal, T.: A review of coronavirus disease-2019 (covid-19). *The indian journal of pediatrics* 87(4), 281–286 (2020)
28. Sladojevic, M.A.S.S.S.: Detection of covid-19 cases by utilizing deep learning algorithms on x-ray images. In: *Proceedings of the 18th International Scientific Conference on Industrial Systems Industrial Innovation in Digital Age*. pp. 1–8
29. Tang, Y.: Deep learning using support vector machines. *CoRR*, abs/1306.0239 2 (2013)
30. Vrbačič, G., Pečnik, Š., Podgorelec, V.: Hyper-parameter optimization of convolutional neural networks for classifying covid-19 x-ray images. *Computer Science and Information Systems* (00), 56–56 (2021)
31. Wang, L., Lin, Z.Q., Wong, A.: Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Scientific Reports* 10(1), 1–12 (2020)
32. Wang, S.H., Zhang, Y.D.: Densenet-201-based deep neural network with composite learning factor and precomputation for multiple sclerosis classification. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 16(2s), 1–19 (2020)
33. Wang, W., Xu, Y., Gao, R., Lu, R., Han, K., Wu, G., Tan, W.: Detection of sars-cov-2 in different types of clinical specimens. *Jama* 323(18), 1843–1844 (2020)
34. Zhang, J., Xie, Y., Pang, G., Liao, Z., Verjans, J., Li, W., Sun, Z., He, J., Li, Y., Shen, C., et al.: Viral pneumonia screening on chest x-ray images using confidence-aware anomaly detection. *arXiv preprint arXiv:2003.12338* (2020)

J. Bhuvana Associate Professor in the Department of Computer Science and Engineering with 22 years of experience in teaching. Before joining SSN in 2006, she worked as Assistant professor in AVC College of Engineering for 8 years. She received her PhD from

Anna University, Chennai in 2015, with master degree, ME in CSE from Annamalai University, Chidambaram in 2004 with First class and Distinction. She completed BE in CSE from University of Madras in 1998. Her research interests include Deep learning, Multiobjective optimization, Memetic Algorithms, Evolutionary Algorithms, Machine Learning.

T. T. Mirnalinee is a Professor at SSN College of Engineering, Chennai, India, and is currently the head of the department of Computer Science and Engineering. She received her B.E. degree from Bharathidasan University, Trichy, M.E. degree from the College of Engineering, Guindy, Anna University, Chennai, and Ph.D. from Indian Institute of Technology Madras (IITM), Chennai, India. Her research interests include Computer vision, Machine learning, Green Networks and Software Defined Networks. Seven research scholars have completed PhD under her supervision and she is currently guiding seven more scholars. Mirnalinee has completed three research projects, and has published about 80 papers in international journals and conferences. She has reviewed several papers in international journals and chaired several sessions in conferences.

B. Bharathi Associate Professor in the Department of Computer Science and Engineering has 24 years of teaching and research experience. She received her PhD. in Computer Science (2014) from Anna University, Chennai, M.E. Computer Science & Engineering (2006) from SSN College of Engineering, Anna University, Chennai, and B.E Computer Science & Engineering (1998), from University of Madras. She has completed 2 externally funded projects in the area of speech processing. She has published around 88 papers in international conferences and journals.

Infant Sneha S. is a Software Engineer currently working in Optum Global Solutions, India. She received a bachelor's degree in Computer Science Engineering at SSN College Of Engineering, Chennai in 2021. Her research interests include detecting COVID 19 with chest X-ray images using deep learning concepts.

Received: February 07, 2022; Accepted: July 05, 2022.

Federating Digital Contact Tracing using Structured Overlay Networks

Silvia Ghilezan^{1,2}, Simona Kašterović², Luigi Liquori⁴, Bojan Marinković^{3,1}, Zoran Ognjanović¹, Tamara Stefanović²

¹ Mathematical Institute of the Serbian Academy of Sciences and Arts
Belgrade, Serbia

{bojanm, zorano}@mi.sanu.ac.rs

² Faculty of Technical Sciences, University of Novi Sad,
Novi Sad, Serbia

{gsilvia, simona.k, tstefanovic}@uns.ac.rs

³ Clarivate, Serbia

⁴ Inria & Université Côte d'Azur, France
Luigi.Liquori@inria.fr

Abstract. In this paper, we present a comprehensive, yet simple, extension to the existing systems used for Digital Contact Tracing in Covid-19 pandemic. The extension, called *BubbleAntiCovid19 (BAC19)*, enables those systems, regardless of their underlying protocol, to enhance their sets of traced contacts and to improve the global fight against pandemic during the phase of opening borders and enabling more traveling. *BAC19* is a *Structured Overlay Network*. Its protocol is inspired by the Chord and Synapse Structured Overlay Networks. We design the architecture of the Overlay Network Federation. We show that the federation can be used as a formal model of *Forward Contact Tracing*. *BAC19* provides a *fully exhaustive* retrieving procedure thanks to avoiding search during peer churn. Furthermore, we give simulation results for the *BAC19* system, the simulator written in Python.

Keywords: Covid-19, Digital Contact Tracing, Distributed Hash Tables, Structured Overlay Networks, Bluetooth, GPS.

1. Introduction

One of the biggest challenges the world is facing since the beginning of Covid-19 pandemic is to slow down the spreading of SARS-CoV-2 virus producing Covid-19 pandemic; *Prevention, Testing and Tracing* are the main pillars of the solution for controlling and slowing down the spread of the virus. Contact Tracing of an infected person is essential to control the spread of the disease.

Tracing. Contact tracing is the process of identifying, notifying, and monitoring people who came in close contact with an individual who was tested positive for an infectious disease, like Covid-19, while he/she was infectious. Contact tracing benefits the fight with the pandemic at multiple levels. Identifying and quarantining close contacts limits their ability to spread the disease. Therefore, in a period in which the disease and its effects are still being investigated, contact tracing plays a key role in preventing the further spread of the disease. In [20] the authors studied the spreading of H1N1 virus and showed that

tracing close contacts can effectively slow down the rate of virus spreading. Furthermore, contact tracing data helps medical experts to find the origin of the virus and learn more about the nature of the virus.

Manual Contact Tracing. Contact tracing has mostly been done manually since many centuries ago just by taking note on a simple piece of paper the list of persons and goods you get in contacts with (see e.g. *La Peste* by A. Camus [8]). In the actual days, manual contact tracing could be exploited using simple telephone calls. Identifying contacts is done through an interview with the person infected with the virus. Each person is then contacted by phone. Health Authorities should quickly alert people who are close contacts that they may have been exposed to the virus. The sooner the contacts are notified, the lower the risk of the spreading further. However, due to the highly contagious nature of the SARS-CoV-2 virus and the fact that symptoms can manifest after many days (or even never, e.g. *asymptomatic cases*), manual contact tracing does not give satisfactory results. Health Departments and National Authorities do not have enough employees to do manual contact tracing. It must be further emphasized that the SARS-CoV-2 can be transmitted not only by direct contact, but also by indirect contact. The reason is that infected people can leave virus droplets on any physical object they touch. In this case, actual digital contact tracing solutions are ineffective. For the reasons stated above, digital contact tracing has been considered already at the beginning of the Covid-19 pandemic.

1.1. Problem

There is a plethora of Digital Contract Tracing (DGT) applications in use all over the world fighting the Covid-19 pandemic [14, 26]. They are developed on very different paradigms, centralized [10] vs. decentralized [35], GPS based (very few indeed because of a clear violation of privacy) vs. Bluetooth Low Energy based (the majority). The rush to make these applications work in the shortest time led to their great diversity. The most important open problem is their interoperability. There are many ongoing efforts to make a “federation” of these different systems. Herein, we address this problem and propose a solution based on mathematical models of overlay networks. We leave to the the reader to envisage the following scenario:

Alice lives in the region which has centralized DCT *System A*, while Bob lives in the region which has centralized DCT *System B*. Bob has spent some time in the region A, and both of them are traveling together side by side with negative RT-PCR tests. However, Bob developed symptoms of Covid-19 after couple of days and was confirmed as positive.

1.2. Contributions

We develop a formal *Federation of Overlay Networks*, called *BubbleAntiCovid19 (BAC19)*, for connecting different digital contact tracing applications, which are currently in use all over the world. The model is based on the well-known model of Structured Overlay Network protocols like e.g. Chord [31, 32], Kademia [27], and Synapse [21]. However, in these well-known cases nodes are volunteer to join/leave a network when ever it want. In *BAC19* we control all the nodes that participate in the system and it is not allowed that

a node from the public domain joins the system. Also, *BAC19* is conceived as a topology composed by trusted overlay networks, in the sense that every node, before enter in one *BAC19* overlay, has been trusted by a *Health Authority* within the original application. There is one and only one *Health Authority per* original application included in the *BAC19* system. Original applications communicate with *Health Authorities* in order to validate the test result of each user. Even if it is possible to choose a different technique to solve this problem, e.g. distributed databases, our choice was Distributed Hash Tables (DHT) since we see it as a natural extension of the basic idea that all contacts of one person should be stored in one overlay network and the contact between persons could be seen as “the synapse nodes”. Further, a distributed database is a federation governed by a “distributed server dictionary/jellow pages” (e.g. Microsoft Azure, Amazon elastic, etc...), whereas *BAC19* can be populated by nodes trusted by the *Health Authority*, so giving, in principle the capability to servers belonging to Municipalities, Departments, etc. to participate to the Overlay.

We prove that *BAC19* provides a fully exhaustive retrieving procedure of people that get in touch with other people having tested positive to the Covid-19 disease. Hence, *BAC19* is proven to be a simple yet powerful *interconnection* of already existing digital contact tracing applications that - by construction - do not communicate with each others as such providing their efficient interoperability.

BAC19 solves the problem presented in the scenario with Alice and Bob. Suppose System A and System B are part of *BAC19*. For Alice and Bob it will be sufficient that only one of them installs the system of the other region during the time of their travel, so that Alice gets informed that she is the first contact of an infected person, namely Bob during their joint travel.

We give simulation results by running a prototypical implementation of the *BAC19* system, written in Python. The simulator creates the *BAC19* topology, generates random requests for the search procedure and calculates the success rate. The simulator does not skip any node, so that the success rate of every iteration is 1, as expected.

As far as we know, the mathematical model and techniques presented in this paper have not been considered in other approaches.

1.3. Overlay networks in a Nutshell

Structured Overlay Networks [13] are suitable models of scalable and efficient organization of resources on the Internet. They represent logical organizations, independent on underlying network infrastructure that physically connects available assets. Structured Overlay Networks have been proven as very resilient tool (they can be reliable as central servers or distributed cloud solutions) in the situation when some parts of the underlying infrastructure fail or become overloaded or corrupted.

1.4. Organization of the Paper

The rest of the paper is organized as follows. Section 2 presents classifications of digital contact tracing applications. Section 3 reviews Chord and Synapse protocols of Structured Overlay Networks. It also briefly reviews basic notions of Abstract State Machines and some related work by the authors. Section 4 introduces the *BAC19* system and proves

the completeness and full exhaustiveness of the retrieving procedure. Section 5 presents the simulation results for the *BAC19* system. Section 6 presents a discussion on other proposals for providing interoperable frameworks for digital contact tracing. Section 7 concludes the paper.

2. Digital Contact Tracing Applications

Advances in digital technology have enabled smartphones and other digital devices to be used for contract tracing. Particularly, more and more countries are showing interest in *Digital Contact Tracing* applications (DCT apps) implemented for smartphones. Despite the great variety among these applications, all contact tracing apps work on the principle of automatic data exchange with nearby devices. When a user of a particular contact tracing app is identified as infected, a special report is uploaded to the DCT app server. Based on that report, close contacts of the infected person (also DCT app users) are informed that they have been in a contact with a positive user and/or the app calculates their exposure risk. The identity of the infected persons is not disclosed in order to protect their privacy. Existing contact tracing apps can be classified based on two criteria:

- System architecture and topology;
- Contact-tracing technology.

More information about the classification of the existing contact tracing apps can be found in [30,33].

2.1. Classification by System Architecture

Since the data is collected from users, their processing should be addressed. When it comes to DCT app data managing, the responsibility can be on a central authority or on each user individually. Therefore, contact tracing apps can be divided into three major categories depending on the architecture of the underlying system:

- Centralized apps - data are solely managed by a central server;
- Hybrid apps - multiple nodes can manage the data, but the control is centralized;
- Decentralized apps - each user is managing his/her data.

In centralized apps, the central server is responsible for ID generation, risk analysis and notifications. Centralized apps can, when using GPS and no encrypted IDs, raise privacy concerns and questions about massive surveillance. People often do not trust servers and as a consequence there is a low adoption rate of these applications. On the other hand, the advantage of centralized apps is the possibility for Health Authorities to make a transmission graph and learn more about the virus. Also, the possibility of false positive users is reduced. As we have already observed, privacy issues lead to low adoption rate and decrease the efficiency of the application. In order to solve the problem of distrust of the central server, decentralized apps were designed. In decentralized apps these functionalities are moved to the user devices and the central server is only an encounter point. However, decentralized methods also raise some privacy concerns, for example the identity of a patient can be easily revealed again when IDs are not encrypted. The hybrid

architecture proposes decentralized ID generation and centralized risk analysis. There are a few proposed hybrid contact tracing protocols, but their implementation in real contact tracing apps is still pending. More about these protocols can be found in [2]. For that reason, we will focus on apps with centralized and decentralized architecture, and we will provide an overview of the existing contact tracing apps based on the above classifications in Section 2.3. The main apps are summarized in Figure 1, which is motivated by [30].

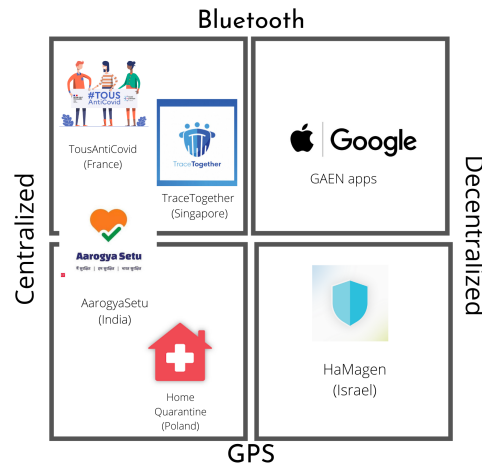


Fig. 1. Classification of analyzed applications

2.2. Classification by Contact-Tracing Technology

In order for two people to be in close contact, they need to stay in the same place, at a short distance, and for a long enough period of time. Therefore, the main data type used by the contact tracing apps is location data. There are various technologies for collecting and tracking location data, and contact tracing apps can be divided into two major categories depending on whether they track absolute or relative location:

- Absolute location apps – Apps that track the absolute location of their users are mostly based on GPS technology. Location data is stored in the form of geolocation coordinate pair. These apps are also known as Geolocation-based DCT apps.
- Relative location apps – Apps that track relative location of their users are mostly based on Bluetooth technology. These apps are also known as proximity DCT apps. A boarding pass or a ticket for a specific event can also be considered as relative location data. In order to use this kind of data, some contact tracing apps deploy QR code technology.

Geolocation-based DCT apps record past geo-trajectories of every user, and the calculation of exposure risk of a user is based on the intersection of its past trajectories and

trajectories of patients. We give a brief review of the existing Geolocation-based apps in Section 2.3.

Two main advantages of geolocation-based DCT apps are the following:

- 1) Geolocation-based DCT apps are compatible with manual contact tracing. Compatibility of geolocation-based DCT apps and manual contact tracing has mutual benefits. On the one hand, past geo-trajectories of a patient can be added to an app by the contact tracer even if the patient did not use the app. This enables the app to warn more users. On the other hand, the app can give the information about places with higher exposure risk to a contact tracer, so that the contact tracer can identify high-risk service workers.
- 2) Another advantage is that geolocation-based DCT apps can recognize patterns of disease's spreading and locations with higher exposure risk, and they can inform health authorities about it.

Nevertheless, geolocation-based DCT apps have also disadvantages. The major challenges are privacy concerns, which cause low adoption rate of these applications. User's privacy can be violated in several ways. Recording all user's trajectories can result in revealing user's personal information such as identity, home address, work address, the identity of the patient and revealing user's exposure risk to other users. These problems have been elaborated in more details in [12].

Bluetooth-based DCT apps record direct contacts of the users. A device generates a unique, randomized identifier and assigns it to a user. There are two kinds of identifiers: static, identifiers do not change over time, and dynamic, identifiers change over time. During a direct contact devices exchange identifiers and save received identifiers. Once a user is identified as positive in the application, other users can calculate their exposure risk by checking whether they received a patient's identifier. We give a brief review of the existing Bluetooth-based apps in Section 2.3.

Depending on whether the exposure risk is calculated by the central server or the user's device, we have centralized and decentralized apps, respectively, see Figure 2, which is motivated by [18].

The major disadvantage is that Bluetooth-based DCT apps work only if both users have installed the *same* application.

In order to take advantage of both types of these apps, Bluetooth-GPS apps were designed, see Section 2.3. Given the different characteristics of DCT apps, the question arises whether it is possible to aggregate their data in order to track contacts more effectively. The answer can be found in Structured Overlay Networks.

2.3. DCT Apps - Overview

In this section we give a review of the existing DCT apps.

Geolocation-based DCT apps. *Home Quarantine.* At the beginning of the Covid-19 pandemic, Ministry of Digital Affairs of Poland developed the Home Quarantine app. More details about this app can be found in [34]. This is a typical example of a centralized app which deploys GPS technology. It is developed to support the authorities,

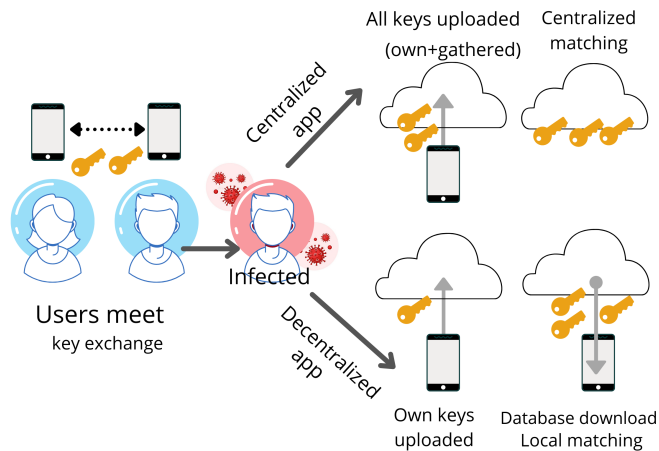


Fig. 2. Centralized vs Decentralized Bluetooth-based apps

especially the police and social services, with adequate information about people undergoing mandatory home quarantine. Users are also required to upload their digital photos. So, aside the GPS technology the app also uses face recognition. The app is mandatory for anyone who has developed coronavirus symptoms. It should be emphasized that Poland also developed the ProteGO Safe app for alerting users of close contact with an infected person based on The (Google/Apple) Exposure Notification (GAEN) system [4].

The Shield (HaMagen). In March 2020, Israeli Ministry of Health developed The Shield app [3]. This is a typical example of a decentralized app which deploys GPS technology. Location data is stored in the phone. If a user tests positive, he/she can upload his/her location history to the central server. Once the user uploaded his/her location history, it is added into a JSON file that is updated with new data on an hourly basis. Matching the locations happens on the phone. If the match is found, the app shows you the exact time and location. The app is later updated to work with Bluetooth technology but on a voluntary basis, every user can choose whether to use the proximity data or not.

Bluetooth-based DCT apps. *Blue-Trace protocol apps.* Singapore’s Government Technology Agency in collaboration with Ministry of Health in March 2020 released the TraceTogether app that allows digital contact tracing using the custom Blue-Trace protocol. Australia has later adopted the protocol and released the CovidSafe app. More details about these apps can be found in [1]. Contact tracing is done using Bluetooth Low Energy and proximity data is encrypted and stored only on the users phone. Users in the contact log are identified using anonymous time-shifting “temporary IDs”. If a user tests positive for the infection, the Ministry of Health requests his/her contact log. The user has the right to choose whether to share the contact log or not. If the user chooses to share the log, the contact log is uploaded to a central server and the Health Authority is then responsible for matching the log to contact detail and informing close contacts of the infected user. These apps are examples of Bluetooth-based centralized apps. It should also be noted that

Singapore solved the problem of tracing people who don't use smartphones by enabling the app to work with Token - a physical Bluetooth-based device.

ROBERT protocol app. Inria (France) and Fraunhofer AIESEC (Germany) released the ROBERT protocol (ROBust and privacy-presERving proximity Tracing) [9]. The French Government released the StopCovid (later renamed to TousAntiCovid) app in May 2020. It also deploys Bluetooth technology and belongs to the category of centralized apps. The difference between this app and apps based on the BlueTrace protocol relates to confirmation of positive users. More precisely, in France when a person is confirmed to be positive, the lab gives a patient a QR code and the scanned code is the proof for the app that you are infected. It is up to you to share this information with the app, and if you choose to share this information with a central server, the server is responsible for alerting your close contacts.

Google/Apple exposure notification apps. In April 2020, Google and Apple announced the joint work on decentralized Bluetooth-based protocol named The Google/Apple Exposure Notification (GAEN) system [4]. Many states then developed different apps using the Google/Apple Exposure Notification framework including Austria (Stopp Corona app), Germany (Corona-Warn-App), Italy (Immuni), Canada (COVID Alert) etc. The principle by which applications work is as follows. During a close contact, user's phones exchange random Bluetooth identifiers. These identifiers change frequently and the information about exchanged ID's is stored on the user's phone. When a user gets infected, he/she can decide to upload ID's he/she was using the last 14 days to the server. Phones of all users periodically download the list of ID's which belong to the infected users and does the matching locally.

Bluetooth-GPS apps. Apps that deploy both Bluetooth and GPS technology are rare. One app of this kind is the Aarogya Setu app [6], developed by National Informatics Centre that comes under the Ministry of Electronics and Information Technology, Government of India. Aarogya Setu is following the centralized approach, and is one of the world's fastest growing applications. The app mainly uses proximity data and GPS data are recorded only once in 30 minutes. The location data is mainly used to identify the locations where you might have caught the infection and identify potential hotspots that may be developing when multiple infected people visit the same place. Interaction between users is recorded by exchange of Device Identification Numbers (DID's) which are static. Contact tracing data is kept on the phone. Council of Medical Research (ICMR) shares the list of Covid-19 positive persons with the Aarogya Setu server, and information about contact tracing is uploaded to the server only if you are tested positive. The central server is then responsible for alerting your close contacts.

3. Overlay Networks

In this section, we briefly review basic notions of Structured Overlay Networks for the purpose of this work. Some Structured Overlay Networks are implemented in a form of Distributed Hash Tables (DHTs). One of DHT protocols is the Chord protocol. We also provide a brief review of Abstract State Machines, since they are used for proving properties of Chord. For more details, please see [7, 16, 21, 23, 25, 31, 32].

3.1. Chord and Synapse

Chord was introduced in [31,32]. Nodes that are part of a Chord system form a ring shaped network. The basic operations of a Chord node are entering and leaving the system and the mapping given key onto the corresponding node of the system using consistent hashing.

The correctness and efficiency of the Chord's protocol lookup procedure was in the focus of several papers, e.g. [23, 25, 31, 32]. In [11] the authors exploited the benefits of Chord, such as accelerating the lookup, and developed an energy efficient routing protocol for Wireless Sensor Networks, called *CHEARP: Chord-based Hierarchical Energy-Aware Routing Protocol for Wireless Sensor Networks*. However, these properties will not be in the focus of this paper. Our goal is to deliver information of every affected node, so we will not use any presented improvements to speed up the process of getting results, but to linearly pass every node in a Chord network, to be sure that no information is missed.

Interconnection of several Structured Overlay Networks is a very hard problem since different networks may use different protocols, and even in the case of several DHT networks that use the same protocol (e.g. Chord) it is enough that every overlay network uses *its own hash function* and information between two of them cannot be exchanged. A proposal to solve this issue was given by defining the Synapse protocol in [21]. Its performances were analyzed in [22], whereas one real-life proof of concept was developed in [24].

In the Synapse protocol, the interconnection of intra-overlay networks is achieved by co-located nodes taking part in several of these intra-overlays, called *Synapses*. Each node acts in accordance with the policies of each of its intra-overlays, but it also has the extra-role of forwarding the requests to some other intra-overlay it belongs to. Every node comes with a proper logical name in each intra-overlay; in particular, synapses have as many logical names as the number of networks they belong to. Each node is responsible for a set of resources it hosts, (key,value) pairs. Every key also comes with a proper logical name peculiar to each intra-overlay.

For the purpose of this paper we will consider the so-called, *white-box* version of the Synapse protocol that, in short, allows to consider all the keys as they were using the same hash table (see [21] for details). Again, since we will use the linear search procedure in one Chord network we can be sure that information will be retrieved if it exists in the system.

3.2. Abstract State Machines

In this subsection, we briefly review basic notions of Abstract State Machines for the purpose of this work. For more details, please look at [7, 16].

Abstract State Machine (ASM) [7, 16] is a formalization method for algorithms at the appropriate abstraction level. An ASM \mathcal{A} is defined as a program *Prog* which consists of:

- an at most countable set of states, its subset of initial states, and
- a finite number of transition rules.

States are first order structures over a fixed signature, whereas transitional rules:

- update ($:=$),
- sequential (seq ... endseq),

- conditional (if ... then ... else ... endif),
- parallel (par ... endpar),
- nondeterministic (choose $v \in U$ satisfying $g(v)$... endchoose) and
- universal (forall v with g ... endforall)

represent next-state functions. An execution of one of the last two types of rules introduces a variable v . In the case of nondeterministic rule, the transition is executed with a value of v which satisfies a guard g , while in the case of universal rule, the transition is executed simultaneously for all values v which satisfy a guard g . An ASM can interact with its environment using external functions (oracles) by providing arguments to oracles and receiving the corresponding results.

In a distributed case with many agents, every agent executes its own program and has its own partial view of a global state. The nullary function Me allows an agent to identify itself among other agents. The global program is the union of all agents' programs, whereas a transition between two states is obtained by an evaluation of transition functions of all agents.

An ASM \mathcal{A} models a real system \mathcal{S} in terms of evolution of states described by runs. A run of \mathcal{A} is a (possible infinite) sequence of S_0, S_1, S_2, \dots , where S_0 is an initial state, and every S_{i+1} is obtained from S_i by executing a transitional rule. In this paper we consider only runs in which states are global and agents' moves are atomic (instantaneous). The most general kind of runs for a distributed ASM are partially ordered runs. To prove properties of partially ordered runs, thanks to the results proved in [16], the attention may be restricted to their linearizations that are sequential runs and satisfy the following fundamental properties:

- All linearizations of the same finite initial segment of a run have the same final state.
- A property holds in every reachable state of a run iff it holds in every reachable state of any of its linearization.

Note that this implies that it is enough to find only one sequence of transitions and the runs that are considered here and start from the same initial state will have the same final state.

The main notions introduced in [23, Definition 5.1] are:

- *stable states* in a Chord network, where a state of a network is stable if the successor (predecessor) pointers of all nodes form an ordered ring, and
- *regular runs*, where a run is regular if it is a linearization such that nodes leave and enter the network only in stable states.

Having these notions, the paper [23] proves that the presented formalization of Chord consistently maintains the topological structure of rings and manipulates with distributed keys. In Section 4 we will explain that our model of *BAC19* satisfies the mentioned constraints and that results from [23] hold also for *BAC19*.

On the other hand, the papers [21, 22] recognize the fact that the search procedure of the Synapse protocol is not complete nor fully exhaustive. This is due to the fact that the lookup procedure of the Chord network can “skip” in routing some of the Synapse nodes, and thus not to spread the query to all networks that are reachable. To avoid this situation we redefine the lookup procedure with Algorithm 1, not to skip any node. This allows to achieve exhaustiveness and keep the routing complexity still be acceptable due to the relatively small number of contacts for each device.

4. System *BAC19*

In this section, we propose the design of the system *BubbleAntiCovid19 (BAC19)*, which is a formal Federation of Structural Overlay Networks for connecting different Digital Contact Tracing applications that are currently in use all over the world.

BAC19 federation (Figure 3) consists of several Chord networks:

- a network for each person/device of his/her first contacts (black circles in Figure 3),
- dedicated *red* network to connect all infected persons (red circle in Figure 3),
- dedicated *amber* network to connect all the first contacts of infected persons (amber circle in Figure 3),

and

- *Gateways* (black rectangles in Figure 3) as the connections to the existing Digital Contact Tracing systems (black rounded corners rectangles in Figure 3).

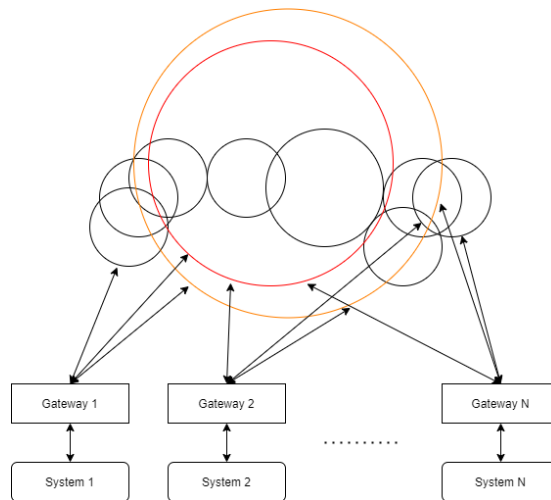


Fig. 3. *BAC19* Federation of Structural Overlay Networks

The first connection between the proposed extension and an existing system for contact tracing is called *Gateway*. The purpose of a *Gateway* is to maintain communication between two parts and to transform messages in a way that both sides can communicate efficiently.

The most important thing is to maintain the mappings between identifiers (IDs) used on both sides of a *Gateway*. As we could see in Section 2, some systems periodically change IDs, so the possibility to trace those changes is vital for functioning of *BAC19*. Regarding IDs, our goal is to have one identifier per one person/device regardless of how many systems it appears in. We argue that this is possible to achieve because of the scalability of Structured Overlay Networks. First, it is possible to use sufficiently large

Algorithm 1: FindSuccessor

```

FINDSUCCESSOR (KEY) =
For Given key
//successor(id(Me)) is responsible for key
if member_of(key, id(Me), successor(id(Me))) then
| Respond With successor(id(Me))
else
| //Me forwards query to its successor
| Forward Query To successor(id(Me))

```

codomain of the hash function (e.g. 2^{128} or greater). Also, it is possible to select enough parameters of a person/device so that it can be uniquely identified. We are not storing any other attribute of a person/device except a newly introduced identifier in our extension.

More precisely, with respect to the specifications that are provided in [22,23], we need to introduce the following changes:

- the set

$$Network = \{red, amber, net_1, \dots, net_N\}, \quad N \in \mathbb{N}$$

to denote all possible networks, where N is the number of possible persons/devices in the proposed extension;

- the set *Time* and the function

$$contact_time() : (Chord \cup \{amber\}) \times Chord \rightarrow Time$$

to denote the time of the contact between two persons;

- the external function *current_date()* to get the current date.

The algorithm FINDSUCCESSOR finds a responsible node for a given ID and it is originally defined in [31]. The lookup procedure of the Chord network can skip some of the synapse nodes, and thus not to spread the query to all networks that are reachable. To avoid this situation and to obtain *full exhaustiveness* of the retrieval procedure, the retrieval procedure is redefined with **Algorithm 1**, not to skip any node.

With this proposal we are not compromising performances of the extension by much. Since the number of contacts of a given person using the system is *relatively small*, it is manageable to allow increasing the complexity of the worst case retrieval from $O(\log N)$ to $O(N)$. In the predefined time-slots, our extension will receive the following information from a system:

- all identified infected cases since the last import (**Algorithm 2**),
- all confirmed cases that are not infected anymore since the last import (**Algorithm 3**),
- all identified contacts since the last import in the form of the tuple $\langle id_i, id_j, t \rangle$ ($i \neq j$) with the meaning that persons id_i and id_j had a risk contact at t timestamp. For the purpose of providing privacy protection timestamp should be kept at the precision of days. Unfortunately, this type of communication is not possible with the systems that are categorized as decentralized Bluetooth systems, since the fact that contact tracing computation is performed at users' devices and not shared with the central storage.

Algorithm 2: Put

```

PUT =
For all  $inf \in NewCases$ 
  Invoke PUT Of Network  $red$  To Store  $inf$ 
For all  $id \in net_{inf}$ 
  if  $contacttime(amber, id) < t$  or  $contacttime(amber, id) = undef$  then
    Set  $contacttime(amber, id) = t$ 

```

Algorithm 3: Leave

```

LEAVE =
For all  $inf \in Healed$ 
  Invoke LEAVE Of Network  $red$  for  $inf$ 

```

These systems can only share newly identified cases and their time of recovery (**Algorithm 4**).

Also, if needed, it is possible to introduce the new *Gateway* with the purpose to enter manually recognized contacts to the system.

When information is received from the origin systems, as the first step *BAC19* will connect all newly recognized infected cases to the *red* network, as well as to remove all cured. A node will remain in the *red* network until its recovering is confirmed. All IDs that are recognized as the risk contacts of a person/device (e.g. id_i) will be added to its bubble. They will stay there until $t + 14$ days, where t is the time of their contact and 14 days is a widely used time frame in Digital Contact Tracing implementations. If the id_i is the member of the *red* network all the members of its network will be added to the *amber* network and stay there during the same time frame $t + 14$ days. If a contact is already in the *amber* network, the timestamp will be updated to the higher value (**Algorithm 5**).

During the opposite way of communication, *BAC19* will pass on information to all nodes in the *amber* network to *Gateways*. If an identifier is recognized in the set of mappings for the particular origin system, the corresponding information is transferred to the origin system to alert (if not already) the person/device that he/she had risk contact with an infected person at stored timestamp. Also, *BAC19* is capable to send information on the second level contacts (the result is stored in the set *Result*, **Algorithm 6**).

Namely, for all nodes of the *amber* network it is possible to go through every origin bubble and pass those identifiers to the *Gateways*. Then the origin systems can inform those persons that they should increase their awareness since they are second level contacts.

Algorithm 4: Set contact time

```

SETCONTACTTIME =
par
  Set  $contact\_time(id_i, id_j) = t$ 
  Set  $contact\_time(id_j, id_i) = t$ 
endpar

```

Algorithm 5: Leave of network

```

LEAVE =
For all  $net_{id_i} \in Network \setminus \{red\}$ 
  For all  $id_j \in net_{id_i}$ 
    if  $contacttime(id_i, id_j) + 14 \text{ days} > currentdate()$  then
      Invoke LEAVE Of Network  $id_j$  for  $id_i$ 

```

Algorithm 6: Get all nodes

```

GET =
seq
  Invoke GET all nodes from amber and store the result in amber
  For all  $id \in amber$ 
    Invoke GET all nodes from  $net_{id}$  and append the result to Result
endseq

```

The “consistency” of the whole *BAC19* system is naturally related to the consistency of the overlay network it employs, namely Chord [32]. Informally, consistency refers to the fact that if a (key-value) is stored in *BAC19* than every query having a subject a stored key will route to the stored value; in other words if a (key-value) is memorised in some Chord node, then `FINDSUCCESSOR(KEY)` will succeed to route the query till the value. More formally, and according to Theorem IV.3 of [32], consistency is defined as follows: “if any sequence of JOIN operations are executed interleaved with stabilization” (performed by the fundamental Chord function `STABILIZE` that ran periodically in every node, checking and correcting that every node does not have *dangling* successor pointers), “then at some time after the last JOIN the successor pointers will form a *cycle* on all the nodes in the network”. This property, cited by Stoica *et al.* as *Inconsistency is a “transient” state* apply also, by construction, to our *BAC19* Federation network.

Using the results from [22, 23] we can prove the following statement:

Theorem 1 *The proposed extension stores and retrieves only consistent information on Covid-19 positive cases (identified by the origin systems) and their contacts and makes it available to all origin systems.*

Proof. It is shown in the paper [23] that it might happen that stable states in a Chord network cannot be achieved if Leave and/or Put algorithms are executed in the unstable states. Thus, to avoid that in *BAC19*, it is necessary to ensure that executions of each of Algorithm 2 to Algorithm 6 do not intertwine. Since all the nodes are under the control of our system *BAC19* we can assure that this issue won’t happen.

Executions of the proposed extension are performed in the controlled environment. Due to the scheduled time intervals for running different tasks, the nodes’ leaving from the bubbles will not happen during the unstable states, i.e., there will be only runs compatible with conditions of [23, Theorems 5.3, 5.4, and 5.7]. Also, the fact that the algorithm `FINDSUCCESSOR` is changed guarantees that all nodes will be contacted during the search procedure, and that the retrieving procedure of the Synapse protocol [22] is fully exhaustive.

As a consequence of the mentioned adaptations of the proposed Chord model, all possible execution of *BAC19* fulfill conditions from [23, Theorems 5.3, 5.4, and 5.7]. So, with these modifications starting from the given state *BAC19* will always reach the stable state, and the retrieved information will be consistent. \square

Consequently, *BAC19* solves the problem presented in the scenario with Alice and Bob, that is raised in the introduction in Problem 1.1. Let both System A and System B be part of *BAC19*. For Alice and Bob it will be sufficient that only one of them installed the system of the other region during the time of their travel, so that Alice gets informed that she is the first contact of an infected person, namely Bob during their joint travel.

5. Simulation Results

To better capture the relevance of our results, we have conducted simulations of *BAC19*. The purpose of conducted simulations is to show that the retrieving procedure of *BAC19* is fully exhaustive, like we stated in Theorem 1. The simulator is written in Python. It creates the *BAC19* topology, generates random requests for the search procedure and calculates the success rate. We expect that the simulator will not skip any node, so that the success rate of every iteration will be obviously 1.

The *BAC19* topology is created in three phases (see Figure 4). The simulator first randomly chooses which nodes will be in contact. When creating the network of each individual, special care was taken to ensure that “being in contact” is a symmetrical relation. Therefore, if persons P1 and P2 were in contact, then id of person P1 will be placed in the network of person P2, and id of person P2 will be placed in the network of person P1. According to [19] and [28], the number of close contacts of one person in a period of 14 days is between 30 and 40. For that reason, we have set in *BAC19* simulator the lower and the upper bound for the number of close contacts to be 30 and 40. Further, a *red* network is created by randomly choosing a certain percentage of nodes that will be infected. Finally, by going through the networks of all infected nodes and placing the results in *amber*, the *amber* network is created. For more details on *red* and *amber* network, please look at Figure 3 and Section 4.

The search requests were also generated completely at random. To achieve statistical significance for the performed simulations, for each configuration of the simulation we have generated 20 random networks and conducted 50 searches on every of them.

We have analyzed the results for two types of simulations:

- fixed network size (number of nodes) and variable percentage of infected nodes;
- variable network size and fixed percentage of infected nodes.

For the network size we chose among 1000, 2000 and 5000 nodes. For the percentage of infected nodes we chose among 1%, 5% and 10%. According to [5], only 13% of the world’s population had been infected by SARS-CoV-2 by the end of 2020, and the same trend seems to be confirmed in 2021. That is why we assume that percentage of infected nodes at a certain point of time can not be more than 10%.

The results from *BAC19* simulator show that the success rate of the retrieving procedures is constantly 1, and it does not depend on the network size or the percentage of infected nodes. The results for specific simulations can be seen in Figure 5. The first part

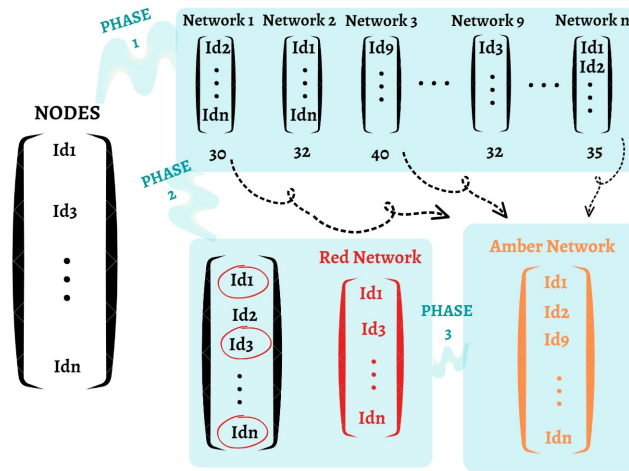


Fig. 4. Creation of *BAC19* topology

of the figure presents the results for a network with 5000 nodes. Although the percentage of infected nodes changes, the success rate remains the same. The second part of the figure presents the results for networks with 5% infected nodes. Although the size of the network changes, the success rate remains the same.

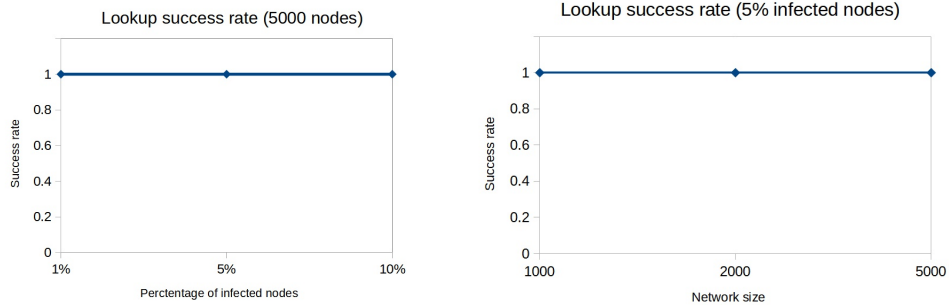


Fig. 5. Simulation results

By changing the original algorithm FINDSUCCESSOR from [31] in the way as it is presented in Algorithm 1, the search procedure slows down but the robustness of the model significantly increases. The simulation results confirm our expectations regarding the success of the search procedure, showing that the search procedure of *BAC19* is fully exhaustive.

6. Discussion

The paper [36] proposes building a common API. This approach is rather similar to the extension proposed in this paper. However, these approaches have also two significant differences:

- while [36] is building API connection points between each of two different origin systems that are connected, our extension proposes a version to common bus where each of the origin systems communicates with the proposed extension and in this way reduces and simplifies the number of connection points that needs to be maintained when several origin systems are connected;
- with *BAC19* we are simplifying also information that is being exchanged, and we do not violate privacy in the origin systems (since our extension does not collect information of an origin DCT system).

ETSI GS-E4P presents in [15] an interoperability framework for pandemic contact tracing systems which allows the centralized and decentralized modes of operation to fully interoperate.

European Community presents in [29] a guideline on interoperability specifications for cross-border transmission chains between approved apps, by using a Federation Gateway Service for synchronizing the diagnosis keys (keys of infected users) across backend servers of each national app. However, this approach focuses only on Google/Apple exposure notification apps because the majority of European countries have developed this kind of apps, and also because one Google/Apple exposure notification app can detect the contact with a user of another Google/Apple exposure notification app. In this paper we do not focus on a certain type of DCT apps, because we want to achieve the connection between them regardless the digital contact tracing technology employed and their system architecture.

However, [15, 29, 36] do not completely solve the problem presented in the scenario with Alice and Bob. Theorem 1 and simulation results show that the *BAC19* represents a much more convenient and efficient solution to this problem.

7. Conclusion and Further Work

In this paper we have presented *BAC19* a new and efficient Structured Overlay Network connecting existing systems for digital contact tracing. The advantages of *BAC19* (its usage) are:

- a person does not install anything new on his/her mobile device (except a new application which is used in the region that this person is visiting);
- the overlay does not store any personal sensitive information;
- the overlay is independent regarding how the origin system calculated contacts or is it based on Bluetooth or GPS technology;
- the overlay supports manual entry of recognized contacts;
- there are no new highly complicated calculations of possible contacts beside those that are performed by the original digital contact tracing systems.

The presented extension *BAC19* is the so-called digital forward tracing system, i.e. finding all direct contacts of an infected person. We plan in the future to explore the possibilities to adapt *BAC19* to also enable digital backward tracing system, i.e. finding the source of infection using contacts.

Furthermore, the development from a simulator to a real working system prototype should encompass the following steps:

- (i) to choose one of the multiple implementations of Chord, existing, e.g., in github,
- (ii) to transform the pseudocode and the simulation code into some real code,
- (iii) to embed this code into an Android app: to do this, Chord implementations in Java or Kotlin are worth to be chosen and “enriched” with our pseudocode, and
- (iv) to try to build a *proof of principle*, a.k.a. *proof of concept* of the whole *BAC19* system.

The proof of concept results should at least overlap our simulation results and so validate our *BAC19* system. The results could be presented in some standardization institute groups, e.g. ETSI ISG E4P⁵. Another potential future work is to make a systematic comparison/integration of our *BAC19* solution using/extending the GAEN, Apple-Google exposure notification solution [17], which, by construction, is fully decentralized. Scientific community should be aware of the fact that Covid-19 pandemics is not finished yet, and our *BAC19* system is fully compatible/parametric/polymorphic with any other kind of future virus and pandemics (e.g. Monkeypox⁶).

Acknowledgments. Authors would like to thanks anonymous referees for their useful comments that help to greatly improve the paper. This work was partly supported by the *Science Fund Republic of Serbia #6526707 AI4TrustBC*.

References

1. Abbas, R., Michael, K.: COVID-19 Contact Trace App Deployments: Learnings From Australia and Singapore. *IEEE Consumer Electronics Magazine* 9(5), 65–70 (2020)
2. Ahmed, N., Michelin, R.A., Xue, W., Ruj, S., Malaney, R., Kanhere, S.S., Seneviratne, A., Hu, W., Janicke, H., Jha, S.K.: A Survey of COVID-19 Contact Tracing Apps. *IEEE Access* 8, 134577–134601 (2020)
3. Altshuler, T.S., Hershkovitz, R.A.: Digital Contact Tracing and the Coronavirus: Israeli and Comparative Perspectives. In: *Foreign Policy at Brookings* (2020)
4. Apple, Google: (google/apple) exposure notification, aka, privacy-preserving contact tracing (2020), <https://www.google.com/covid19/exposurenotifications/>; <https://www.apple.com/covid19/contacttracing/>
5. Ayoub, H.H., Mumtaz, G.R., Seedat, S., Makhoul, M., Chemaitelly, H., Abu-Raddad, L.J.: Estimates of global sars-cov-2 infection exposure, infection morbidity, and infection mortality rates in 2020. *Global Epidemiology* 3, 100068 (2021), <https://www.sciencedirect.com/science/article/pii/S2590113321000225>
6. Basu, S.: Effective Contact Tracing for COVID-19 Using Mobile Phones: An Ethical Analysis of the Mandatory Use of the Aarogya Setu Application in India. *Cambridge Quarterly of Healthcare Ethics* 30(2), 262–271 (2021)

⁵ See ETSI Industry specification group “Europe for Privacy-Preserving Pandemic Protection” (ISG E4P) <https://www.etsi.org/committee-activity/activity-report-e4p>.

⁶ See <https://www.cdc.gov/poxvirus/monkeypox/about.html>.

7. Börger, E., Stärk, R.F.: Abstract State Machines. A Method for High-Level System Design and Analysis. Springer (2003), <http://www.springer.com/computer/swe/book/978-3-540-00702-9>
8. Camus, A.: La peste. Gallimard (1947)
9. Castelluccia, C., Bielova, N., Boutet, A., Cunche, M., Lauradoux, C., Le Métayer, D., Roca, V.: ROBERT: ROBust and privacy-presERving proximity Tracing (May 2020), <https://hal.inria.fr/hal-02611265>, working paper or preprint
10. Castelluccia, C., Bielova, N., Boutet, A., Cunche, M., Lauradoux, C., Métayer, D.L., Roca, V.: ROBERT (ROBust and privacy-presERving proximity Tracing protocol). Tech. rep., Inria and Fraunhofer AISEC (2020), <https://hal.inria.fr/hal-02611265>
11. Cheklat, L., Amad, M., Omar, M., Boukerram, A.: CHEARP: chord-based hierarchical energy-aware routing protocol for wireless sensor networks. *Comput. Sci. Inf. Syst.* 18(3), 813–834 (2021), <https://doi.org/10.2298/osis200308043c>
12. Cheng, X., Yang, H., Krishnan, A.S., Schaumont, P., Yang, Y.: KHOVID: interoperable privacy preserving digital contact tracing. *CoRR abs/2012.09375* (2020), <https://arxiv.org/abs/2012.09375>
13. El-Ansary, S., Haridi, S.: An overview of structured p2p overlay networks. *Handbook on Theoretical and Algorithmic Aspects of Sensor, Ad Hoc Wireless, and Peer-to-Peer Networks* (08 2005)
14. ETSI: Comparison of existing pandemic contact tracing systems. Tech. Rep. DGS E4P-002, ETSI (2021), https://www.etsi.org/deliver/etsi_gr/E4P/001_099/002/01.01.01_60/gr_E4P002v010101p.pdf
15. ETSI: Pandemic proximity tracing systems: Interoperability framework. Tech. Rep. DGS E4P-007, ETSI (2021), https://www.etsi.org/deliver/etsi_gs/E4P/001_099/007/01.01.01_60/gs_e4p007v010101p.pdf
16. Gurevich, Y.: Evolving algebras 1993: Lipari guide. In: Börger, E. (ed.) *Specification and validation methods*, pp. 9–36. Oxford University Press (1993)
17. Hoepman, J.H.: A Critique of the Google Apple Exposure Notification (GAEN) Framework. *ArXiv abs/2012.05097* (2020)
18. Huang, J., Yegneswaran, V., Porras, P., Gu, G.: On the privacy and integrity risks of contact-tracing applications (2020)
19. Keeling, M., Hollingsworth, T., Read, J.: The Efficacy of Contact Tracing for the Containment of the 2019 Novel Coronavirus (COVID-19). *J Epidemiol Community Health* (10 2020)
20. Kong, X., Qi, Y., Song, X., Shen, G.: Modeling disease spreading on complex networks. *Comput. Sci. Inf. Syst.* 8(4), 1129–1141 (2011), <https://doi.org/10.2298/CSIS110312061K>
21. Liquori, L., Tedeschi, C., Vanni, L., Bongiovanni, F., Ciancaglini, V., Marinkovic, B.: Synapse: A scalable protocol for interconnecting heterogeneous overlay networks. In: Crovella, M., Feeney, L.M., Rubenstein, D., Raghavan, S.V. (eds.) *NETWORKING 2010, 9th International IFIP TC 6 Networking Conference, Chennai, India, May 11-15, 2010. Proceedings. Lecture Notes in Computer Science*, vol. 6091, pp. 67–82. Springer (2010), https://doi.org/10.1007/978-3-642-12963-6_6
22. Marinković, B., Ciancaglini, V., Ognjanović, Z., Glavan, P., Liquori, L., Maksimović, P.: Analyzing the exhaustiveness of the synapse protocol. *Peer Peer Netw. Appl.* 8(5), 793–806 (2015), <https://doi.org/10.1007/s12083-014-0293-z>
23. Marinković, B., Glavan, P., Ognjanović, Z.: Proving properties of the chord protocol using the ASM formalism. *Theor. Comput. Sci.* 756, 64–93 (2019), <https://doi.org/10.1016/j.tcs.2018.10.025>
24. Marinković, B., Liquori, L., Ciancaglini, V., Ognjanović, Z.: A distributed catalog for digitized cultural heritage. In: Gusev, M., Mitrevski, P. (eds.) *ICT Innovations 2010 - Second International Conference, ICT Innovations 2010, Ohrid, Macedonia, September 12-15, 2010. Revised*

- Selected Papers. Communications in Computer and Information Science, vol. 83, pp. 176–186 (2010), https://doi.org/10.1007/978-3-642-19325-5_18
25. Marinković, B., Ognjanović, Z., Glavan, P., Kos, A., Umek, A.: Correctness of the chord protocol. *Comput. Sci. Inf. Syst.* 17(1), 141–160 (2020), <https://doi.org/10.2298/CSIS181115017M>
 26. Martin, T., Karopoulos, G., Hernández-Ramos, J.L., Kambourakis, G., Fovino, I.N.: Demystifying COVID-19 Digital Contact Tracing: A Survey on Frameworks and Mobile Apps. *Wireless Communications and Mobile Computing* 2020(8851429), 29 (2020), <https://www.hindawi.com/journals/wcmc/2020/8851429/>
 27. Maymounkov, P., Mazières, D.: Kademia: A peer-to-peer information system based on the XOR metric. In: Druschel, P., Kaashoek, M.F., Rowstron, A.I.T. (eds.) *Peer-to-Peer Systems, First International Workshop, IPTPS 2002, Cambridge, MA, USA, March 7-8, 2002, Revised Papers. Lecture Notes in Computer Science*, vol. 2429, pp. 53–65. Springer (2002), https://doi.org/10.1007/3-540-45748-8_5
 28. Mcaloon, C., Wall, P., Butler, F., Codd, M., Gormley, E., Walsh, C., Duggan, J., Murphy, T., Nolan, P., Smyth, B., O'Brien, K., Teljeur, C., Green, M., O'Grady, L., Culhane, K., Buckley, C., Carroll, C., Doyle, S., Martin, J., More, S.: Numbers of close contacts of individuals infected with SARS-CoV-2 and their association with government intervention strategies. *BMC Public Health* 21 (12 2021)
 29. eHealth Network: Interoperability specifications for cross-border transmission chains between approved apps (2020)
 30. Ocheja, P., Cao, Y., Ding, S., Yoshikawa, M.: Quantifying the privacy-utility trade-offs in COVID-19 contact tracing apps (2020)
 31. Stoica, I., Morris, R.T., Karger, D.R., Kaashoek, M.F., Balakrishnan, H.: Chord: A scalable peer-to-peer lookup service for internet applications. In: Cruz, R.L., Varghese, G. (eds.) *Proceedings of the ACM SIGCOMM 2001 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication, August 27-31, 2001, San Diego, CA, USA*. pp. 149–160. ACM (2001), <https://doi.org/10.1145/383059.383071>
 32. Stoica, I., Morris, R.T., Liben-Nowell, D., Karger, D.R., Kaashoek, M.F., Dabek, F., Balakrishnan, H.: Chord: a scalable peer-to-peer lookup protocol for internet applications. *IEEE/ACM Trans. Netw.* 11(1), 17–32 (2003), <https://doi.org/10.1109/TNET.2002.808407>
 33. Tang, Q.: Privacy-preserving contact tracing: current solutions and open questions (2020)
 34. Taylor, L., Sharma, G., Martin, A., Jameson, S. (eds.): *Data justice and COVID-19: Global perspectives*. Meatspace Press (aug 2020)
 35. Troncoso, C., et al.: *Decentralized Privacy-Preserving Proximity Tracing*. Tech. rep., École Polytechnique Fédérale de Lausanne, ETH Zurich, KU Leuven, Delft University of Technology, University College London, Helmholtz Centre for Information Security, University of Torino, ISI Foundation (2020), <https://github.com/DP-3T/documents/blob/master/DP3T%20White%20Paper.pdf>
 36. Vukolic, M.: On the interoperability of decentralized exposure notification systems. *CoRR abs/2006.13087* (2020), <https://arxiv.org/abs/2006.13087>

Silvia Ghilezan is a Professor of mathematics with the University of Novi Sad and Mathematical Institute of the Serbian Academy of Sciences and Arts. On several occasions she has held visiting positions at University of Oregon, École Normale Supérieure de Lyon, Université Paris Diderot - Paris 7, University of Turin, Radboud University and McGill University. The major lines of her research are in mathematical logic with application to programming languages, concurrency and mathematical linguistics. Her current research interests include formal methods for new challenges in privacy protection and artificial

intelligence. She has collaborated with over seventy co-authors on publications in leading scientific journals and conferences (POPL, LPAR, TLCA, PDP), books and editorials. She acts as a SC member of FSCD, an advisor for ARVR enterprises and industry, a popularizer of science and a promoter of gender balance in science. She was awarded the distinction Chevalier (2013) and Officier (2021) de l'Ordre des Palmes Académiques of the French Republic.

Simona Kašterović is a teaching assistant at the Faculty of Technical Sciences, University of Novi Sad. She received her B.Sc. degree at the Faculty of Sciences, University of Novi Sad in 2015. In 2017 she received her M.Sc. degree at the Faculty of Technical Sciences, University of Novi Sad. Currently, she is a Ph.D. student in applied mathematics at the same faculty. In 2018 she spent three months as visiting researcher at University Paris Diderot (Paris 7) in Paris, France. Her research interests include mathematical logic and its application in computer science: proof theory, lambda calculus, type theory; uncertain reasoning; probabilistic logic; computer assisted mathematical reasoning, formal methods for artificial intelligence.

Luigi Liquori got his MS in 1990 at Udine University, Italy. He got his Ph.D. in 1996 at University of Turin, Italy, and his H.d.R. in 2007 at Institut National Polytechnique de Lorraine, France. He served as Lecturer at the Ecole Nationale des Mines de Nancy from 1999. Since 2001, he is a senior researcher at French Institute for Research in Computer Science and Automation. His research's fields range from logics, type theory, lambda calculus, foundations of interactive proof assistants, to semantics of object oriented programming languages, until foundations of overlay networks, IoT protocols, and recently digital contact tracing against Covid-19.

Bojan Marinković got his PhD during 2014 at The Faculty of Technical Sciences University of Novi Sad, Serbia. Currently, he is Data Architect at Clarivate, Serbia. Until October 2018, he has been Research Assistant Professor at Mathematical Institute of the Serbian Academy of Sciences and Arts, however still interested in research in the following domains: distributed systems, applications of non-classical mathematical logic in computer science, and digitization of cultural and scientific heritage. During 2009, he spent three months as visiting researcher at Inria Sophia Antipolis, France.

Zoran Ognjanović is a research professor at the Mathematical Institute of the Serbian Academy of Sciences and Arts. He received his PhD degree in mathematical logic from University of Kragujevac, Serbia, in 1999. He has authored or coauthored two monographs, and a number of technical papers in major international journals and conferences. His research interests concern: applications of mathematical logic in computer science, artificial intelligence and uncertain reasoning, automated theorem proving, applications of heuristics to satisfiability problem, and digitization of cultural and scientific heritage. He is a recipient of the Serbian Academy of Sciences and Arts Award in the field of mathematics and related sciences for 2013 and the annual award of Serbian Ministry of Science for results in fundamental research in 2004.

Tamara Stefanović is a teaching assistant at the Faculty of Technical Sciences, University of Novi Sad, Serbia. She received her B.Sc. degree at the Faculty of Sciences,

University of Novi Sad in 2016. In 2019 she received her M.Sc. degree at the same faculty. Currently, she is a Ph.D. student in applied mathematics at the Faculty of Technical Sciences, University of Novi Sad. Her current scientific focus is on mathematical logic and its application in computer science, especially mathematical models for data privacy.

Received: August 25, 2021; Accepted: July 30, 2022.

Nearest Close Friend Query in Road-Social Networks

Zijun Chen*, Ruoyu Jiang, and Wenyuan Liu

¹ School of Information Science and Engineering, Yanshan University,
Qinhuangdao 066004, China

² The Key Laboratory for Computer Virtual Technology and
System Integration of Hebei Province,
Qinhuangdao 066004, China
zjchen@ysu.edu.cn
jiangruoyu@stumail.ysu.edu.cn
wylu@ysu.edu.cn

Abstract. Nearest close friend query ($k\ell$ NCF) in geo-social networks, aims to find the k nearest user objects from among the ℓ -hop friends of the query user. Existing efforts on $k\ell$ -NCF find the user objects in the Euclidean space. In this paper, we study the problem of nearest close friend query in road-social networks. We propose two methods. One is based on Dijkstra algorithm, and the other is based on IS-Label. For the Dijkstra-based method, Dijkstra algorithm is used to traverse the user objects needed. For the label-based method, we make use of IS-Label to calculate the distance between two vertices to avoid traversing the edges that do not contain the desired user object. For each method, we propose effective termination condition to terminate the query process early. Finally, we conduct a variety of experiments on real and synthetic datasets to verify the efficiency of the proposed methods.

Keywords: road-social networks, R-tree, IS-Label index, nearest neighbor query.

1. Introduction

With the development of location-aware smart devices, location-based applications have received extensive attention. Smart devices can allow users to obtain their own location information in location-based social networks, such as Foursquare, Facebook, Twitter, and Weibo.

The k -Nearest ℓ -Close Friends ($k\ell$ -NCF) query [22] retrieves the k nearest data objects to a query point p_q from among the ℓ -hop friends of a query user u . The $k\ell$ -NCF query is proposed in the Euclidean space. In real life, from the location of these returned data objects to the query point is limited by the road network. So, in this paper, we would propose the k -nearest neighbor ℓ -close friend query in road-social networks, which can also be applied in scenarios proposed in [22], such as making new friends, spatial crowdsourcing, blind dates, ridesharing, etc. We give one of the application scenarios in the following example.

Example 1. Spatial crowdsourcing. Spatial crowdsourcing is a platform, in which human workers can be assigned tasks related to a location. A requester q may issue a request to collect pictures in a specific location p_q . The worker who is assigned the task should be

* Corresponding author

close to p_q . To win the requester's trust, the worker should have acceptable social links to the requester. The 1-hop friends of the requester may be far away from p_q , or they could not accept the task. In this case, the requester may need to search the ℓ -hop friends of q .

In the road network shown in Fig. 1(a), there are four human workers B , C , D and E , who could accept the task. According to the social network shown in Fig. 1(b), q has two 1-hop friends B and F , and has 2-hop friends B , F , C and D . If only search 1-hop friends of q , F is the nearest neighbor of p_q . But F could not accept the task for some reason. B is far away from p_q . If search 2-hop friends of q , D is the nearest neighbor of q . In the Euclidean space, C is the nearest neighbor of p_q . But in the road network, C is farther away from p_q than D .

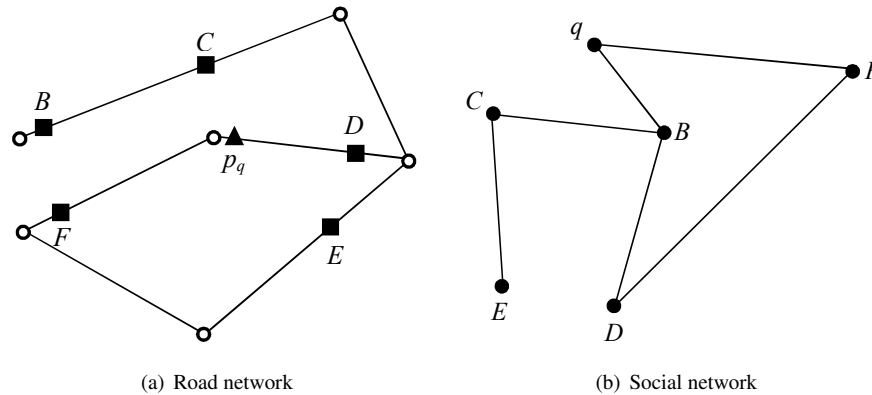


Fig. 1. Example of application scenario

Since the computation cost of the road network distance is much higher than that in Euclidean space, the methods proposed in [22] cannot be used to solve the problem of $k\ell$ -NCF in road-social networks directly.

For the $k\ell$ -NCF query in road-social networks, we will propose two methods. One is based on Dijkstra algorithm, and the other is based on IS-Label. For the Dijkstra-based method, Dijkstra algorithm [4] is used to traverse the user objects needed. The other is based on IS-Label, first use the R-Tree index to store the edges containing the user objects needed, then use the Best-First algorithm to search the edges. IS-Label is used to calculate the shortest distance between two vertices.

Our major contributions are summarized as follows.

- (1) We define the problem of $k\ell$ -NCF query in road-social networks.
- (2) We will propose two methods to tackle $RSk\ell$ -NCF query. One is based on Dijkstra algorithm, the other is based on IS-Label.
- (3) We conduct extensive experiments on different datasets to verify the efficiency of two algorithms.

The rest of the paper is organized as follows: Section 2 describes the related work and Section 3 formalizes the problem. Section 4 presents the method based on Dijkstra

algorithm. Section 5 presents the method based on IS-Label. Section 6 presents the experimental results and analysis. Finally, we conclude this paper in Section 7.

2. Related work

2.1. k NN query on road network

Roussopoulos et al. [21] designed a branch-and-bound R-tree [9] traversal algorithm to find the nearest neighbor object to the query point, and then generalize it to finding the k nearest neighbors (k NN). For k NN query on road networks, there are many studies. Hu et al. [12] simplified the network by replacing the graph topology with a set of interconnected tree-based structures called SPIE's, and proposed a lightweight nd index for the SPIE. Lee et al. [16] designed a new system framework ROAD for the spatial object search on road networks. Jiang et al. [14] studied the top- k nearest keyword search problem in a massive graph and proposed algorithms that return the exact answers. Inspired by R-tree, Zhong et al. [29] proposed a height-balanced and scalable index, namely G-tree, to efficiently support three types of location-based queries on road networks. Zhao et al. [28] studied the problem of group nearest compact POI set (GNCS) query and showed that this problem is NP-hard. Ouyang et al. [20] studied the problem of top- k nearest neighbors search on road networks. They proposed an efficient and progressive query processing algorithm to output each result in well-bounded delay. He et al. [10] proposed a framework on correctness-aware k NN queries, which aims at optimizing the system throughput while guaranteeing query correctness on moving objects. Dong et al. [5] presented a direction-aware KNN (DAKNN) query covering moving objects on road networks. Kim et al. [15] proposed the moving view field nearest neighbor (MVFNN) query, which continuously retrieves the nearest object in the query's view field with the change of query location.

2.2. Geo-Social query

Geo-social queries consider the location and social relationship. Liu et al. [18] proposed a new type of query called Circle of Friend Query (CoFQ), which returns a group of friends in a Geo-Social network whose members are close to each other both socially and geographically. Emrich et al. [6] studied the problem of geo-social skyline queries. The returned users are closely connected to the query user, and close to the query location. Ahuja et al. [1] proposed geo-social keyword (GSK) search, and presented three specific GSK queries. Jiang et al. [13] proposed the top- k local user search (TkLUS) query in geo-tagged social media. Sohail et al. [24] proposed Top- k Famous Places (T_k FP) query and Socio-Spatial Skyline Query (SSSQ). For the queries, they proposed three approaches, called Social-First, Spatial-First and Hybrid. There are also researches on group queries. Zhu et al. [30] proposed a family of geo-social group queries (GSGQs) with minimum acquaintance constraints, and devised two index structures, namely SaR-tree and SaR*-tree. Sohail et al. [25] proposed the Geo-Social Group preference Top- k (SG-Top $_k$) query, which retrieves nearby places popular among a particular group of users based on spatial and social relevance. Ma et al. [19] proposed the personalized geo-social group (PGSG) query, which aims to retrieve a user group and a venue, where each user in the group is socially connected with at least c other users in the group and the maximum distance of

all the users in the group to the venue is minimized. Ghosh et al.[8] proposed a novel Top k Flexible Socio Spatial Group Query (Top k -FSSGQ) to find the top k groups of various sizes w.r.t. multiple POIs. Shim et al. [23] proposed the ℓ -cohesive m -ridesharing group (ℓm CRG) query, which retrieves a cohesive ridesharing group by considering spatial, social, and temporal information.

2.3. Road-Social networks query

Road-Social networks query has drawn lots of attention in recent years. Zhao et al. [26] proposed the Reverse Top- k Geo-Social Keyword (RkGSK) query on road networks, and designed the GIM-tree index for the query. Attique et al.[2] proposed geo-social top- k keyword (GSTK) query and geo-social skyline keyword (GSSK) query on road networks. They proposed appropriate indexing frameworks and algorithms to efficiently process these queries. Zhao et al.[27] proposed the diversified top- k geo-social keyword (DkGSK) query on road networks, which considers not only the relevance but also the diversity of the result. Li et al.[17] studied the skyline cohesive group query problem in road-social networks.

In this paper, k -nearest neighbors ℓ -close friends ($k\ell$ -NCF) query in road-social networks is closely related to the research of [22]. Shim et al.[22] studied the $k\ell$ -NCF query and proposed three approaches for the query: Neighboring Cell Search, Friend-Cell Search, and Personal-Cell Search. The methods proposed in [22] cannot be directly applied to road networks. Therefore, we would study the problem of $k\ell$ -NCF query on road networks.

3. Problem definition

The road-social network is composed of a pair of networks, a road network G_r and a social network G_s , denoted as $G = (G_r, G_s)$. The road network is modeled as an undirected weighted graph $G_r = (V_r, E_r, W)$, where V_r is the vertex set, E_r is the edge set, and W is a function, such that $w(n_i, n_j)$ is the weight of edge $(n_i, n_j) \in E_r$. For $(n_i, n_j) \in E_r$, if the id of n_i is less than that of n_j , we call n_i and n_j the starting vertex and the ending vertex of (n_i, n_j) respectively, and vice versa. The social network is modeled as an undirected graph $G_s = (V_s, E_s)$, where V_s is the vertex set (representing users), and E_s is the edge set (representing social relations). The user objects in social network are mapped to the nearest intersection or edge on the road network based on their location. We use $rdist(a, b)$ to represent the shortest path length between a and b on the road network, where a or b could be a query point, vertex or user object.

Definition 1. (ℓ -hop friend list [22]) $V_v^\ell \subseteq V_s$ denotes the ℓ -hop friend list of v such that:

$$V_v^\ell = \begin{cases} \{v' | \exists e(v, v') \in E_s\} & (\ell = 1) \\ V_v^{\ell-1} \cup \{v' | \exists e(v', v'') \in E_s \wedge v'' \in V_v^{\ell-1}\} \setminus \{v\} & (\ell > 1) \end{cases} \quad (1)$$

Example 2. Fig. 2 describes the social network containing the user vertex v_0 . The 1-hop friend list of v_0 is v_1, v_3 . The 2-hop friend list of v_0 consists of v_2, v_4 , and the friends of v_0 , because v_2 and v_4 are the friends of v_1 . In the same way, we can get the 3-hop and 4-hop friends list of v_0 , which are shown in Fig. 2.

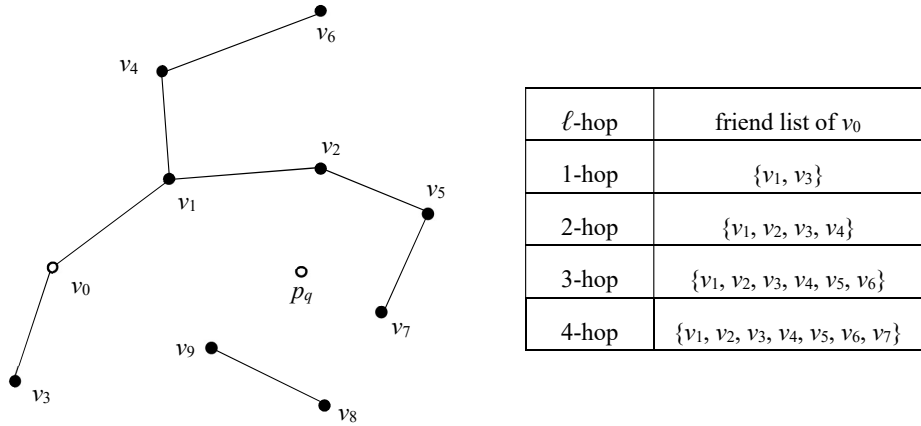


Fig. 2. Social network and ℓ -hop friend lists of v_0

The $k\ell$ -NCF query is defined in [22]. Based on this, we would give the definition of $k\ell$ -NCF query on road networks as follows:

Definition 2. (*$k\ell$ -NCF on road networks*) Given a road-social network $G = (G_r, G_s)$, a query point p_q , a query user u , the number of result elements k , and a friendship degree ℓ , the k -nearest ℓ -close friends ($k\ell$ -NCF) query $q = (p_q, u, k, \ell)$ on road networks finds a result list $R = (v_1, v_2, \dots, v_k)$, such that $(1 \leq i < k)$:

$$R \subseteq V_v^\ell \wedge rdist(p_q, v_i) \leq rdist(p_q, v_{i+1}) \wedge v' \in V_v^\ell \setminus R \wedge rdist(p_q, v_k) \leq rdist(p_q, v') \tag{2}$$

where, friendship degree represents the minimum number of edges (hops) between two data objects in the graph [22].

Example 3. After mapping, we get the road network shown in Fig. 3, where v_0, \dots, v_9 are the user objects shown in Fig. 2. Given a $k\ell$ -NCF query $q = (p_q, u = v_0, k = 2, \ell = 3)$ on road networks, the result list is (v_3, v_5) . Although v_7 is closer to p_q than v_5 , v_7 is not in the 3-hop friend list of v_0 , so v_7 is not in the result list.

4. Method based on Dijkstra algorithm

In this section, we will propose a method based on Dijkstra algorithm. Before the $k\ell$ -NCF query on road networks is issued, we can prepare some information to speed up query processing. We create adjacency lists for the social network and the road network respectively. For all the user objects in the social network, we create a hash table $UEHm$ with the structure $(v, (e, len))$, where e is the edge on which v locates, and len is the road network distance between v and the starting vertex of e .

In the following, we would introduce some data structures for the method.

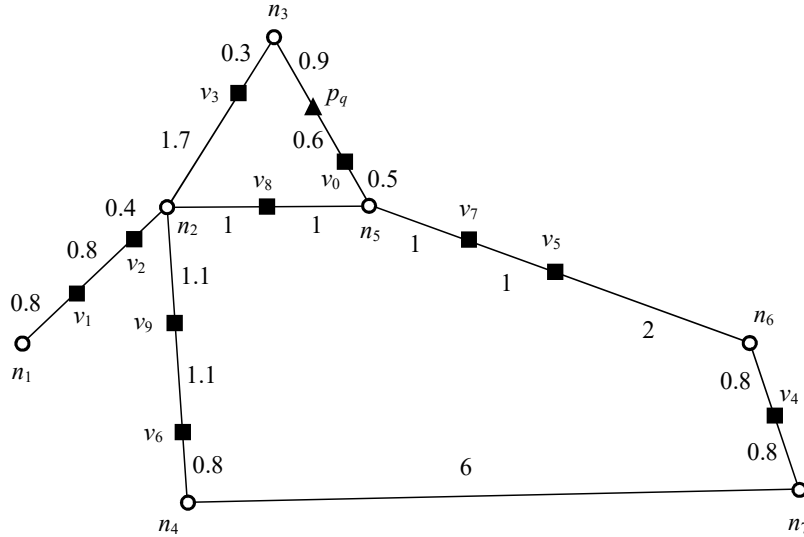


Fig. 3. Road network

closed: *closed* is a hash table to store the vertices which are closed. We call a vertex closed if it has been extracted from the min heap.

clounvis: *clounvis* is a hash table to store the vertices that are closed but not visited. We call a vertex visited if the user objects on all of its adjacent edges have been visited.

hE: *hE* is a hash table with the structure $(e, UList)$, where *UList* is a list to store the user object, such that the user object locates on the edge *e* and belongs to V_u^ℓ (*u* represents the query user).

The overall procedure of D-RSCNF is summarized as follows:

- (1) Create the ℓ -hop friend list V_u^ℓ of a query user *u*.
- (2) Create the hash table *hE*.
- (3) Dijkstra algorithm is used to expand from p_q .

Algorithm 1 describes the query process based on Dijkstra algorithm. Lines 1-2 are initialization. *R* is a max-heap with a maximum size of *k*, which is used to store the results. *H* is a min heap, in which the key is the road network distance from a vertex (or an object) to p_q . For *edgecount*, it is used to record the number of edges visited in *hE*. In lines 4-5, *hE* is created by using the user object $v \in V_u^\ell$ and the edge where the object locates.

In line 26, for the user object *v* on the edge (n_i, n_j) , we could calculate $rdist(p_q, v)$ with Formula (3). As shown in Fig. 4, *v* represents the required user object, *svid* is the starting vertex, and *tvid* is the ending vertex. The shortest path length from p_q to *v* in Algorithm 1 is defined as:

$$rdist(p_q, v) = \min\{rdist(p_q, svid) + d(svid, v), rdist(p_q, tvid) + d(tvid, v)\} \quad (3)$$

Algorithm 1: D-RSNCF Query

Input: $RSk\ell$ -NCF query $q = (p_q, u, k, \ell)$, the road-social network $G = (G_r, G_s)$, $UEHm$

Output: Result list R

```

1  $R \leftarrow \emptyset, H \leftarrow \emptyset, edgcount \leftarrow 0, count \leftarrow 0;$ 
2  $hE \leftarrow \emptyset, closed \leftarrow \emptyset, clounvis \leftarrow 0;$ 
3 compute  $V_u^\ell$  with the adjacency list of the social network  $G_s$ ;
4 for each user object  $v \in V_u^\ell$  do
5    $\lfloor$  find the edge  $e$  on which  $v$  locates using  $UEHm$  and update  $hE$ ;
6 get the edge  $(sqid, tqid)$  on which  $p_q$  locates;
7 insert  $(rdist(sqid, p_q), sqid)$  and  $(rdist(tqid, p_q), tqid)$  into  $H$ ;
8 while  $H \neq \emptyset$  do
9    $(rdist(n_i, p_q), n_i) \leftarrow H.delMin();$ 
10   $closed[n_i] \leftarrow n_i;$ 
11   $clounvis[n_i] \leftarrow rdist(n_i, p_q);$ 
12   $flag \leftarrow 0;$ 
13  for each adjacent vertex  $n_j$  of  $n_i$  do
14    if  $n_j$  is not in  $closed$  then
15       $flag \leftarrow 1;$ 
16      if  $n_j$  does not exist in  $H$  then
17         $\lfloor H.add(rdist(n_j, p_q), n_j);$ 
18      else
19         $\lfloor$  update  $(rdist(n_j, p_q), n_j)$  in  $H$ ;
20    else
21      if all the adjacent vertices of  $n_j$  are in  $closed$  then
22         $\lfloor$  remove  $n_j$  from  $clounvis$  table;
23      if edge  $(n_i, n_j)$  is in  $hE$  then
24         $edgcount ++;$ 
25        for each user object  $v$  in  $hE[(n_i, n_j)]$  do
26           $\lfloor R.add(rdist(p_q, v), v);$ 
27           $count ++;$ 
28        if  $hE.size() \leq edgcount$  then
29           $\lfloor R$  is sorted in ascending order by the value of  $rdist()$ ;
30           $\lfloor$  return  $R$ ;
31        if  $count \geq k$  then
32           $\lfloor$  if  $R.getRoot().rdist \leq \min_{rdist}(clounvis)$  then
33             $\lfloor R$  is sorted in ascending order by the value of  $rdist()$ ;
34             $\lfloor$  return  $R$ ;
35  if  $flag = 0$  then
36     $\lfloor$  remove  $n_i$  from  $clounvis$  table;
37  $R$  is sorted in ascending order by the value of  $rdist()$ ;
38 return  $R$ ;
```

where $d(svid, v)$ represents the length from $svid$ to v on the edge $(svid, tvid)$, and $d(tvid, v)$ represents the length from $tvid$ to v on the edge $(svid, tvid)$.

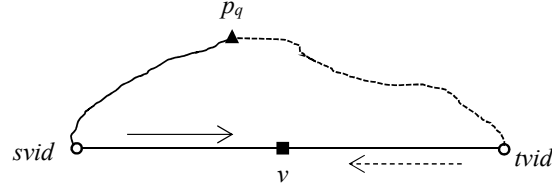


Fig. 4. Paths to be chosen for method based on Dijkstra algorithm

In lines 8-36, Dijkstra algorithm is used to expand. For an object v on the edge $(svid, tvid)$, in order to calculate $rdist(p_q, v)$, both $svid$ and $tvid$ need to be closed before calculation. If a vertex is closed, it needs to enter *closed* table and *clounvis* table (lines 10-11). If all adjacent vertices of a vertex have been closed, the vertex can be removed from *clounvis* table. In line 12, we use the value of *flag* as 0 to indicate that all adjacent vertices of n_i have been closed. If one of adjacent vertices of n_i is not closed, *flag* is set to 1 in line 15. In line 19, the update is to find the shortest path length from p_q to n_j and store it in H .

For a vertex in *clounvis* table, in order to remove it from *clounvis* table, there are two cases: (1) it is closed after all its adjacent vertices; (2) it is not closed after all its adjacent vertices. For case (1), in lines 35-36, n_i is removed from *clounvis* table, where n_i is closed after all its adjacent vertices. For case (2), in lines 21-22, n_j is removed from *clounvis* table, where n_j is closed before its adjacent vertex n_i .

In the following theorem, we would prove the correctness of the termination condition in lines 31-32 of Algorithm 1. In the condition, *count* is the size of R , $R.getRoot().rdist$ is the maximum distance in R and $min_{rdist}(clounvis)$ is the minimum distance in *clounvis*.

Theorem 1. *If $count \geq k$ and $R.getRoot().rdist \leq min_{rdist}(clounvis)$, then Algorithm 1 could sort and return R to terminate the query correctly.*

Proof. From Algorithm 1, we can see that *clounvis* is used to store the vertices that are closed but not visited. In order to terminate the query, we should focus on the unvisited user objects. For any unvisited user object v locating on edge (a, b) , there are two cases: (1) a or b is closed; (2) Neither a nor b is closed.

For case (1), without loss of generality, let a be in *clounvis*. Since b is not closed, $rdist(p_q, b) \geq rdist(p_q, a)$. Then, we have $rdist(p_q, v) \geq rdist(p_q, a) \geq min_{rdist}(clounvis)$. Therefore, if $count \geq k$ and $R.getRoot().rdist \leq min_{rdist}(clounvis)$, we have $rdist(p_q, v) \geq R.getRoot().rdist$, which indicates that v cannot or need not replace the user object in R .

For case (2), we have $rdist(p_q, v) \geq \min(rdist(p_q, a), rdist(p_q, b)) \geq max_{rdist}(clounvis) \geq min_{rdist}(clounvis)$, where $max_{rdist}(clounvis)$ is the maximum distance in *clounvis*. Therefore, if $count \geq k$ and $R.getRoot().rdist \leq min_{rdist}(clounvis)$, we have $rdist(p_q, v) \geq R.getRoot().rdist$. ■

Example 4. For Fig. 3, given a query $q = (p_q, u = v_0, k = 2, \ell = 2)$, we illustrate the query process of Algorithm 1. According to the social network in Fig. 2, the 2-hop friend list of v_0 is $\{v_1, v_2, v_3, v_4\}$. Then the hash table hE is created. Based on $UEHm$, we get the edges on which the 2-hop friend list of v_0 locate. The edges are $(n_1, n_2), (n_2, n_3), (n_6, n_7)$, which are stored in hE . And get the edge (n_3, n_5) where p_q locates, so $(0.9, n_3)$ and $(1.1, n_5)$ are put into the min-heap H .

(1) The first removed from H is $(0.9, n_3)$, and n_3 is put into *closed* and *clounvis*. Then we find the adjacent vertices of n_3 , and $(2.9, n_2)$ is put into H . Note that $(2.9, n_5)$ would not replace $(1.1, n_5)$ in H .

(2) The second removed from H is $(1.1, n_5)$. Similarly, n_5 is put into *closed* and *clounvis*. Then we find the adjacent vertices of n_5 and $(5.1, n_6)$ is put into H . At the moment, n_3 and n_5 are in *closed*, and the edge (n_3, n_5) is not in hE , so we can conclude that there is no 2-hop friend of v_0 on this edge.

(3) The third removed from H is $(2.9, n_2)$, and n_2 is stored in *closed* and *clounvis*. We find the adjacent vertices of n_2 , then $(4.9, n_1)$ and $(5.9, n_4)$ are put into H . For the adjacent vertices of n_2, n_3 and n_5 are in *closed*. For the adjacent vertices of n_3, n_2 and n_5 are in *closed*, so n_3 will be removed from *clounvis*. Since the edge (n_2, n_3) is in hE , we find the user object v_3 on (n_2, n_3) , and calculate $\min_{rdist}(p_q, v_3)$ according to Formula (3). So $(1.2, v_3)$ is added to R .

(4) The fourth removed from H is $(4.9, n_1)$, and n_1 is put into *closed* and *clounvis*. Since the adjacent vertex n_2 of n_1 is in *closed*, we find the user objects v_1 and v_2 on the edge (n_1, n_2) . Calculate $\min_{rdist}(p_q, v_1)$ and $\min_{rdist}(p_q, v_2)$ according to Formula (3). So $(3.3, v_2)$ is stored in R . Because n_1 has only one adjacent vertex n_2 , and n_2 is in *closed*, we can remove n_1 from *clounvis*.

(5) The fifth removed from H is $(5.1, n_6)$, and n_6 is put into *closed* and *clounvis*. For the adjacent vertex n_7 of n_6 , $(6.7, n_7)$ is put into H . Since the adjacent vertex n_5 of n_6 is in *closed*, and the edge (n_5, n_6) is not in hE , no user object is found. Because all the adjacent vertices of n_5 are in *closed*, we can remove n_5 from *clounvis*.

(6) The sixth removed from H is $(5.9, n_4)$, and n_4 is put into *closed* and *clounvis*. Since the adjacent vertex n_2 of n_4 is in *closed*, and the edge (n_2, n_4) is not in hE , no user object is found. Because all the adjacent vertices of n_2 are in *closed*, we can remove n_2 from *clounvis*. Now, we have $count \geq 2$ and $R.getRoot().rdist \leq \min(clounvis)$, where $R.getRoot().rdist = \min_{rdist}(p_q, v_2) = 3.3$ and $\min(clounvis) = 5.1$, so we can terminate the query. The process of this example is shown in Table 1.

Table 1. The process of Example 4

Order	$H.delMin()$	<i>closed</i>	<i>clounvis</i>	R
1	$(0.9, n_3)$	$\{n_3\}$	$\{(n_3, 0.9)\}$	\emptyset
2	$(1.1, n_5)$	$\{n_3, n_5\}$	$\{(n_3, 0.9), (n_5, 1.1)\}$	\emptyset
3	$(2.9, n_2)$	$\{n_3, n_5, n_2\}$	$\{(n_5, 1.1), (n_2, 2.9)\}$	$\{(1.2, v_3)\}$
4	$(4.9, n_1)$	$\{n_3, n_5, n_2, n_1\}$	$\{(n_5, 1.1), (n_2, 2.9)\}$	$\{(1.2, v_3), (3.3, v_2)\}$
5	$(5.1, n_6)$	$\{n_3, n_5, n_2, n_1, n_6\}$	$\{(n_2, 2.9), (n_6, 5.1)\}$	$\{(1.2, v_3), (3.3, v_2)\}$
6	$(5.9, n_4)$	$\{n_3, n_5, n_2, n_1, n_6, n_4\}$	$\{(n_6, 5.1), (n_4, 5.9)\}$	$\{(1.2, v_3), (3.3, v_2)\}$

Time complexity of Algorithm 1: First, create the ℓ -hop friend list V_u^ℓ for user u , which is equivalent to a breadth-first search process. So, it takes at most $O(|V_s| + |E_s|)$ to find V_u^ℓ . Next, it takes at most $O(|V_s|)$ to create hE . Then, it takes at most $O(|E_r|)$ to find the edge on which p_q locates. The top- k nearest user objects are found with the help of Dijkstra algorithm, so, the time cost is $O((|V_r| + |E_r|) \cdot \log|V_r| + |V_s| \cdot t_o)$, where t_o is the time used to compute the shortest distance between p_q and the user object according to Formula (3). To sum up, the time complexity of Algorithm 1 is $O(|V_s| + |E_s| + (|V_r| + |E_r|) \cdot \log|V_r|)$.

5. Method based on IS-Label

Because the method based on Dijkstra algorithm traverses the road network from near to far, it is not very advantageous that the user object found by query is far from the query point. Based on this situation, we propose a label-based method. IS-label index [7] is one of the label indexes, and it is also applicable to large graphs. So, we use IS-label index to calculate the minimum distance between p_q and the vertex.

Algorithm 2: L-RSNCF Query

Input: RSkI-NCF query $q = (p_q, u, k, \ell)$, the road-social network $G = (G_r, G_s)$, $UEHm$, IS-Label Index of G_r

Output: Result list R

- 1 $R \leftarrow \emptyset, hE \leftarrow \emptyset, count \leftarrow 0;$
- 2 $Queue \leftarrow \text{NewPriorityQueue}();$
- 3 compute V_u^ℓ with the adjacency list of the social network $G_s;$
- 4 **for** each user object $v \in V_u^\ell$ **do**
- 5 \lfloor find the edge e on which v locates using $UEHm$ and update $hE;$
- 6 create the R-Tree $index$ for all edges in $hE;$
- 7 get the edge $(sqid, tqid)$ where p_q locates;
- 8 $Queue.Enqueue(index.Root, MinDist(p_q, index.Root));$
- 9 **while** not $Queue.isEmpty()$ **do**
- 10 $temp \leftarrow Queue.Dequeue();$
- 11 **if** $temp$ is an object **then**
- 12 **for** each user object v on the edge $temp$ **do**
- 13 $R.add(rdist(p_q, v), v);$
- 14 $count ++;$
- 15 **if** $count \geq k$ **then**
- 16 **if** $R.getRoot().rdist \leq Queue.getRoot().edist$ **then**
- 17 \lfloor break;
- 18 **else**
- 19 **for** each child c of $temp$ **do**
- 20 \lfloor $Queue.Enqueue(c, MinDist(p_q, c));$
- 21 R is sorted in ascending order by the value of $rdist();$
- 22 return $R;$

Algorithm 2 describes the query process based on IS-Label index, which adopts the best-first traversal [11]. In line 1, R is the same as that in Algorithm 1. In line 2, $Queue$ is a min heap, in which the key is the minimum Euclidean distance from a node or an object to p_q (Definition 3). Lines 3-5 is the same as lines 3-5 in Algorithm 1. Line 6 is to create the R-Tree $index$ with the edges in hE . In lines 8-20, best-first traversal is used to search user objects with the R-Tree $index$. In lines 11-17, if the element removed from the priority queue $Queue$ is an edge, we access the user objects on this edge, and calculate the minimum distance between p_q and user objects using the IS-Label index. In lines 15-17, we could jump out of the while loop early, which is proved in Theorem 2. In lines 19-20, we process each child c of $temp$.

We should note that for each edge (a, b) on the road network, we create an object with Minimum Bounding Rectangle (MBR) containing a and b for the R-Tree $index$.

Definition 3. (*MinDist Distance [21]*) In Euclidean space of dimension n , the minimum distance between a point q and MBR $N(s, u)$ is denoted by $MinDist(q, N(s, u))$, which is defined as follows:

$$MinDist(q, N) = \sum_{i=1}^n |q_i - r_i|^2, r_i = \begin{cases} s_i, & q_i < s_i \\ u_i, & q_i > u_i \\ q_i, & \text{otherwise} \end{cases} \quad (4)$$

As shown in Fig. 5, $sqid$ and $tqid$ are the starting vertex and ending vertex of the edge on which p_q locates respectively. For the user object v , $svid$ and $tvid$ are the starting vertex and ending vertex of the edge on which v locates respectively. The shortest path length from p_q to v in Algorithm 2 is defined as:

$$rdist(p_q, v) = \min\{rdist1(p_q, v), rdist2(p_q, v), rdist3(p_q, v), rdist4(p_q, v)\} \quad (5)$$

where

$$\begin{aligned} rdist1(p_q, v) &= d(p_q, sqid) + rdist(sqid, svid) + d(svid, v) \\ rdist2(p_q, v) &= d(p_q, tqid) + rdist(tqid, tvid) + d(tvid, v) \\ rdist3(p_q, v) &= d(p_q, sqid) + rdist(sqid, tvid) + d(tvid, v) \\ rdist4(p_q, v) &= d(p_q, tqid) + rdist(tqid, svid) + d(svid, v) \end{aligned}$$

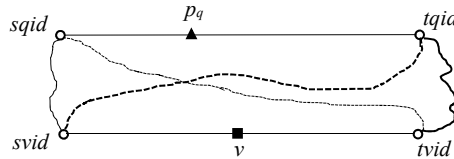


Fig. 5. Paths to be chosen for method based on IS-Label

In the following theorem, we would prove the correctness of the termination condition in lines 15-16 of Algorithm 2. In the condition, $count$ is the size of R , $R.getRoot().rdist$

is the maximum distance in R and $Queue.getRoot().edist$ is the minimum Euclidean distance in $Queue$.

Theorem 2. *If $count \geq k$ and $R.getRoot().rdist \leq Queue.getRoot().edist$, then Algorithm 2 could sort and return R to terminate the query correctly.*

Proof. R is a max-heap with a maximum size of k , so $R.getRoot().rdist$ is the maximum distance in R . $Queue$ is a min heap, $Queue.getRoot().edist$ is the minimum Euclidean distance in $Queue$. In order to terminate the query, we should focus on the unvisited user objects. For any unvisited user object v locating on the edge (a, b) , we have $rdist(p_q, v) \geq \min(rdist(p_q, a), rdist(p_q, b)) \geq Queue.getRoot().edist$. Therefore, if $count \geq k$ and $R.getRoot().rdist \leq Queue.getRoot().edist$, we have $rdist(p_q, v) \geq R.getRoot().rdist$, which shows that v cannot or need not replace the user object in R . ■

Example 5. For Fig. 3, given a query $q = (p_q, u = v_0, k = 2, \ell = 2)$, we illustrate the query process of Algorithm 2. The 2-hop friend list of v_0 and hE are the same as those in Example 4. Next, we create an R-Tree index for all edges in hE . The MBR of the edge in hE is shown in Fig. 6 using dashed rectangle.

(1) The first object to be removed from $Queue$ is (n_2, n_3) , and we find the user object v_3 on the edge (n_2, n_3) . Then we calculate $rdist(p_q, v_3)$. There are four paths from p_q to v_3 , which are $p_q \rightarrow n_3 \rightarrow n_2 \rightarrow v_3$, $p_q \rightarrow n_5 \rightarrow n_3 \rightarrow v_3$, $p_q \rightarrow n_3 \rightarrow n_3 \rightarrow v_3$, $p_q \rightarrow n_5 \rightarrow n_2 \rightarrow v_3$. $rdist(n_2, n_3)$, $rdist(n_2, n_5)$, $rdist(n_3, n_5)$ can be calculated using the IS-Label index. According to Formula (5), we get $rdist1(p_q, v_3) = 4.6$, $rdist2(p_q, v_3) = 3.4$, $rdist3(p_q, v_3) = 1.2$, $rdist4(p_q, v_3) = 4.8$. So we get $rdist(p_q, v_3) = rdist3(p_q, v_3) = 1.2$. Now we have $R = \{(1.2, v_3)\}$.

(2) The second object to be removed from $Queue$ is (n_1, n_2) , and we find v_1 and v_2 on the edge (n_2, n_3) . Using the IS-Label index, we get $rdist(p_q, v_1) = 4.1$ and $rdist(p_q, v_2) = 3.3$. Now we have $R = \{(1.2, v_3), (3.3, v_2)\}$.

(3) When (n_6, n_7) is the root of $Queue$, we have $count \geq 2$ and $R.getRoot().rdist \leq Queue.getRoot().edist$. Then we can jump out of the loop early, although v_4 on the edge (n_6, n_7) has not been processed.

Time complexity of Algorithm 2: In Algorithm 2, we use the traversal of R-Tree to replace the traversal of the road network using Dijkstra algorithm. According to [11], it takes at most $O(|E_r| \cdot \log|E_r|)$ to traverse R-Tree. The time of other parts are similar to that of Algorithm 1, so the time complexity of Algorithm 2 is $O(|V_s| + |E_s| + |E_r| \cdot \log|E_r|)$.

6. Experiments

6.1. Datasets and setting

This experiment uses two social network datasets and three road network datasets for testing. The social network datasets are Brightkite(BR) and Gowalla(GA) [3]. They come from <http://snap.stanford.edu/data/>. There are three road network datasets: (1)BAY; (2)San Francisco (SF); (3)City of San Joaquin County (TG). BAY comes from <http://users.diag.uniroma1.it/challenge9/download.shtml>. SF and TG come from

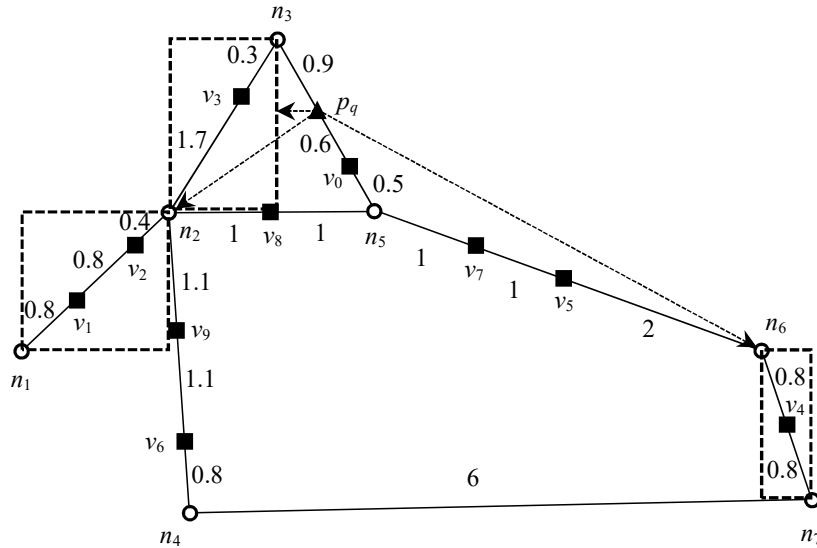


Fig. 6. Road network with MBRs

lifeifei/SpatialDataset.htm. The information of the road networks are shown in Table 2. Standardize the latitude and longitude of the user locations in the two social networks into a flat two-dimensional space, and then map the user to the nearest intersection or edge on the road network according to the coordinates. BR_rangeBAY and GA_rangeBAY represents the social network dataset of Brightkite and Gowalla within the range of BAY respectively. The information of the social networks are shown in Table 3. We use two real datasets:(1) BR_rangeBAY + BAY;(2) GA_rangeBAY + BAY. BR_rangeBAY + BAY represents the road-social network formed by BR_rangeBAY and BAY. GA_rangeBAY + BAY represents the road-social network formed by GA_range BAY and BAY. In these two real data sets, the user objects are sparse, so we also use synthetic datasets for testing.

The synthetic datasets retain the number of vertices in the two social networks and friendship relationship between users. The uniform function and the Zipf function are used to randomly allocate the location information of all user vertices, and the range of the horizontal and vertical coordinates is [0, 10000]. Uniform distribution is used for BR and GA in Fig. 11(a) and (c) respectively. Zipf distribution is used for BR and GA in Fig. 11(b) and (d) respectively. We get four synthetic datasets: (1)UBR+TG; (2)ZBR + TG; (3)UGA + SF; (4)ZGA + SF. UBR+TG represents the dataset formed by BR and TG, where uniform distribution is used for BR. ZBR+TG represents the dataset formed by BR and TG, where Zipf distribution is used for BR. UGA + SF represents the dataset formed by GA and SF, where uniform distribution is used for GA. ZGA + SF represents the dataset formed by GA and SF, where Zipf distribution is used for GA. Table 4 shows the density of road-social networks, where the density denotes the ratio of the number of vertices in the social network to the number of edges on the road network.

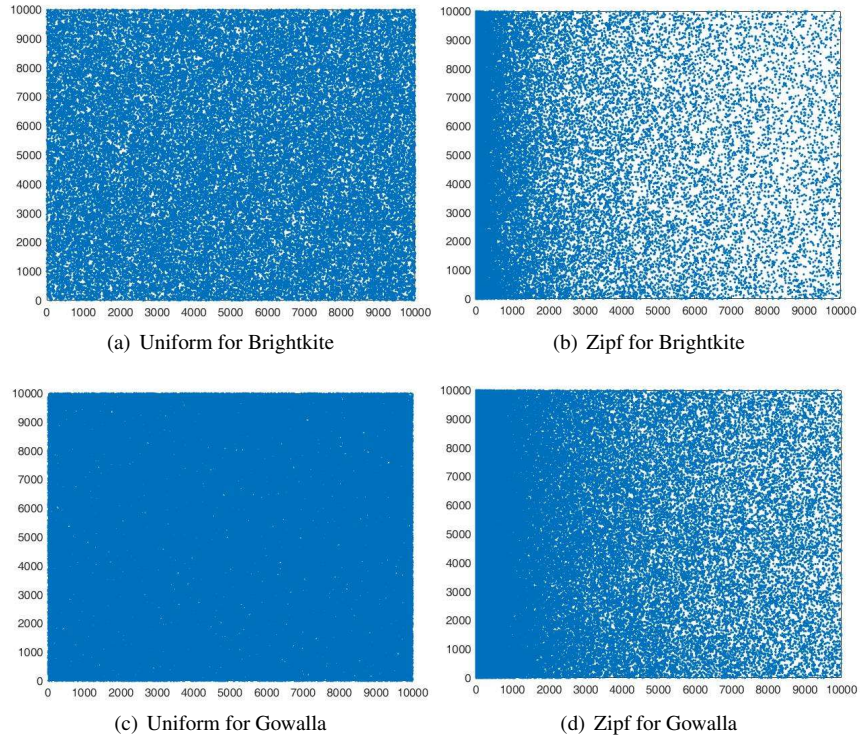


Fig. 7. Data object distributions of the synthetic datasets

Table 2. Statistics of the road network datasets

Road network	Vertices	Edges
BAY	321270	400086
SF	174956	223001
TG	18263	23874

Table 3. Statistics of the social network datasets

Social network	Vertices	Edges
BR_rangeBAY	2756	3819
GA_rangeBAY	4794	11086
Brightkite(BR)	58228	214078
Gowalla(GA)	196591	950327

Table 4. Density of the road-social network datasets

Road-social network	Density
BR_rangeBAY + BAY	0.007
GA_rangeBAY + BAY	0.012
UBR + TG	2.439
ZBR + TG	2.439
UGA + SF	0.882
ZGA + SF	0.882

Implementation: We implement all the algorithms on the Eclipse platform using Java. The experimental machine configuration is the Windows 10 operating system, Intel(R) Core(TM) i5-10500 CPU @ 3.10GHz and 8G RAM. In this experiment, we measure the average value at each experiment performed 100 times with random query users and vertices.

Parameters setting: Parameters setting are shown in Table 5, where ℓ is the friendship degree, k is the number of results, and rd is the Euclidian distance between the query point location p_q and the location of the query user u . The length unit of the data set BR.RangeBAY + BAY and GA.RangeBAY + BAY is kilometer. The social network datasets of the other four datasets are synthesized based on the range of the road network TG and SF. The coordinate range of TG and SF is [0, 10000], and the range of parameter rd for the synthetic datasets is [100, 200], [300, 400], [700, 800], [1500, 1600], [3100, 3200], and the default value is [1500, 1600].

Table 5. Parameters setting

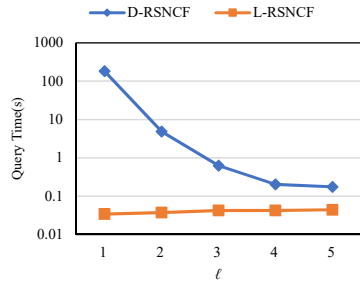
Parameter	Range	Default
ℓ	1, 2, 3, 4, 5	3
k	10, 20, 30, 40, 50	20
rd (for real dataset)	[1, 2], [3, 4], [7, 8], [15, 16], [31, 32]	[15, 16]

6.2. Performance Evaluation

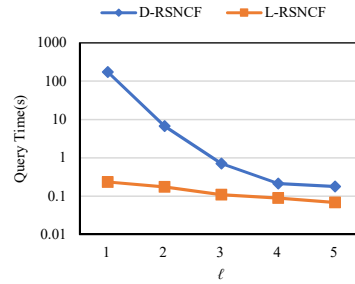
In order to test the effect of the experiment, $RSk\ell$ -NCF query algorithm based on Dijkstra (D-RSNCF) and $RSk\ell$ -NCF query algorithm based on IS-Label index (L-RSNCF) were compared on the datasets of six road-social networks.

(1) Effect of friendship degree ℓ on query time

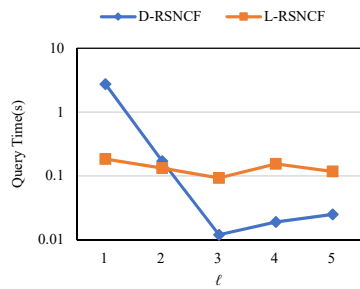
Fig. 8 demonstrates the effect of varying ℓ . In Fig. 8(a), we set $k = 10$. As shown in Fig. 8(a) and (b), with the increase of ℓ , the query speed of D-RSNCF algorithm is accelerated. This is because with the increase of ℓ , more user objects will be added to the ℓ -hop friend list of the query user u . Then the query range will become smaller. The query speed of L-RSNCF is much faster than that of D-RSNCF. The road network used in Fig. 8(a) and (b) is the largest in the three road network data sets, while few user objects



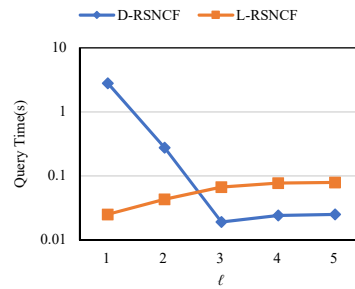
(a) BR_rangeBAY + BAY



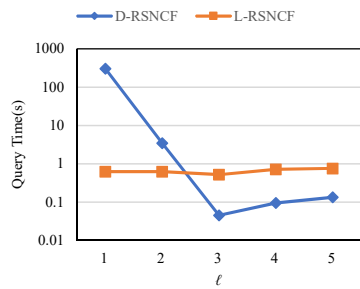
(b) GA_rangeBAY + BAY



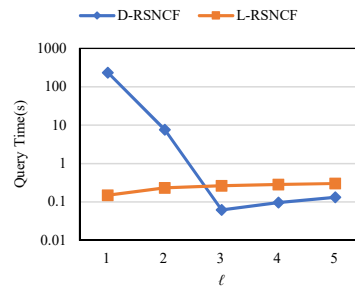
(c) UBR + TG



(d) ZBR + TG



(e) UGA + SF



(f) ZGA + SF

Fig. 8. Effect of ℓ on query time

are mapped to the road network. This situation has a greater impact on D-RSNCF query algorithm, so D-RSNCF will be much slower in this case. From Fig. 8(a) and (b), we can see that ℓ has little effect on L-RSNCF.

As shown in Fig. 8(c) and (d), when ℓ reaches a certain value, D-RSNCF is faster than L-RSNCF. The reason may be that UBR + TG and ZBR + TG have the largest density in Table 4. Then D-RSNCF may traverse less edges for these two datasets than other datasets. For larger ℓ , there may be more user objects that meet the friendship degree. So with ℓ increasing, D-RSNCF may traverse less edges. In Fig. 8(e) and (f), the experimental effect is similar to that shown in Fig. 8(c) and (d).

(2) Effect of numbers of result k on query time

Fig. 9 demonstrates the effect of varying k . In general, the query time of both methods increases with the increase of k . The reason is that both methods need to traverse more edges with k increasing. We can see that the influence of k on L-RSNCF is not obvious. In Fig. 9(c)-(f), D-RSNCF is faster than L-RSNCF in most cases, which is similar to the case in Fig. 8(c)-(f). From Fig. 8 and 9, we can see that for the six datasets, the higher the density of the data set, the shorter the query time, and vice versa.

(3) Effect of distance rd on query time

Fig. 10 shows the efficiency of the two query algorithms by changing the parameter rd . As shown in Fig. 10(a) and (b), the query time of D-RSNCF increases with the increase of rd . With rd increasing, most of the ℓ -hop friends of the query user u may be far away from the location of p_q . D-RSNCF is based on Dijkstra algorithm, so it will traverse more edges to find the result in the ℓ -hop friends of u .

As shown in Fig. 10(c) and (e), the user objects are uniformly distributed. Then the number of user objects in a given range is relatively stable, independent of the value of rd . So the query time of D-RSNCF is relatively stable with the increase of rd . In Fig. 10(d) and (f), the user objects follow the Zipf distribution. The number of edges needs to be traversed by D-RSNCF is uncertain, so the query time of D-RSNCF is uncertain with rd increasing.

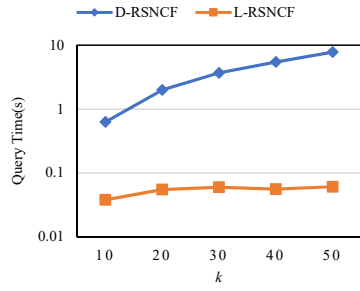
For L-RSNCF, by using R-Tree it traverses only the edges containing the ℓ -hop friends of the query user u , so the running time is uncertain, as shown in Fig. 10. With the increase of rd , the running time of L-RSNCF does not change significantly.

6.3. Discussion

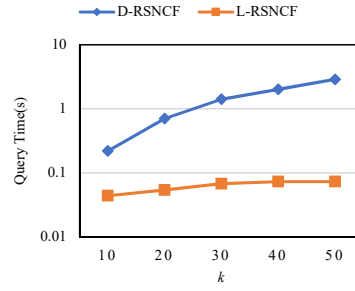
In this section, 6 datasets are used to test the two algorithms. D-RSNCF is not as efficient as L-RSNCF in most cases, but D-RSNCF is more efficient than L-RSNCF on those four synthetic datasets in most cases. The reason is that these four synthetic datasets have much higher density than the two real datasets and D-RSNCF is based on Dijkstra algorithm. For high-density dataset, the search range becomes smaller, so D-RSNCF is more efficient. For dataset with lower density, the search range of D-RSNCF will become larger and the efficiency will become lower. L-RSNCF is based on IS-Label index, so the change of dataset density has no obvious impact on the running time of L-RSNCF.

7. Conclusion

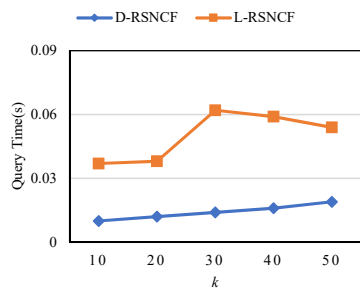
This paper makes an in-depth exploration of the k -nearest ℓ -close friends ($k\ell$ -NCF) query in road-social networks. The $RSk\ell$ -NCF query algorithm based on Dijkstra (D-RSNCF)



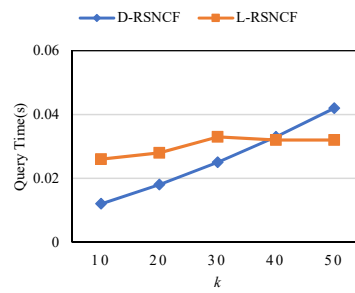
(a) BR_rangeBAY + BAY



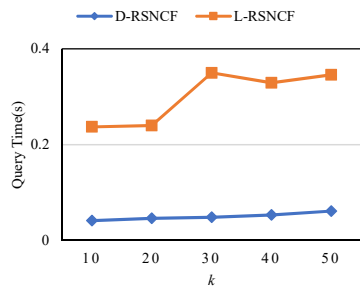
(b) GA_rangeBAY + BAY



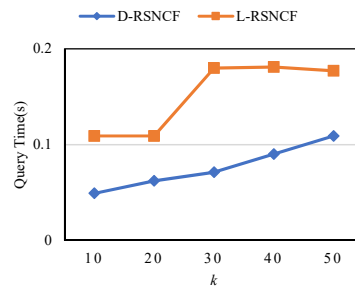
(c) UBR + TG



(d) ZBR + TG

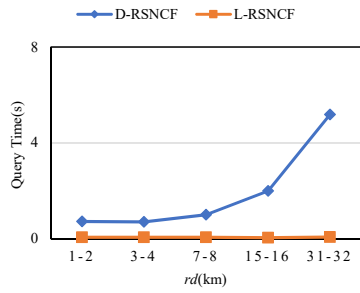


(e) UGA + SF

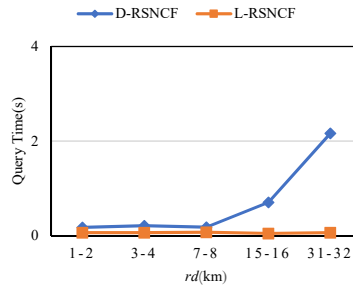


(f) ZGA + SF

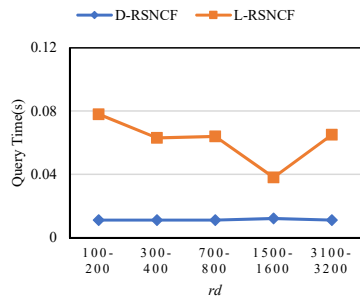
Fig. 9. Effect of k on query time



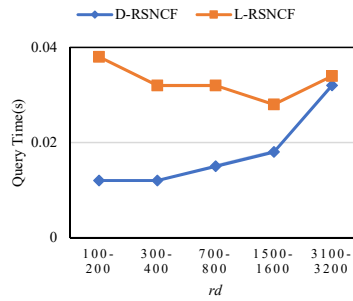
(a) BR_rangeBAY + BAY



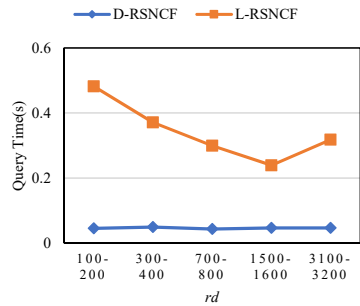
(b) GA_rangeBAY + BAY



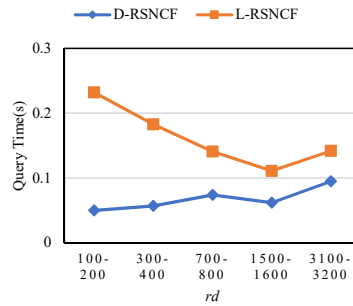
(c) UBR + TG



(d) ZBR + TG



(e) UGA + SF



(f) ZGA + SF

Fig. 10. Effect of *rd* on query time

and the $RSk\ell$ -NCF query algorithm based on IS-Label index (L-RSNCF) are proposed. For both methods, several hash tables are used to speed the query. D-RSNCF is based on Dijkstra algorithm to traverse the user objects needed. L-RSNCF is based on IS-Label and R-Tree to traverse the user objects needed. Real datasets and synthetic datasets are used to test the two algorithms. Through experiments, we find that D-RSNCF is more suitable for dataset with high user object density, while L-RSNCF is just the opposite.

References

1. Ahuja, R., Armenatzoglou, N., Papadias, D., Fakas, G.J.: Geo-social keyword search. In: Proceedings of the 14th International Symposium on Advances in Spatial and Temporal Databases, SSTD 2015, Hong Kong, China. pp. 431–450 (2015)
2. Attique, M., Afzal, M., Ali, F., Mehmood, I., Ijaz, M.F., Cho, H.: Geo-social top-k and skyline keyword queries on road networks. *Sensors* 20(3), 798 (2020)
3. Cho, E., Myers, S.A., Leskovec, J.: Friendship and mobility: user movement in location-based social networks. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA. pp. 1082–1090 (2011)
4. Dijkstra, E.W.: A note on two problems in connexion with graphs. *Numerische Mathematik* 1, 269–271 (1959)
5. Dong, T., Lulu, Y., Cheng, Q., Cao, B., Fan, J.: Direction-aware KNN queries for moving objects in a road network. *World Wide Web* 22(4), 1765–1797 (2019)
6. Emrich, T., Franzke, M., Mamoulis, N., Renz, M., Züfle, A.: Geo-social skyline queries. In: Proceedings of the 19th International Conference on Database Systems for Advanced Applications, DASFAA 2014, Part II, Bali, Indonesia. vol. 8422, pp. 77–91 (2014)
7. Fu, A.W., Wu, H., Cheng, J., Wong, R.C.: IS-LABEL: an independent-set based labeling scheme for point-to-point distance querying. *Proceedings of the VLDB Endowment* 6(6), 457–468 (2013)
8. Ghosh, B., Ali, M.E., Choudhury, F.M., Apon, S.H., Sellis, T., Li, J.: The flexible socio spatial group queries. *Proceedings of the VLDB Endowment* 12(2), 99–111 (2018)
9. Guttman, A.: R-trees: A dynamic index structure for spatial searching. In: Proceedings of the ACM SIGMOD Annual Meeting on Management of Data, SIGMOD 1984, Boston, Massachusetts, USA. pp. 47–57 (1984)
10. He, D., Wang, S., Zhou, X., Cheng, R.: An efficient framework for correctness-aware kNN queries on road networks. In: Proceedings of the 35th IEEE International Conference on Data Engineering, ICDE 2019, Macao, China. pp. 1298–1309 (2019)
11. Hjalton, G.R., Samet, H.: Distance browsing in spatial databases. *ACM Transactions on Database Systems* 24(2), 265–318 (1999)
12. Hu, H., Lee, D.L., Xu, J.: Fast nearest neighbor search on road networks. In: Proceedings of the 10th International Conference on Extending Database Technology, EDBT 2006, Munich, Germany. pp. 186–203 (2006)
13. Jiang, J., Lu, H., Yang, B., Cui, B.: Finding top-k local users in geo-tagged social media data. In: Proceedings of the 31st IEEE International Conference on Data Engineering, ICDE 2015, Seoul, South Korea. pp. 267–278 (2015)
14. Jiang, M., Fu, A.W., Wong, R.C.: Exact top-k nearest keyword search in large networks. In: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Melbourne, Victoria, Australia. pp. 393–404 (2015)
15. Kim, W., Shim, C., Heo, W., Yi, S., Chung, Y.D.: Moving view field nearest neighbor queries. *Data & Knowledge Engineering* 119, 58–70 (2019)
16. Lee, K.C.K., Lee, W., Zheng, B., Tian, Y.: ROAD: A new spatial object search framework for road networks. *IEEE Transactions on Knowledge and Data Engineering* 24(3), 547–560 (2012)

17. Li, Q., Zhu, Y., Yu, J.X.: Skyline cohesive group queries in large road-social networks. In: Proceedings of the 36th IEEE International Conference on Data Engineering, ICDE 2020, Dallas, TX, USA. pp. 397–408 (2020)
18. Liu, W., Sun, W., Chen, C., Huang, Y., Jing, Y., Chen, K.: Circle of friend query in geo-social networks. In: Proceedings of the 17th International Conference on Database Systems for Advanced Applications, DASFAA 2012, Part II, Busan, South Korea. pp. 126–137 (2012)
19. Ma, Y., Yuan, Y., Wang, G., Bi, X., Wang, Y.: Personalized geo-social group queries in location-based social networks. In: Proceedings of the 23rd International Conference on Database Systems for Advanced Applications, DASFAA 2018, Part I, Gold Coast, QLD, Australia. pp. 388–405 (2018)
20. Ouyang, D., Wen, D., Qin, L., Chang, L., Zhang, Y., Lin, X.: Progressive top-k nearest neighbors search in large road networks. In: Proceedings of the 2020 International Conference on Management of Data, SIGMOD Conference 2020, online conference [Portland, OR, USA]. pp. 1781–1795 (2020)
21. Roussopoulos, N., Kelley, S., Vincent, F.: Nearest neighbor queries. In: Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data, San Jose, California, USA. pp. 71–79 (1995)
22. Shim, C., Kim, W., Heo, W., Yi, S., Chung, Y.D.: Nearest close friend search in geo-social networks. *Information Sciences* 423, 235–256 (2018)
23. Shim, C., Sim, G., Chung, Y.D.: Cohesive ridesharing group queries in geo-social networks. *IEEE Access* 8, 97418–97436 (2020)
24. Sohail, A., Cheema, M.A., Taniar, D.: Social-aware spatial top-k and skyline queries. *The Computer Journal* 61(11), 1620–1638 (2018)
25. Sohail, A., Hidayat, A., Cheema, M.A., Taniar, D.: Location-aware group preference queries in social-networks. In: Proceedings of the 29th Australasian Database Conference on Databases Theory and Applications, ADC 2018, Gold Coast, QLD, Australia. pp. 53–67 (2018)
26. Zhao, J., Gao, Y., Chen, G., Jensen, C.S., Chen, R., Cai, D.: Reverse top-k geo-social keyword queries in road networks. In: Proceedings of the 33rd IEEE International Conference on Data Engineering, ICDE 2017, San Diego, CA, USA. pp. 387–398 (2017)
27. Zhao, J., Gao, Y., Ma, C., Jin, P., Wen, S.: On efficiently diversified top-k geo-social keyword query processing in road networks. *Information Sciences* 512, 813–829 (2020)
28. Zhao, S., Xiong, L.: Group nearest compact POI set queries in road networks. In: Proceedings of the 20th IEEE International Conference on Mobile Data Management, MDM 2019, Hong Kong, SAR, China. pp. 106–111 (2019)
29. Zhong, R., Li, G., Tan, K., Zhou, L., Gong, Z.: G-tree: An efficient and scalable index for spatial search on road networks. *IEEE Transactions on Knowledge and Data Engineering* 27(8), 2175–2189 (2015)
30. Zhu, Q., Hu, H., Xu, C., Xu, J., Lee, W.: Geo-social group queries with minimum acquaintance constraints. *The VLDB Journal* 26(5), 709–727 (2017)

Zijun Chen received the bachelor’s degree from the Northeast Heavy Machinery Institute, China, the master’s degree from Yanshan University, and the PhD degree from Fudan University in 2002, all in computer science. Since 1995, he has been with the School of Information Science and Engineering, Yanshan University, Qinhuangdao, China, where he is currently a professor. His research interests include moving object databases, spatio-temporal databases and graph databases.

Ruoyu Jiang received the bachelor’s degree in software engineering from Hebei Normal University, China, in 2018. She received the master’s degree in computer technology from Yanshan University, China, in 2021. Her research interest includes geo-social networks.

Wenyuan Liu received the bachelor's and master's degrees from the Northeast Heavy Machinery Institute, China, and the PhD degree from the Harbin Institute of Technology in 2000, all in computer science. Since 1996, he has been with the School of Information Science and Engineering, Yanshan University, Qinhuangdao, China, where he is currently a professor. His research interests include wireless sensor networks and mobile networks.

Received: August 30, 2021; Accepted: July 30, 2022.

Crowdsourcing Platform for QoE Evaluation for Cloud Multimedia Services

Asif Ali Laghari¹, Hui He², Asiya Khan³, Rashid Ali Laghari⁴, Shoulin Yin^{5*}, and Jiachi Wang⁶

¹ Department of Computer Science, Sindh Madressatul Islam University
Karachi, Pakistan
asif.laghari@smiu.edu.pk

² School of Computer Science & Technology, Harbin Institute of Technology
Harbin, China
hehui@hit.edu.cn

³ School of Engineering, Computing and Mathematics, University of Plymouth
United Kingdom
asiya.khan@plymouth.ac.uk

⁴ Interdisciplinary Research Center for Intelligent Manufacturing and Robotics (IRC-IMR), King Fahd University of Petroleum and Minerals, Dhahran, 31261, Saudi Arabia
rashidalilaghari@gmail.com

⁵ School of Information and Communication Engineering, Harbin Engineering University
Harbin China
352720214@qq.com

⁶ Software College, Shenyang Normal University
Shenyang 110034, China
853757309@qq.com

Abstract. This paper presents a novel web-based crowdsourcing platform for the assessment of the subjective and objective quality of experience (QoE) of the video service in the cloud-server environment. The user has the option to enter subjective QoE data for video service by filling out a web questionnaire. The objective QoE data of the cloud-server, network condition, and the user device is automatically captured by the crowdsourcing platform. Our proposed system collects both objective and subjective QoE simultaneously in real-time. The paper presents the key technologies used in the development of the platform and describes the functional requirements and design ideas of the system in detail. The system collects real-time comprehensive data to enhance the quality of the user experience to provide a valuable reference. The system is tested in a real-time environment and the test results are given in terms of the system performance. The crowdsourcing platform has new features of real-time network monitoring, the client device, and cloud monitoring, which currently has not been provided by existing web platforms and crowdsourcing frameworks. The results show that 1MB buffer is filled 100% very soon after starting watching videos from the crowdsourcing platform.

Keywords: Crowdsourcing platform, Video service, Quality of Experience (QoE), Cloud computing.

* Corresponding author

1. Introduction

Today with the rapid development of the Internet and mobile devices, organizations have been able to provide a variety of services for users. One of them is multimedia cloud computing, which has been mainly offered on the Internet [1]. Cloud computing organizations provide multimedia services to end-users on pay per use. Multimedia-streaming technology has been widely used in Internet TV, online cinema, live events, video sessions, short video sharing, and so on [2]. Besides that, more and more multimedia streaming services are being created and developed. Meanwhile, the competition among the service providers becomes fiercer, if the organization wants to win in such a fierce competition, they will need to be recognized by the users and satisfy them. At the same time they provide services, cloud organizations are paying attention to the QoE and hoping to improve their service level by users satisfaction with their products [3]. Previous research has provided QoE based solutions for multimedia services based on the QoE to provide quality of service (QoS) for video streaming to end-user [4,5], but did not provide a satisfactory solution that differentiates the positive and negative feedback of users for service management [6]. Users' high-quality experience will improve their awareness of business and application which will enhance the organization's brand value [7]. Therefore, it is necessary to collect information about QoE for users who use streaming media services. Through the analysis of collected QoE data, they can develop systems to provide better services for users [8].

One of the biggest challenges that the service providers face is how to observe and improve the business quality in real-time [9], and hence, gain more user groups. At present, most organizations take the form of questionnaires or user complaints to obtain the user's subjective QoE data and then go to optimize and adjust the network [10,11]. While this is the way to get the user's QoE data, it is time-consuming and less comprehensive [12,13]. So, organizations need a way to get the objective data of QoE quickly and in real-time. Still, there is no crowdsensing framework is proposed by any researcher, which collects subjective and objective QoE and differentiates between negative and positive QoE [41,42]. For multimedia streaming services, while the user is watching a crowdsourcing platform can capture the objective metrics of the service performance data (such as network bandwidth, network latency) and send them to the server to record [14,15].

QoE is defined as "a blueprint of all human subjective and objective quality needs and experiences arising from the interaction of a person with technology and with business entities in a particular context" [16]. The QoE is all about the human perception by using any product or service [43]. The QoE methods are used for capturing user opinions about video quality, video streaming and network services and products [17,18]. The QoE can be captured by using subjective and objective approaches, one is a subjective way, which is conducted by using interviews, questionnaires, and web survey methods. The subjective QoE is captured by the mean opinion score (MOS). The second method is objective QoE, which is based on QoS and human physiological data [17].

The combination of subjective and objective QoE data has a greater reference value to the service providers to provide services according to user needs with QoS for those users who use mobile devices [19]. As the final link in the whole mobile communication industry chain, the mobile device directly affects the user's perception and QoE of the service and application provided by the organization [20], so the QoE data of the mobile device is very important for organizations [21]. To solve these problems, we developed

mobile-oriented multimedia streaming services and a QoE evaluation platform, which can provide video services to users and statistically analyze the user's QoE subjective and objective data simultaneously in real-time. Previous platforms were limited in capturing only one type of QoE, subjective or objective [22-25] but did not have the functionality to capture objective QoE in real-time and made comparative analysis for accurate QoE and service level agreement (SLA) [26].

The main contribution of this paper is to present the crowdsourcing platform for the assessment of subjective and objective QoE video services in the cloud-server environment. The proposed platform is made up of two subsystems which include a video service website subsystem and a QoE data statistics management subsystem. Video website provides relatively complete video services; users can register and log onto the site, also can upload and manage video on the site. The crowdsourcing platform contains a questionnaire and this questionnaire is based on a QoE-related research foundation design. Users can submit the questionnaire for feedback on their subjective experience data. On the back-end, we collect objective QoE/QoS data of the user, which does not affect the user's normal use. The administrator can view the user's subjective and objective QoE data by logging into the QoE Statistics Management System. Through the crowdsourcing platform's QoE data the video service system will optimize to provide users with better service. To our knowledge, the crowdsourcing platform has new features of real-time network monitoring of the client device and cloud monitoring, which currently has not been provided by existing web platforms and crowdsourcing frameworks.

The rest of the paper is organized into 6 sections. In section 2, we provide related work based on the overview of the existing QoE/QoS based platforms for cloud multimedia services. Section 3 provides the crowdsourcing platform requirement, analysis, and design. Section 4 presents the video subsystem website and implementation. Section 5 presents testing of the system and results of the QoE crowdsourcing platform for cloud multimedia services and finally, in section 6, we conclude the work and provide future directions.

2. Related Work

This section is divided in two sub-sections giving an overview of the existing QoE/QoS based platforms for cloud multimedia services and a comparison of existing crowdsourcing platforms.

2.1. Overview of Existing QoS/QoE Based Platforms for Cloud Multimedia Services

Web platforms and frameworks for QoE capture and assessment were introduced by several researchers [4, 27, 28, 29]. A survey of web-based crowdsourcing frameworks based on subjective QoE is given by Hobfeld et al. [30], which monitors the objective QoE/QoS data from the cloud to the user's device. Web frameworks are limited for the evaluation of multimedia services to get users' perceptions about the services and multimedia contents, but they did not compare service delivery status to SLA. Also, they did not focus on the negative responses of the end-users during the submission of QoE when proposed and developed web-based platforms. Wu-Hsiao Hsu and Chi-Hsiang Lo [4] proposed QoS/QoE mapping and adjust their model for cloud service providers to monitor and adjust the

user's QoE. The proposed model translates QoS parameters into QoE in a cloud computing environment. The model is tested by capturing QoE of users by setting a simulated platform of video streaming, which consists of three parts, the GA, NS-2, and monitoring process and compares it to the monitored QoS parameters. During the experiment, two parameters were used as buffering time (BT) and streaming video discontinuity (SVD) to measure user perception about video streaming. Forty-eight videos were used with different lengths and forty users were invited to view the videos and assign mean opinion score (MOS) for each video. The result of the experiment shows that network QoS and the user's QoE are consistent with each other.

Jordi et al. developed the Wersync web platform [25], which enables distributed media synchronization and social interaction across remote users. The development of the Wersync web platform is based on four key technologies, The first one Node.js is a cross-platform runtime environment and open source, which uses a networking based application and is written in JavaScript for the server-side. Second is the HTML 5 video component, which supports embedding full-fledged media players into web pages. The third element is the clock synchronization between all the involved entities to ensure a coherent notion of time in the shared session. Fourth is Socket.IO, which is a lightweight JavaScript library that enables real-time bidirectional communication between web clients and a (Node.js) webserver. This platform provides a facility for users to create and join sessions, which are ongoing and use a the same media contents with remote users in a synchronized manner. This platform also supports cross-platform, cross-device support and cross-network, which is a key point in the current heterogeneous media delivery ecosystem.

Ahammad et al. present a flexible web platform for QoE-driven delivery of image-rich web applications [22]. The platform is based on content-aware optimization and instead of delivering the whole image once to the client, it delivers in partitions. The partitions image approach is based on creating content-aware parts of the image and reorganizing the bytes on their significance and using timestamp in the application. At the client-side delivered image parts were recombined in the full image and did not affect the visual perception of the end-user. This approach is different from the content delivery network (CDN) architecture because cloud behavior launches optimization process offline, and when it is ready then starts serving to deliver contents C thus delivering an improved QoE overall. The other major difference of this architecture for improving the performance of the application is that it avoids passive response to browser requests, it breaks requests in pieces or delays in a request to avoid negative impact on the user's other resource download.

A Data-Driven Platform for QoE Visualization and System Performance Monitoring (QOEYE) has been proposed by Chao Zhou [23]. The proposed platform is based on the QoE metrics of rule-based QoS monitoring. During the experiments, a real trace of the BesTV content provider is used as the data source, and video logs were collected from the BesTV server to monitor server and log data which were statistically analyzed to find fault location. QoE monitoring metrics were set by dividing the playtime of the video into 5 percentage points, such as 0-20% and 80- 100%. If the load on the server is at the peak, the user will not watch the complete video and leave within 0 C 20% resulting in a decrease in QoE. Organizations earn from long video watching of the user by adding advertisements in the video. If users watch the complete video then the log will store

on the server from 80 C 100% and the QoE of the user is excellent. QOEYE platform provides the facility to use content without limitation of geographical boundaries and the user can easily view usage statistics via the service provider. This platform also enables us to monitor large scale servers and find the fault location.

Cross-Layer Multi-Cloud Application Monitoring as a Service Framework (CLAMS) is a proposed framework for QoS monitoring based on agent technology, which monitors applications and big data analytics in a multi-cloud environment and addresses the issue of cross-layer monitoring of applications [33]. CLAMS framework has limited functionality which only supports QoS monitoring and does not support subjective QoE, user device monitoring and external network monitoring outside the cloud organization. Our proposed crowdsourcing platform captures both network QoS (NQoS) and application QoS (AQoS) parameters. It is based on subjective and objective QoE/QoS assessment and data analysis.

2.2. Comparison of Existing Crowdsourcing Frameworks with Crowdsourcing Platform

Ribeiro et al. [34] proposed an open-source project crowdMOS framework, which only focused on the audio quality assessment of users and this framework can easily be modified or installed on any suitable web server. Initially, crowdMOS was developed for audio quality assessment but later this was extended to support image quality assessment methodology. For assessing the accuracy of MOS of users it uses a simple correlation coefficient between the MOS. crowdMOS rejected users if the correlation is less than a definite threshold such as 0.25 suggested in [6]. crowdMOS accepted a large number of fake user's MOS, to avoid this threshold can be increased to eliminate a large number of users. This framework did not support QoS monitoring to analyze the user's given MOS and system set preferences.

The QualityCrowd [35] framework is an open-source project for a crowdsourcing framework designed for QoE evaluation, which can be easily modified and installed with minimal effort on any web server. This framework supports multiple questions to set and design tests for the image, video, and audio with multiple combinations and methodologies. Crowdsourcing framework is based on two parts: A back-end can handle test results and provide features for new test design and the front-end provide a user interface where the actual subjective test takes place. This framework also has limitations such as it did not provide automatic monitoring of contents and objective QoE assessment features, therefore, it is unable to distinguish between the positive and negative feedback of end-users.

A web-based subjective QoE open-source evaluation platform has been proposed by Benjamin Rainer [24] and is available at [32]. The proposed platform was developed for the assessment of subjective QoE for both lab experiments as well as for crowdsourcing. The proposed crowdsourcing platform is based on the HTTP server with the support of PHP and a MySQL database. The platform follows the Model View Controller (MVC) pattern, the description of test and questions can be configured separately as the requirement of the field for subjective QoE assessment will be conducted. This platform can easily be extended and used for the QoE assessment of multimedia applications and by adding adobe flash player and HTML5 as it supports a wide range of codecs and browsers.

Kraft and Zolzer designed the BeagleJS crowdsourcing framework for subjective QoE assessment of audio files [36]. This framework is built on PHP, JavaScript, and HTML5,

which supports numerous audio file formats for subjective QoE assessment. This framework can easily be extended to add more evaluation methodology by extending a simple code. This framework does not support database to store user data, however, results are emailed in the format of text files to the organization.

In-momento crowdsourcing framework for the assessment of QoE was introduced by Gardlo et al. [37], which is best known for QoE conducting a test with a user interface facility to end-users. This framework provides reliable ratings because the user has an interface facility to view and understand scales and discard unreliable ratings due to strict a-posteriori filtering which reduces the amount of large work from the administration. The framework also provides a rapid feedback section for direct communication with test participants if any suspicious behavior is found. This also provides a facility for users to submit their ratings if the previous rating is not satisfied or due to low performance and can be continued or stopped the testing process.

The proposed crowdsourcing platform is intended to perform the functions of monitoring the user system and middle network traffic and cloud infrastructure from the client device to the cloud. The user's submitted subjective QoE and automatically collected objective QoS/QoE data will be examined for service delivery according to SLA. The crowdsourcing platform also has the functionality of analyzing the difference between the positive and negative QoE by comparing it with the current service delivery report and QoE submitted by the user, which is to the best of our knowledge, these functions have not been provided by previous crowdsourcing platforms. A comparison of features is given in Table 1.

3. Web Platform Requirements Analysis and Design

This section gives an overview of proposed system design and QoE data collection.

3.1. System Design

The structure of the platform is shown in Figure 1, which shows the core functions and the system partitioning of the project macroscopically. The following will be the three subsystems of the project's functional aspects of the introduction.

The system consists of three subsystems, including a video service site, an Android video application, and QoE data statistics management system. Through this system, the administrator can access the subjective and objective QoE data of the streaming media from Android end, to summarize and analyze these data. Through this platform, cloud service providers can collect and evaluate the QoS/E in real-time and provide more data support for user QoE evaluation. This platform can more fully reflect the true QoE, thus ensuring the user experience. The structure of the crowdsourcing platform is shown in Figure 1.

3.2. Demand Analysis of Website

This part is a relatively complete video service system, which provides users with video service, subjective QoE survey and interface function for data submission. The use case diagram is shown in Figure 2.

Table 1. Comparison of exiting crowdsourcing Framework with New Crowdsourcing Platform

Framework Feature	CrowdMOS [34]	QualityCrowd2 [35]	WESP [24]
Media type	Image, audio	Image, video and audio	Image, video, audio, sensory effects
Methodology	ACR, DCR, Mushra	ACR, flexible: single, double stimulus; discrete, continuous scales	All (flexible), e.g.,ACR, ACR-HR, DSCQE, Double stimulus for sensory effects
Questionnaire	Embedded in evaluation	Separated tasks	Embedded in evaluation
Task Design	Custom template all tasks have the same template	Custom template Tasks configured in script file	All tasks have the same template
Task order	Random Full set or subset of all stimuli	Fixed	Flexible
Open source	Yes	Yes	Yes
Data storage	Text files	Text files CSV format	Database
Programming Language	Ruby	PHP+own script language	JavaScript+PHP
Monitoring	No	No	No
Remarks	Subjective Evaluation	Subjective Evaluation	Subjective Evaluation
Framework Feature	BeagleJS [36]	in-momento [37]	Proposed Crowdsourcing Platform
Media type	Audio	Image, Video	Image, Video and Audio
Methodology	ACR, DCR, Mushra	ACR, flexible: single, double stimulus; discrete, continuous scales	All (flexible), e.g.,ACR, ACR-HR, DSCQE, Double stimulus for sensory effects
Questionnaire	Embedded in evaluation	Separated tasks	Embedded in evaluation
Task Design	Custom template all tasks have the same template	Custom template Tasks configured in script file	All tasks have the same template
Task order	Fixed	Random based on actual number of ratings	Flexible
Open source	Yes	Yes	Yes
Data storage	Text files	Database	MySQL Database
Programming Language	JavaScript+PHP	PHP	JavaScript, HTML5, CSS+PHP, MySQL
Monitoring	No	Limited	Overall
Remarks	Subjective Evaluation	Subjective Evaluation	Subjective, Objective (QoS)

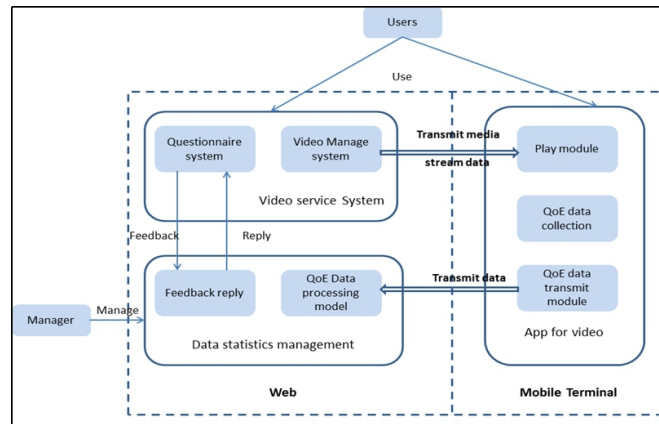


Fig. 1. Crowdsourcing Platform structure diagram

Users need to register and login before they can use the video service site. The video site defaults to provide users with 2GB of cloud storage space. Users can manage their own cloud storage space; they can do management like video upload, playback and delete operation. The cloud storage space also has a video folder management system and users can create or delete folders for users to organize their uploaded videos. Users can modify the basic attribute information (name, description, label, whether public) of the uploaded videos and modify the folder to which the video belongs by the management operation. Open-book management provides an option for the user to manage their videos public to other users of the system.

This site provides user feedback capabilities, and the user in the use of the video service can fill out the questionnaire on the site to provide feedback for the web services. This section also lists the administrator's response to the user's questionnaire, and the user can view the administrator's response to solving their problems, to better make use of this site.

3.3. Demand Analysis of QoE System

QoE data statistics management platform for the administrator is presented here. Administrators log on to the platform to view and manage the system's QoE data. The system consists of QoE capture, assessment, and QoE data display. The use-case diagram for the management platform subsystem is shown in Figure 3.

When the administrator logs in to the system, they can view the subjective and objective QoE data feedback from the user. Subjective QoE data is submitted via the questionnaire and objective QoE is captured automatically by various video playbacks, objective QoE data contains information of the user's device, including the video buffer status delay time and some basic parameters of the device. Both subjective and objective QoE data will be stored on the server in the database for future processing. The administrator can view each questionnaire and can respond to the user's questionnaire feedback.

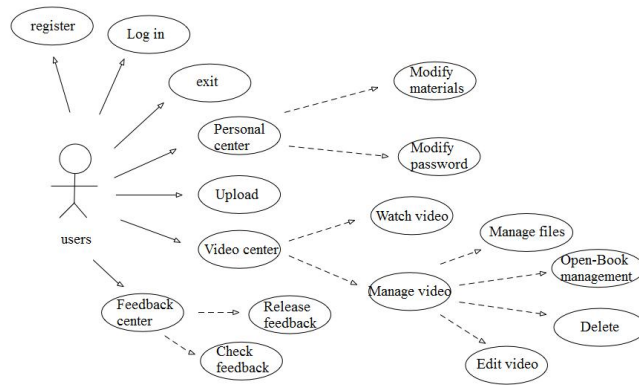


Fig. 2. Video Website Subsystem Use Case Diagram

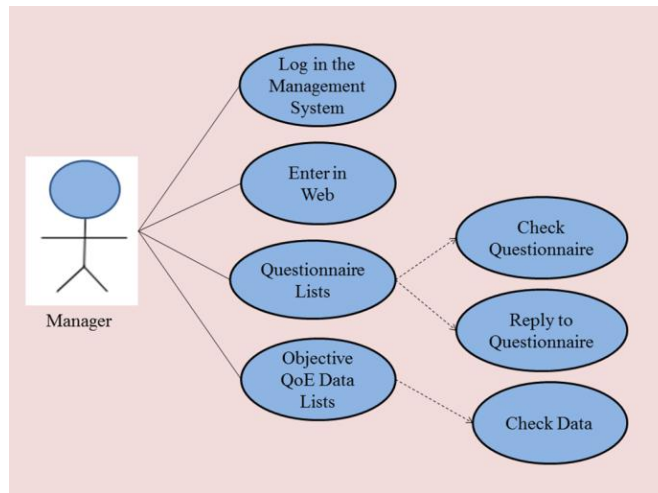


Fig. 3. QoE data management platform use case diagram

4. Video subsystem Website Design and Implementation

This section gives an overview of the video subsystem and QoE statistics management.

4.1. Video Subsystem

The video subsystem includes a user site service module, video module, questionnaire modules and app interface module. The cloud side view of the subsystem is shown in Figure 4 and is described in detail below.

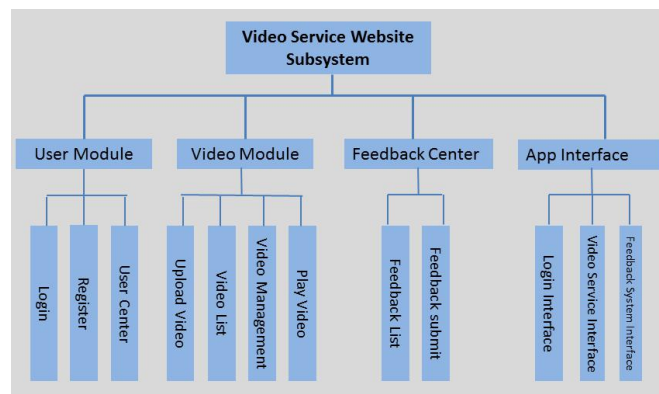


Fig. 4. Video Service Website Subsystem Cloud server

Video module (1) Video upload

The user has the option to browse the local system to upload the video file to his account, but file selection has been limited because the crowdsourcing platform accepts only .mp4, .mkv, .flv and .mov video formats. Taking into account the larger video files PHP server-side configuration has restrictions on the upload file size, slow upload, and other factors so there is a need for video files on the client-side for the concurrent upload. The project uses Web up-loader components, integration fragment and concurrency to upload large files split into multiple pieces (each 5MB) as shown in Figure 5.

When the user clicks upload platform uses Ajax technology to send the size of the video information to the back-end to verify the user's default storage space for 2GB if the user's available space allows then a video is uploaded, if space is not sufficient the user is prompted: not enough storage!.

PHP background on the block to upload the video processing has been uploaded to merge the video clips when the merger is successful then set the tag parameter \$ done is true and a unique file name is assigned and stored in the upload directory. The platform that contains the FFmpeg component has been successfully uploaded to the video clip from the beginning of the video cover (preview image) PHP background command is as follows:

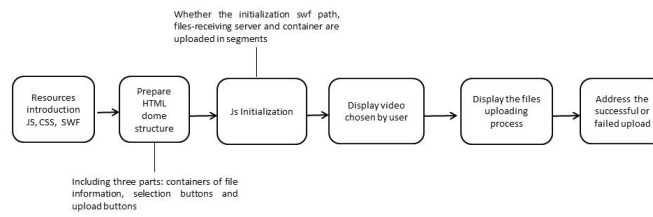


Fig. 5. Web Up-loader component integration

\$ File. "-y -f mjpeg -ss 3 -t". \$ Time. "-s 320x180". \$ Imgpath); // Interception (\$ str = "/usr/local/bin/ffmpeg -i" The cover image of the video; We also need to obtain the total length of the video, php background execution command is as follows: \$ Vtime = exec ("/usr/local/bin/ffmpeg -i". \$ File. "2i & 1 — grep 'Duration' — cut -d " -f 4 — sed s // Get the total length of the video; We use the file size (\$ DIR) function to get the size of the video and store the video's data in the database and the user's free space is updated.

(2) Video list

The user's media library is divided into the main list and folder system; the user can view the uploaded video and then manage them. On the main list page uploaded videos of the user will show displaying 5 videos per page and counting the total number of videos. The user can select one or more videos to delete, move to a folder, share or cancel the videos for the public. The operation of the flowchart is shown in Figure 6.

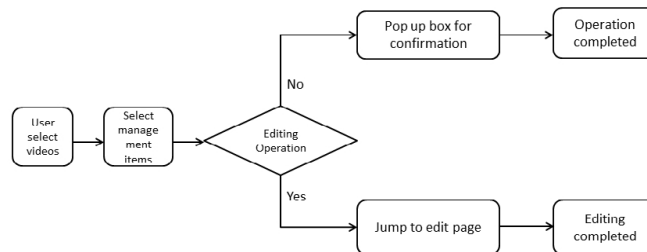


Fig. 6. Schematic diagram of the operation flow

When the user acts to delete the video, open or cancel then a pop-up prompt box appears and the user can choose to confirm or cancel the operation. The user has the option to move the video folder by using the drop-down bar. If the user chooses to confirm the number of id arrays submitted to the background of the site, it verifies video id and is first deleted from the disk storage and then deleted from the database record and the delete field is set to 1. Other operations can directly modify the corresponding field in the database.

When the user selects a video for editing the video editing page will display, the user can rename the video and also the label the description of the corresponding editing and

modification work. When the user clicks the Save button the form will be submitted to the background post, because the video id is passed as a parameter, the background of the effectiveness of the id is verified and if the video information is correct the data of the video is updated as shown in Figure 7.



Fig. 7. JWplayer components integration

Feedback Center Module (1) Feedback list

Users click to enter the feedback list and enter their subjective QoE. The administrator will give feedback to the list of all user questions and the user can click on one of the feedbacks to view the administrator's response to the problem. Through this list, users can better understand how to use the site and can easily solve the problems encountered when using the system.

(2) The user questionnaire

The crowdsourcing platform contains a questionnaire (form) for the user to enter their feedback (subjective QoE) about the services, which they receive. International Telecommunication Union (ITU) provides a table for subjective experience indicators to collect the user experience, which is shown in Table 2 to allow the user to rate the assessment [40].

Table 2. Selection of Experience Indicators [40]

MOS	Quality	Perception
5	Excellent	Imperceptible
4	Good	Perceptible
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

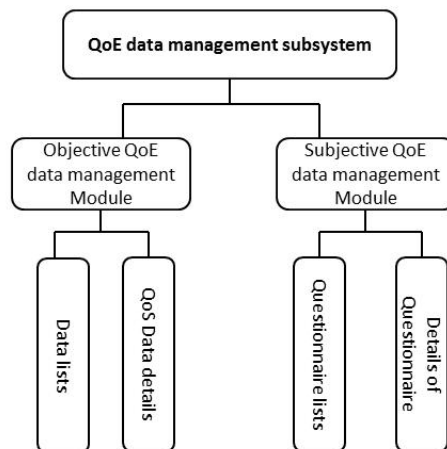
In combination with the above scoring method, we designed a user feedback questionnaire as shown in Table 3. The questionnaire can be completed and submitted by the user. The user needs to fill in the required fields with * and other feedback information. During the experiment it was verified the validity of the feedback data submitted by the user and marked the source for the PC will be deposited in the database. Through the questionnaire, we can get the user's subjective experience of the platform.

Table 3. Feedback Questionnaire Design

Item	Type	Required fields or not	Description
User name	text	yes	User name
Email	text	yes	User Email
Phone	text	no	User phone
Priority	option	no	Urgency degree of feedback 1.low 2.normal 3.high
Issue	text	no	Problems encountered by users
Network satisfy	option	no	Satisfaction of users
Network type	option	no	Network type 1.3G/4G 2.Gable network 3.unknown
Network speed	Text,option	no	Network speed, User can firstly write the speed or choose unknown
Buffer waiting	option	no	Dose the buffer get stuck
Video quality	comment	no	Overall comments of video which has 5 levels
Comments	text	no	Other users' feedback

4.2. QoE data statistics management subsystem design and implementation

The administrator can enter the QoE data management background through the "domain name/admin" connection and the non-administrator will be prompted with "access forbidden". This part includes the user's subjective QoE data and objective QoE data in two parts. The subsystem cloud server diagram is shown in Figure 8.

**Fig. 8.** QoE data statistics management subsystem Cloud server diagram

Subjective QoE data management module The subjective QoE data is the feedback questionnaire of data that the user fills. This section will display each page in the form of 10 pages of all user feedback questionnaires, the list in reverse chronological order and the administrator does not consult the questionnaire as it will be hiding. We can get the basic information from the list of the questionnaire the platform source (PC) and the administrator's response. When the administrator clicks on one of the data view buttons can be viewed. Administrators can see the details of the questionnaire and can reply to it.

Objective QoE data management module Objective QoE data is collected from the user's device, network, and cloud server. This section is similar to the feedback list, the same to the page in the form of 10-page display, according to the time in reverse order. When the administrator clicks the view button to jump to the details of the data page and this page consists of four parts: User info, Video info, Device info and QoE Data. User info shows the user information, including user id, account number, nickname, and phone number. Video info shows the recorded video information, including the video's name, duration, and the cover image of the video displayed. Device info shows the user device information obtained by Android, including mobile phone manufacturer, handset model, CPU information, battery power information, network type, and location information.

QoE data first draws a buffering curve, which is plotted using the High Charts component, which reflects the percentage of video buffered and network conditions. We also give the device memory usage, the current application memory usage, video preparation delay time, and the buffer 100% of the proportion of time, the higher the proportion of the user area the less the situation, the better the playback.

4.3. Database Design

The project database contains five tables, namely the member_user table, the folder table, the video table, the feedback table, and the QoE table. The database E-R diagram is shown in Figure 9.

5. System Testing

5.1. Experimental Environment

The Experimental hardware used is the MacBook pro CPU 2.5 GHz Intel Core i7 16G memory 512G hard drives Red rice note3 CPU Qualcomm Xiaolong 650RAM 2GB, ROM 32GB QVGA 1080x1920, 100M LAN and software packages used during the experiment are Server use Mac OS other software and their versions Mysql5.5, PHP5.6.2, Nginx-1.4.2; Android use MIUI 7 Android OS.

The topology is based on the cloud application server, which hosted the crowdsourcing platform, SQL database server is integrated with the cloud application server for performing an operation related to the database. The user terminal is the device where users can access the crowdsourcing platform via the Internet. The crowdsourcing platform was hosted on the architecture in the Computer Network and Security Laboratory at Harbin Institute of Technology and users were invited to access the crowdsourcing platform to create their accounts, upload videos, manage videos and provide feedback by using a

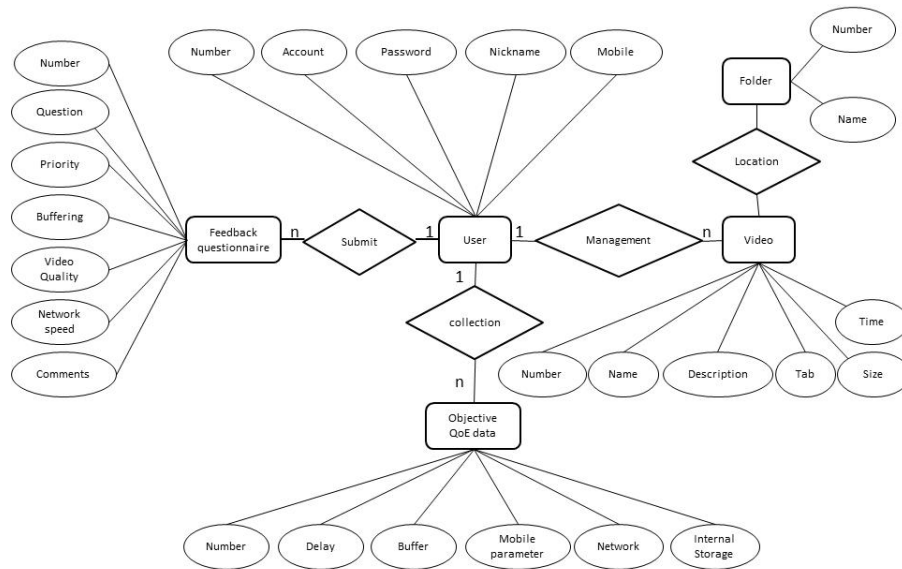


Fig. 9. Database E-R

questionnaire of a crowdsourcing platform. The cloud server architecture and user access diagram are shown in Figure 10.

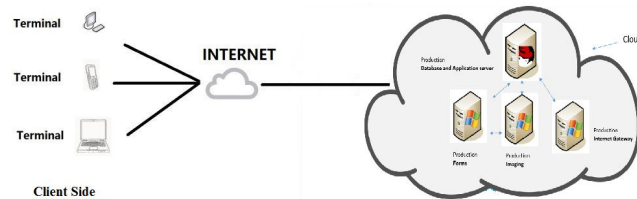


Fig. 10. The Topology of Experiment

5.2. Video Service Website Subsystem Testing

Feedback Questionnaire Module The feedback questionnaire is shown in Figure 11. It comprises the 'feedback' button to enter the feedback list page to expand the view administrator feedback and the 'submit feedback' button to enter into the feedback questionnaire page as shown in Figure 12. This enables the user to submit the feedback questionnaire.

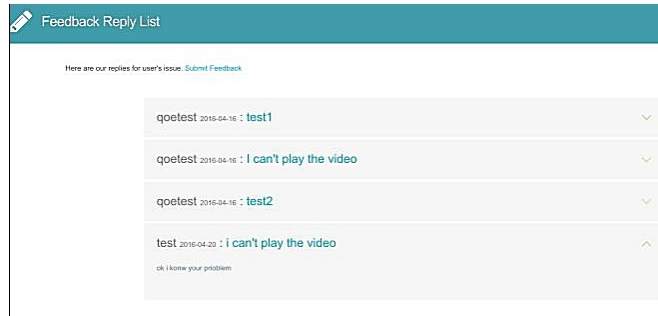


Fig. 11. Feedback list

We're here to help. Please fill out the form and we'll get back to you asap.

User name *: nicehng
Email *: qoetest
Phone *: 1863333333
Priority : Normal Low High

What type issue are you having?

Are you facing Network issues? Then submit below information.
Are you satisfied with quality of Networking services? Yes No
What type of Network you are using? 3G/4G Cable Network Unknown Wifi
Network connection speed? Mb/s Unknown

Video Quality information
Submit information about video quality which you perceive
Are Buffering/Waiting? Yes No
select level of video quality ☆☆☆☆
please score!

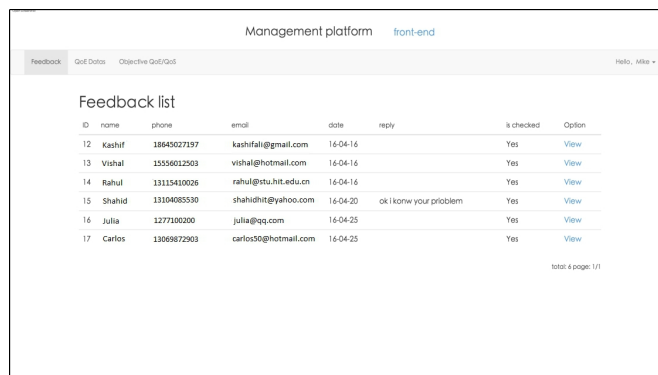
Comments:

[Reset](#)

Fig. 12. Feedback Questionnaire

5.3. QoE data statistics management subsystem testing

The QoE statistics management subsystem is accessible only by the administrator and is accessed by entering the HTTP: // domain name/admin. If the administrator login to the page, the login is successful and the Feedback List page is displayed in Figure 13. Otherwise, the user is prompted to access illegally.



ID	name	phone	email	date	reply	is checked	Option
12	Kashif	18445027197	kashifali@gmail.com	16-04-16		Yes	View
13	Vishal	15556012503	vishal@hotmail.com	16-04-16		Yes	View
14	Rahul	13115410026	rahul@stu.hit.edu.cn	16-04-16		Yes	View
15	Shahid	13104085530	shahidhit@yahoo.com	16-04-20	ok i know your problem	Yes	View
16	Julia	1277100200	julia@qq.com	16-04-25		Yes	View
17	Carlos	13069872903	carlos50@hotmail.com	16-04-25		Yes	View

total 4 page: 1/1

Fig. 13. Feedback Reply list

By clicking on one of the feedback lists, the feedback details page shown in Figure 14, can be accessed, where the administrator can respond to the user.

Agent technology-based function is developed to monitor objective QoE in the crowdsourcing platform and a simple network management protocol (SNMP) is used to QoS data collection from the environment [33]. The SNMP agents are responsible for collecting QoS data such as network type and routing path for data transmission from the cloud to the client. The SIGAR is responsible for system data collection such as RAM and CPU resources consumed by the process, free memory, used and overall memory of the system and overall CPU utilization for all tasks e.g. [38] The platform has the functionality to monitor internal cloud infrastructure for idle resources such as load on the internal network, processing power, and storage. QoS data retrieval of the user such as distance from user to cloud, data routes from several routers between the user and cloud, delay of the router, network throughput, wireless/wired network, user device OS, memory and CPU usage (high impact on the performance), browser and overall capability of user device information for management purpose to the administration for comparison of QoS with SLA.

The monitoring function of objective QoS/QoE is divided into three sections, such as the user device and usage data, middle network environment and internal cloud environment. Additional sections of objective QoS/QoE comprised of the information of the task management (assessed time of accomplishment and remain a time of the task, start and end time, current and previous task).

The user starts using video services, if s/he feels the quality of the video is low or video does not play smoothly and video playback is waiting for a short time then play

Are you facing Network issues? Then submit below information.

Are you satisfied with quality of Networking services?
 Yes

What type of Network you are using?
 Cable Network

Network connection speed?
 1111 Mb/s

Video Quality information
 Submit information about video quality which you perceive

Are Buffering/Waiting?
 Yes

select level of video quality
 5 points

Comments:
 i like this application!

Reply

Thx for your feedback! We will fix this problem soon

Submit reply

Fig. 14. Feedback Details Page

(buffering) then the user can submit his complaint (subjective QoE data/feedback) using feedback questionnaire shown in Figure 12. Users watching videos from the cloud server, same time crowdsourcing platform automatically collect objective QoE/QoS data by using agent technology and store in MySQL database for data analysis. Objective data contains information on the user device's resources such as free RAM and memory for cache files, CPU utilization and network information (such as speed, usage, packet delay, and loss, etc.). This data will be analyzed when a user submits his experience about the services because insufficient resources of the user's device and the network are a problem to receive QoS. Objective QoE data is compared with submitted feedback of user's if the QoS delivery of video services degraded from SLA then crowdsourcing will search for a particular problem. If the problem is found at the user side for insufficient resources to access the cloud multimedia services or external network usage is peak then it will send an alert to the user for the particular problem. If the problem is found at an internal cloud server such as virtual machine (VM) migration, internal cloud network usage is peak due to high traffic then it will send an alert to the administrator for the problem and upgrade the services more than the SLA package as compensation for the time and the problem is solved by the administrator.

QoE is all about subjective user experience. User experience can be both positive negatives. To fully understand the negative feedback, the objective QoE/QoS data collection feature of the crowdsourcing platform will analyze the submitted feedback of the user with objective QoE/QoS to ascertain whether the submitted feedback is true feedback or not. The QoS data collection functionality extended read client device's buffer status and gets information on the current and remaining time of the video, overall information of

video such as the size of the video, total playing time during the playing and after submission of subjective QoE from the user. The buffer checking agent runs across the firewall of the client device in the same way as agent work in the Globus toolkit of grid computing for resource discovery [39]. The result is shown in Figure 15 that the 1Mb buffer is not filled due to the network delay and playing the content of video when the buffer code is tested by using the Wi-Fi network.

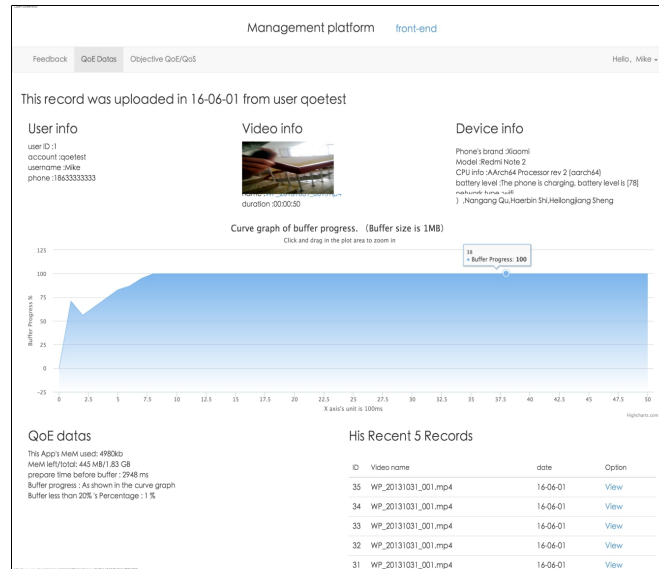


Fig. 15. QoE Data Details page

5.4. User Experience After the Trial Tests

The use case diagram shown in Figure 16 shows the user interaction, accessing services and feedback and on the other hand, cloud management controls the operation of the user. User login in crowdsourcing platform, if the user ID is corrected then s/he can access the services and submit the feedback about the services. If the user ID is wrong then cloud management denies access to the crowdsourcing platform. When the user accesses the services of crowdsourcing such as videos upload, host and share, same time crowdsourcing platform which automatically monitors cloud internal monitor for CPU, utilization, available resources and currently utilization resources, media contents, internal communication network delay, and error rate. Crowdsourcing platform also monitors the middle network between the user and cloud where the crowdsourcing platform is hosted; all objective QoE (QoS data) will be stored in the database of the crowdsourcing platform as well as user-submitted feedback, which are given in the system boundary of the use case diagram. The Cloud management actor handles the overall process of data collection, analysis and sends an alert if a problem is found at the user side. Cloud management

analyzes the client reports and profile and will forward the request to the crowdsourcing platform to produce a report from the QoE database (DB). Cloud administration can select any client from the management section and check his report. The report about the user's problem for getting QoS from the cloud or errors that occurred during the access of service or degrade the performance, which violates the SLA, will be forwarded to the user for information purposes.

The proposed crowdsourcing provides features of monitoring objective QoE/QoS data of client devices, middle network, and cloud environment and compares with subjective QoE submitted data of the user to measure service quality as mentioned in SLA. The result shows that the crowdsourcing framework monitor overall service delivery from client to cloud and Figure 15 illustrates the client device, user's personal, contents it access, and middle network information. Further section 5.4 illustrates system operations from user login to cloud management in the use case diagram.

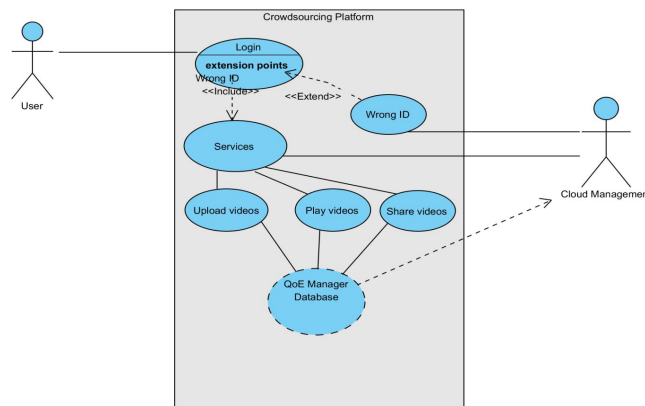


Fig. 16. Use Case Diagram of Crowdsourcing Platform

6. Conclusion

In the paper, we designed and developed the crowdsourcing platform based on the subjective and objective QoE. The platform was designed such that the user subjective QoE data is collected via the questionnaire, simultaneously the platform can collect the objective QoE data automatically without affecting the user to use video services. Using a management platform administrators can view and analyze the subjective and objective QoE data comprehensively. The results of analysis and evaluation can better reflect the quality of user experience and provide a valuable reference for improving the quality of user experience. The platform provides storage space to upload, manage and share their videos to other social media clouds.

The design and development of this platform have improved the video services to create an application-level video cloud service system to provide video fluency switching,

video caching optimization and other new features. We have improved the QoE evaluation system to collect user objective data and monitor the operation of the server and the QoE data quantitative analysis for the negative and positive feedback of the user. In the future, we will provide the design and development architecture of Android mobile to access video services of crowdsourcing platforms from a remote location giving freedom of mobility to access cloud-based video services. This work is proposed for video streaming but did not support gaming streaming, so in the future, a QoE crowdsensing platform will be proposed for gaming, which will provide services according to SLA and track record of user activities.

Acknowledgments. This study was supported by the Scientific Research Funds of Education Department of Liaoning Province in 2021 (General Project) (LJKZ1311).

References

1. Laghari, Asif Ali, Hui He, Shahid Karim, Himat Ali Shah, and Nabin Kumar Karn. "Quality of Experience Assessment of Video Quality in Social Clouds," *Wireless Communications and Mobile Computing*, pp. 1-10, 2017.
2. Vaishnavi, Ishan, Pablo Cesar, Dick Bulterman, and Oliver Friedrich. "From IPTV services to shared experiences: Challenges in architecture design," *In Proc of Multimedia and Expo (ICME), 2010 IEEE International Conference on*, pp. 1511-1516. IEEE, 2010. Suntec City, Singapore.
3. Aldhaibani O A, Al-Jumaili M H, Raschella A, et al. "A centralized architecture for autonomic quality of experience oriented handover in dense networks," *Computers & Electrical Engineering*, vol. 94, no. 2, pp. 107352, 2021.
4. Hsu, Wu-Hsiao, and Chi-Hsiang Lo. "QoS/QoE mapping and adjustment model in the cloud-based multimedia infrastructure," *IEEE Systems Journal*, vol. 8, no. 1, pp. 247-255. (2014)
5. C. G. Bampis, Z. Li, I. Katsavounidis, T. -Y. Huang, C. Ekanadham and A. C. Bovik. "Towards Perceptually Optimized Adaptive Video Streaming-A Realistic Quality of Experience Database," *IEEE Transactions on Image Processing*, vol. 30, pp. 5182-5197, 2021.
6. Laghari, Asif Ali, Hui He, Shehnila Zardari, and Muhammad Shafiq. "Systematic Analysis of Quality of Experience (QoE) Frameworks for Multimedia Services," *IJCSNS*, vol. 17, no. 5. (2017)
7. Feldman, Michael, and Abraham Bernstein. "Behavior-Based Quality Assurance in Crowdsourcing Markets," *In Second AAAI Conference on Human Computation and Crowdsourcing*. (2014)
8. Ligu Wang, Yin Shoulin, Hashem Alyami, et al. "A novel deep learning-based single shot multibox detector model for object detection in optical remote sensing images," *Geoscience Data Journal*, 2022. <https://doi.org/10.1002/gdj3.162>
9. Samet, Nouha, Asma Ben Leta fa, Mohamed Hamdi, and Sami Tabbane. "Real-Time User Experience Evaluation for Cloud-Based Mobile Video," *In Advanced Information Networking and Applications Workshops (WAINA), 2016 30th International Conference on*, pp. 204-208. IEEE. (2016)
10. Wang, Yumei, Xiaojiang Zhou, Mengyao Sun, Lin Zhang, and Xiaofei Wu. "A new QoE-driven video cache management scheme with wireless cloud computing in cellular networks," *Mobile Networks and Applications*, vol. 22, no. 1, pp. 72-82. (2017)
11. Laghari, Asif Ali, Hui He, Muhammad Shafiq, and Asiya Khan. "Assessing effect of Cloud distance on end user's Quality of Experience (QoE)," *In Computer and Communications (ICCC), 2016 2nd IEEE International Conference on*, pp. 500-505. IEEE. (2016)

12. De Pessemier, Toon, Isabelle Stevens, Lieven De Marez, Luc Martens, and Wout Joseph. "Quality assessment and usage behavior of a mobile voice-over-IP service," *Telecommunication Systems*, vol. 61, no. 3, pp. 417-432. (2016)
13. Noor, Tawfeeg S., Niemah I. Osman, and Is-Haka M. Mkwawa. "The impact of gender on the Quality of Experience for video services," *In Automation and Computing (ICAC), 2016 22nd International Conference on*, pp. 488-492. IEEE. (2016)
14. Xia, Huichuan, and Brian McKernan. "Privacy in Crowdsourcing: a Review of the Threats and Challenges," *Computer Supported Cooperative Work (CSCW)*, pp. 1-39. (2020)
15. Hansson, Karin, and Thomas Ludwig. "Crowd dynamics: Conflicts, contradictions, and community in crowdsourcing," *Computer Supported Cooperative Work (CSCW)*, vol. 28, no. 5, pp. 791-794. (2019)
16. Laghari, Khalil Ur Rehman, and Kay Connelly. "Toward total quality of experience: A QoE model in a communication ecosystem," *IEEE Communications Magazine*, vol. 50, no. 4. (2012)
17. Agostino D, Brambilla M, Pavanetto S, et al. "The Contribution of Online Reviews for Quality Evaluation of Cultural Tourism Offers: The Experience of Italian Museums," *Sustainability*, vol. 13, 2021.
18. GuBo, AlazabMamoun, LinZiqi, et al. "AI-Enabled Task Offloading for Improving Quality of Computational Experience in Ultra Dense Networks," *ACM Transactions on Internet Technology (TOIT)*, vol. 22. no. 3, pp. 1-17, 2022.
19. Chen, Yanjiao, Kaishun Wu, and Qian Zhang. "From QoS to QoE: A tutorial on video quality assessment," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 1126-1165. (2015)
20. Jain, Dhruv, Swapnil Agrawal, Satadal Sengupta, Pradipta De, Bivas Mitra, and Sandip Chakraborty. "Prediction of quality degradation for mobile video streaming apps: A case study using YouTube," *In Communication Systems and Networks (COMSNETS), 2016 8th International Conference on*, pp. 1-2. IEEE. (2016)
21. Kimura, Takuto, Masahiro Yokota, Arifumi Matsumoto, Kei Takeshita, Taichi Kawano, Kazumichi Sato, Hiroshi Yamamoto, Takanori Hayashi, Kohei Shiimoto, and Kenichi Miyazaki. "QUVE: QoE Maximizing Framework for Video-Streaming," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 1, pp. 138-153. (2017)
22. Ahammad, Parvez, R. Gaunker, Brian Kennedy, Mehrdad Reshadi, K. Kumar, A. K. Pathan, and Hariharan Kolam. "A flexible platform for QoE-driven delivery of image-rich web applications," *In Multimedia and Expo (ICME), 2015 IEEE International Conference on*, pp. 1-6. IEEE. (2015)
23. Zhou, Chao, Lifeng Sun, Wenming Shi, and Shiqiang Yang. "QOEYE: A Data Driven Platform for QoE Visualization and System Performance Monitoring," *In Proceedings of the 23rd ACM international conference on Multimedia*, pp. 741-742. ACM. (2015)
24. Rainer, Benjamin, Markus Wlzl, and Christian Timmerer. "A web based subjective evaluation platform," *In Quality of Multimedia Experience (QoMEX), 2013 Fifth International Workshop on*, pp. 24-25. IEEE. (2013)
25. Ding, Yifan, Yang Geng, Ruiyi Wang, Yang Yang, and Wenjing Li. "QoE-oriented resource management strategy by considering user preference for video content," *In Network Operations and Management Symposium (APNOMS), 2014 16th Asia-Pacific*, pp. 1-4. IEEE. (2014)
26. Laghari, Asif Ali, Hui He, Muhammad Ibrahim, and Salahuddin Shaikh. "Automatic Network Policy Change on the Basis of Quality of Experience (QoE)," *?Procedia Computer Science*, vol. 107, pp. 657-659. (2017)
27. Garella, Juan Pablo, Eduardo Grampä n, Rafael Sotelo, Javier Baliosian, Jose Joskowicz, Gustavo Guimerans, and Maria Simon. "Monitoring QoE on digital terrestrial TV: a comprehensive approach," *In Broadband Multimedia Systems and Broadcasting (BMSB), 2016 IEEE International Symposium on*, pp. 1-6. IEEE, 2016.
28. Ghadiyaram, Deepti, and Alan C. Bovik. "Massive online crowdsourced study of subjective and objective picture quality," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 372-387. (2016)

29. Li, Xiaofei, and Yue You. "Kano Model Analysis Required in APP Interactive Design based on Mobile User Experience," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 11, no. 11, pp. 247-258. (2016)
30. Hofeld, Tobias, Matthias Hirth, Pavel Korshunov, Philippe Hanhart, Bruno Gardlo, Christian Keimel, and Christian Timmerer. "Survey of web-based crowdsourcing frameworks for subjective quality assessment," *In Multimedia Signal Processing (MMSP), 2014 IEEE 16th International Workshop on*, pp. 1-6. IEEE. (2014)
31. Belda, Jordi, Mario Montagud, Fernando Boronat, Marc Martinez, and Javier Pastor. "Wersync: A web-based platform for distributed media synchronization and social interaction," *ACM TVX*. (2015)
32. Sensory Experience Lab, <http://selab.itec.aau.at>. Accessed 12. 24. 2016.
33. Alhamazani, Khalid, Rajiv Ranjan, Prem Prakash Jayaraman, Karan Mitra, Meisong Wang, Zhiqiang George Huang, Lizhe Wang, and Fethi Rabhi. "Real-time qos monitoring for cloud-based big data analytics applications in mobile environments," *In Mobile Data Management (MDM), 2014 IEEE 15th International Conference on*, vol. 1, pp. 337-340. IEEE. (2014)
34. Ribeiro, Flavio, Dinei Florencio, Cha Zhang, and Michael Seltzer. "Crowdmos: An approach for crowdsourcing mean opinion score studies," *In Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pp. 2416-2419. IEEE. (2011)
35. Keimel, Christian, Julian Habigt, Clemens Horch, and Klaus Diepold. "Qualitycrowd: a framework for crowd-based quality evaluation," *In Picture Coding Symposium (PCS)*, pp. 245-248. IEEE. (2012)
36. Kraft, Sebastian, and Udo Z lzer. "BeagleJS: HTML5 and JavaScript based framework for the subjective evaluation of audio quality," *In Linux Audio Conference, Karlsruhe, DE*. (2014)
37. Gardlo, Bruno, Sebastian Egger, Michael Seufert, and Raimund Schatz. "Crowdsourcing 2.0: Enhancing execution speed and reliability of web-based QoE testing," *In Communications (ICC), 2014 IEEE International Conference on*, pp. 1070-1075. IEEE. (2014)
38. <https://support.hyperic.com/display/SIGAR/Home>.
39. Schopf, Jennifer M., Laura Pearlman, Neill Miller, Carl Kesselman, Ian Foster, Mike D'Arcy, and Ann Chervenak. "Monitoring the grid with the Globus Toolkit MDS4," *In Journal of Physics: Conference Series*, vol. 46, no. 1, p. 521. IOP Publishing. (2006)
40. ITU- T Recommendation P. 910: "Subjective video quality assessment methods for multimedia applications," *International telecommunication Union, Geneva, Switzerland*. (1996)
41. Jisi A and Shoulin Yin. "A New Feature Fusion Network for Student Behavior Recognition in Education," *Journal of Applied Science and Engineering*, vol. 24, no. 2, pp. 133-140, 2021.
42. Laghari, Asif Ali, and Mureed Ali Laghari. "Quality of experience assessment of calling services in social network," *ICT Express*, vol. 7, no. 2, pp. 158-161. (2021)
43. Baloch, Jawwad Ali, Awais Khan Jumani, Asif Ali Laghari, Vania V. Estrela, and Ricardo T. Lopes. "A Preliminary Study on Quality of Experience Assessment of Compressed Audio File Format," *In 2021 IEEE URUCON*, pp. 161-165. IEEE. (2021)

Asif Ali Laghari received the B.S. degree in Information Technology from the Quaid-e-Awam University of Engineering Science and Technology Nawabshah, Pakistan, in 2007 and Master degree in Information Technology from the QUEST Nawabshah Pakistan in 2014. From 2007 to 2008, he was a Lecturer in the Computer and Information Science Department, Digital Institute of Information Technology, Pakistan. In 2015, he joined the school of the Computer Science & Technology, Harbin Institute of Technology, where he is now a PhD student. Currently he is Assistant professor in Sindh Madressatul Islam University, Karachi, Pakistan He has published more than 55 technical articles in scientific journals and conference proceedings. His current research interests include Computer networks, cloud computing, and multimedia QoE management.

He Hui is born in 1974. Dr., associate professor, doctoral tutor. IEEE & IEEE Computer Member, China Computer Society, ACM Association. Harbin Institute of Technology School of computer science and technology. Mainly engaged in computer network, network measurement and simulation, network active defense technology, mobile network security, cloud computing, migration learning and so on. Presided over or participated in the national network information security project key projects.

Asiya Khan received the B.Eng. degree (Hons) in electrical and electronic engineering from the University of Glasgow, Glasgow, U.K., in 1992, the M.Sc. degree in communication, control, and digital signal processing from Strathclyde University, Glasgow, in 1993, and the Ph.D. degree in multimedia communication from the University of Plymouth, Plymouth, U.K. currently Dr. Asiya Khan is a lecturer in Control Systems Engineering, School of Engineering, University of Plymouth, Plymouth United Kingdom. She worked with British Telecommunication Plc from 1993 to 2002 in a management capacity developing various products and seeing them from inception through to launch. She has been Research Assistant in Perceived QoS Control for New and Emerging Multimedia Services (VoIP and IPTV) CFP7 ADAMANTIUM project at the University of Plymouth. She has published several papers in international journals and conferences. Her research interests include video quality of service over wireless networks, adaptation, perceptual modeling, and content-based analysis.

Rashid Ali Laghari received B.E in industrial engineering and Management in 2012 and M.E in 2015 both from Mehran University of Engineering and Technology Jamshoro, Sindh, Pakistan. He has worked 2.5 years of industrial experience as Shift Engineer in the Manufacturing process of MDF BOARD. He joined the School of Mechatronic Engineering, Department of Mechanical Manufacturing and Automation in 2015 in Harbin Institute of Technology where he has achieved the PhD degree in April 2021, and have published number of research articles. His current research interests include FEA and predictive modeling and optimization of Metal Matrix Composite materials machining process, Cutting Simulation and machining process of Titanium alloys, Fabrication of SiCp/Al Metal Matrix Composite Materials through powder metallurgy and functionally graded ceramic cutting tool inserts. Multi Sensing IIoT platform for process monitoring of CNC machines.

Shoulin Yin received the B.S. in Software Engineering from Shenyang Normal University, China, in 2013 and M.S., Computer Applied Technology from Shenyang Normal University, China, in 2015. In 2014, he joined the School of Electronic and Information Engineering, Harbin Institute of Technology, where he was a PhD student. Currently he is an Assistant professor in Shenyang Normal University, Shenyang, China and the Deputy Director of Intelligent Information Processing Laboratory in Shenyang Normal University. He has published more than 100 technical articles in scientific journals and conference proceedings. His current research interests include Software Engineering, AI, Cloud Computing, Network Security, Image processing, Remote Sensing, Pattern Recognition.

Jiachi Wang received the M.S., Computer Applied Technology from Shenyang Normal University, China, in 2020. Mainly engaged in Software Engineering, AI, Cloud Computing, Network Security, Image processing, Remote Sensing, Pattern Recognition.

Received: March 22, 2022; Accepted: September 11, 2022.

A Novel Motion Recognition Method Based on Improved Two-stream Convolutional Neural Network and Sparse Feature Fusion

Chen Chen

Sports Institute, Henan University of Technology
Zhengzhou City, 470001 China
byoungholee@qq.com

Abstract. Motion recognition is a hot topic in the field of computer vision. It is a challenging task. Motion recognition analysis is closely related to the network input, network structure and feature fusion. Due to the noise in the video, traditional methods cannot better obtain the feature information resulting in the problem of inaccurate motion recognition. Feature selection directly affects the efficiency of recognition, and there are still many problems to be solved in the multi-level feature fusion process. In this paper, we propose a novel motion recognition method based on an improved two-stream convolutional neural network and sparse feature fusion. In the low-rank space, because sparse features can effectively capture the information of motion objects in the video, meanwhile, we supplement the network input data, in view of the lack of information interaction in the network, we fuse the high-level semantic information and low-level detail information to recognize the motions by introducing attention mechanism, which makes the performance of the two-stream convolutional neural network have more advantages. Experimental results on UCF101 and HMDB51 data sets show that the proposed method can effectively improve the performance of motion recognition.

Keywords: motion recognition, two-stream convolutional neural network, attention mechanism, sparse feature fusion, low-rank space.

1. Introduction

Motion recognition is a challenging task. Influenced by various factors such as different illumination, complex backgrounds, multiple perspectives, and large intra-class differences [1,2], motion recognition algorithms are mainly divided into two categories: 1) based on traditional machine learning [3-5]; 2) based on deep learning [6-8]. The key of motion recognition algorithm based on traditional machine learning is feature extraction. In the process, it often takes effort to design features that meet the requirements and are easy to implement. However, its ability to represent motions is also limited by the extracted features. Deep learning-based motion recognition algorithms can automatically learn features. But it needs a lot of training data. The effectiveness of automatic feature extraction is closely related to network structure design and network parameter selection.

The most direct method of applying deep learning in motion recognition is to use convolutional neural network (CNN) to recognize each frame of a video, but this method does not take the motion information between continuous video frames into account. Ji et

al., [9] proposed the concept of 3D convolution for the first time, and used 3D convolution kernel to extract spatial and temporal features for motion recognition. Feichtenhofer et al., [10] proposed a two-stream convolutional neural network for motion recognition, which was divided into two parts: spatial flow convolutional network and temporal flow convolution network. The spatial stream convolutional network took a single frame of RGB image as input to represent the static apparent information at a certain moment in the video. Time-flow convolutional network took several successive frames of optical flow images stacked together as input to represent the motion information of objects. Finally, the classification results of the two networks were fused to get the final results. This proposed model broke the leading position of improved dense trajectory extraction algorithm (IDT)[11] in the field of motion recognition.

Tran et al. [12] proposed a new 3D convolutional 3 dimension (C3D), where continuous video frames were stacked as the network input. The 3D convolutional kernel was used to make convolution in the cube formed after the stacking, which had more time dimensions than the 2D convolutional kernel. In this way, motion information could be obtained from continuous frames. The biggest advantage of this algorithm was that the recognition speed was much higher than that of the two-stream algorithm. So far, the motion recognition algorithm had formed two main directions: one was based on two-flow convolutional neural network for motion recognition; The other was based on 3D convolutional neural network for motion recognition.

At present, the mainstream motion recognition network input data sets are RGB images and optical flow images. For the spatial stream convolutional network, the input data is RGB images, and the initial spatial stream network adopts frame by frame input. However, the current publicly available data sets can often be identified by a single frame image. In this case, there is a lot of redundant information in the input of the spatial flow convolutional network. In order to reduce the frame by frame input redundancy between successive frames, Zhu et al. [13] put forward a key frame method to dig the decisive frame and key areas in video, which could improve the accuracy and efficiency. Although the extraction method of key frames could be integrated into one training network, it was similar to the object detection network RCNN, it first extracted the candidate boxes and then selected key frames. Kar et al. [14] proposed an ADASCAN feature aggregation method to judge the importance degree of different frames, and accordingly to achieve the purpose of improving accuracy and efficiency. The overall model of this method was simpler than the previous one. For the time flow convolutional network, the input data was the optical flow image, and the optical flow extraction was time-consuming and labor-consuming. The motion features contained in the optical flow might not be the optimal features.

Many researchers have improved optical flow for motion recognition. Zhu et al. [15] proposed a dual-flow convolutional network, which added MotionNet before the time-flow network to generate optical flow images and served as the input of the time-flow convolutional network. This method improved the quality of optical flow. Sevilla-Lara et al. [16] proved that the optical flow was effective for motion recognition because of its invariant apparent features, and the end-point-error (EPE) had no strong correlation with the accuracy of motion recognition. It could be seen from the tested optical flow algorithm that the accuracy of optical flow at the boundary and small displacement had a strong correlation with the performance improvement of the motion recognition algo-

rithm. Meanwhile, the loss function value of motion recognition was used to improve the optical flow, so that the recognition accuracy could be improved. Similarly, due to the disadvantages of optical flow images, many researchers have done some works in finding features that can replace optical flow. Zhang et al. [17] used motion vectors to replace optical flow. The motion vector was originally used for video compression, and it could be extracted directly without extra calculation, which greatly accelerated the recognition speed of the two-stream convolutional network, but the accuracy was reduced. Choutas et al. [18] proposed a new posture feature, which could be used for motion recognition by extracting the trajectory of the key joints of the human body, which formed posture feature map for motion recognition. It was complementary to the features provided by RGB and optical flow images, but performed poorly with single feature. Only by changing the interaction mode of dual-stream network and extracting new motion features as network input, the problem of accuracy and speed could not be solved at the same time. The change of network structure also played a decisive role in the improvement of algorithm performance.

In recent years, the main structure of motion recognition network is based on dual-stream network and 3D convolution network. Wang et al. [19] proposed a temporal segment network (TSN), which used multiple dual-stream networks to extract and fuse short-term motion information at different timing positions, so as to solve the problem that the traditional dual-stream only paid attention to apparent features and short-term motion information. Lan et al. [20] inherited the excellent characteristics of TSN and carried out weighted fusion for short-term motion information at different temporal positions. Zhou et al. [21] proposed the temporal inference network, which was based on TSN and added the three-layer fully connected network to learn the weight of video frames with different lengths, and carried out temporal inference for video frames with different lengths. Finally, it obtained the results by fusion. Xu et al. [22] proposed R-C3D(region convolutional neural network) by combining C3D and Faster-RCNN[23]. R-C3D used 3D convolution to extract video features. The idea of the Faster-RCNN is adopted, that is, the proposal is first generated, then the candidate region was pooled. Finally, the classification and boundary regression were performed. The network could recognize the behavior of video with any length. The speed was high and accuracy was improved. Chen et al. [24] modified the 3D convolution used in behavior recognition and proposed P3D residual net (pseudo 3D residual Net). 133 convolution and 311 convolution were used to replace 333 convolution. The former was similar to 2D convolution to extract spatial flow features, while the latter was used to obtain temporal flow features. This method greatly reduced the amount of calculation. Two-stream convolutional network and 3D convolutional neural network could extract time stream information, while long-short-term-memory (LSTM)[25] could also conduct time dimension modeling, which was also a popular direction in the field of motion recognition at present. Jiang et al. [26] proposed a multi-modal LSTM structure combining with attention mechanism with high stability. Du et al. [27] introduced the attitude attention mechanism combining with LSTM and CNN structure, which could effectively extract space-time features. In addition, other researchers have studied the common deep networks. Duan et al. [28] proposed a new non-local network structure, which regarded non-local operations as an efficient, simple and universal component that could be used to capture long-distance dependencies in neural networks. Deep learning algorithms mainly include dual-stream structure and 3D

convolution, the dual-stream structure has high accuracy and slow speed. However, the 3D convolution is faster and has slightly low accuracy. They are all higher than the traditional machine learning algorithms on the whole. It has great advantages over traditional algorithms in dealing with complex backgrounds and large changes within the class.

Aiming at the limitation that the input data of mainstream two-stream convolution network is RGB image and optical flow image, this paper uses sparse features in low-rank space to effectively capture the information characteristics of motion objects in the video and supplement the network input data. At the same time, in view of the lack of information mutual characteristics in the network, we combine the high-level semantic information and low-level detail information to jointly identify motions, so that the network performance has more advantages.

The main contribution of this paper has three aspects:

- We research the dual-stream convolutional neural network based on temporal segmentation network, a temporal segmentation network combining sparse features is proposed to better focus on motion objects.
- A multi-layer feature fusion temporal segmentation network for motion recognition is proposed to solve the problem of low feature utilization.
- We fuse the high-level semantic information and low-level detail information to recognize the motions, which makes the performance of the two-stream convolutional neural network have more advantages.

2. Related Works

2.1. Two-stream convolutional neural network

Two-stream convolutional neural network is divided into spatial flow convolutional neural network and time flow convolutional neural network. The two convolutional neural networks process spatial dimension and temporal dimension of video, and extract spatial information and temporal information, respectively. The basic structure of two-stream convolutional neural network is shown in figure 1. Here, spatial information refers to the scene, object and other information in the video. Time information refers to the motion information of objects in the video.



Fig. 1. Two-stream convolutional neural network

The input of spatial flow convolutional neural network is a single frame RGB image, which can effectively recognize human motion in static image. The network structure is similar to the commonly used image classification network, usually using Alexnet, VGG16, GoogleNet and other deep models as spatial flow convolutional neural network. Generally, pre-training is performed on ImageNet, and then the pre-trained parameters

are transferred to spatial flow network to improve the speed and performance of network training. The input of the time stream convolutional neural network is the stacked continuous frame optical flow image, which can represent the motion information of the object in the video. It is a way to show the motion of the object by using the change of pixel in the time domain and its correlation in the continuous frame. By using this characteristic of optical flow, the human motion between successive frames can be recognized effectively. In order to match the feature dimension of spatio-temporal network fusion, the structure of time-flow network is usually the same as that of spatial flow convolution network.

The fusion of two-stream network refers to the fusion between spatial flow network and time flow network, which is generally divided into two forms. 1) The spatial flow and temporal flow convolution networks carry out result fusion after their Softmax layers. Usually, the average method and the weighted method are used to fuse the scores of different categories to get the final result. 2) The spatio-temporal network is fused at the middle feature layer. Generally, a hybrid spatio-temporal convolution network is formed after the fusion of spatio-temporal features at a certain network layer. Another fusion method is to retain pure spatial flow convolutional network or temporal flow convolutional network after forming a mixed spatio-temporal convolutional network. After the Softmax layer, the scores of different categories are fused again to get the final result.

2.2. 3D convolutional neural network

The most direct way to use convolutional neural network in video sequence is to use convolutional neural network to recognize each frame image. However, the processing of single frame image does not consider the information between successive frames. In motion recognition, the occurrence of motion generally lasts a process, and there is motion information between successive frames. Therefore, in order to effectively utilize the motion information between successive frames, reference [29] proposed a 3D convolutional neural network method, that is, 3D convolutional kernel was used in the structure of convolutional neural network for convolution. Compared with 2D convolution kernel, 3D convolution kernel increases the time dimension and can obtain the features of both time and space dimension at the same time, which is better than 2D convolution in the aspect of motion recognition feature representation. 2D convolution is carried out on the basis of a single frame image. Usually, a convolution kernel with the size of 33 is selected, and 2D convolution is applied to the image to output the image. Therefore, 2D convolution network will lose the time information of input signal after each convolution operation. 3D convolution is carried out on several adjacent frames, and the size of the convolution kernel is generally 333. Only 3D convolution can retain the time information of input signal, as shown in figure 2.

3D convolutional neural network reflects the time dimension by stacking multiple continuous image frames together to form a cube, and then using 3D convolutional kernel for convolution in the cube. The depth of the convolutional kernel is less than the number of stacked image frames. Therefore, each feature in 3D convolution is connected by the features of adjacent frames, and the representation on successive frames can obtain the motion information of objects in the video.

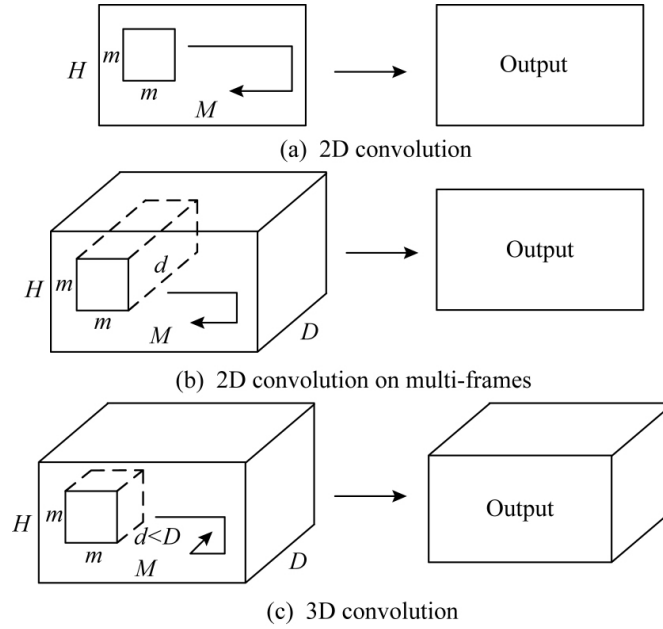


Fig. 2. 2D convolution and 3D convolution

2.3. Temporal segmentation network

Given one video V , it is divided into K segments S_1, S_2, \dots, S_K , each segment is the same, then the temporal segmentation network can be expressed as:

$$Q_{TSN}(T_1, T_2, \dots, T_K) = H(g(F(T_1; W), F(T_2; W), \dots, F(T_K; W))). \quad (1)$$

Where (T_1, T_2, \dots, T_K) is a sequence composed of a single frame in video V . T_k is generated by random sampling of frames in its corresponding video sub-segment S_k , $k \in 1, 2, \dots, K$. $F(T_k, W)$ is the input score prediction function belonging to different categories, that is, the video frame T_k gets a C -dimension vector through the convolutional neural network with parameter W . It represents the predicted number of motions that T_k belongs to class C . $g(\cdot)$ is a segment consensus function, and the prediction results obtained by multiple sub-videos via convolutional neural network are fused to obtain the consistent prediction results $G = (G_1, G_2, \dots, G_C)^T$ about the categories of videos. C represents the number of categories. Based on the above consistent prediction results, function $H(\cdot)$ is used to predict the probability of the entire video belonging to each behavior category. In here, $H(\cdot)$ uses Softmax function, the category with the highest probability is the category that video V belongs to. Combined with the cross-entropy loss commonly used in classification, the category prediction loss function of the final video V can be expressed as:

$$L(y, G) = - \sum_{i=1}^C y_i (G_i - \sum_{j=1}^C \exp G_j). \tag{2}$$

Where, y_i represents the true value of category i . This temporal segmentation network is differentiable determined by the function $g(\cdot)$. The back propagation algorithm and multiple sub-video frames can be used to jointly optimize the model parameter W . G is the consistent prediction result. In the process of back propagation, the gradient of model parameter W for the loss value L is:

$$\frac{\partial L(y, G)}{\partial W} = \frac{\partial L}{\partial G} \sum_{k=1}^K \frac{\partial G}{\partial F(T_k)} \frac{\partial F(T_k)}{\partial W}. \tag{3}$$

Where, K is the number of sub-video segments used by TSN. TSN learns model parameters from the entire video. At the same time, a sparse time sampling strategy is adopted for K , in which the sampled segment only contains a small portion of frames. Compared with the previous method using dense sampling frames, this method greatly reduces the computational overhead, and the structure of timing segmentation network is shown in figure 3.

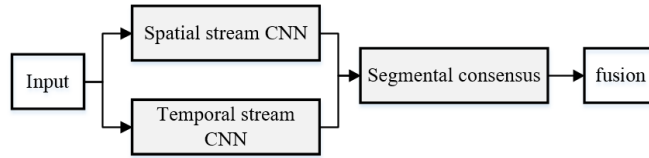


Fig. 3. Temporal segment network

3. Proposed Motion Recognition Method

This section will detailed introduce the network input data and network structure from two aspects. 1) The network input data of sparse features fusion is studied. The purpose is to focus the sparse features on the foreground target in the video, which can effectively extract the motion objects in the image, reduce the redundant information, and complement the information contained in RGB image and optical flow image. 2) It uses convolutional neural network visualization to verify that the shallow convolution can extract detailed features and deep convolution can extract semantic features. The combination of semantic information of high level features and detailed information of low level features in the deep network, and the complementary advantages of features between different convolution layers are helpful for the network to capture the overall features of human behavior and the detail features between different categories. It can improve the accuracy of motion recognition. Figure 4 is the flow chart of the proposed algorithm. The specific steps are as follows: (1) Evenly dividing the input video into three sub-videos, randomly sampling

the three sub-videos to obtain the RGB, optical flow and sparse images of the samples, and input them into the convolutional network respectively; (2) Extracting the features of different convolutional layers of each data type, and fusing the features extracted by the convolutional network according to different sample types; (3) Using the Softmax function for motion classification.

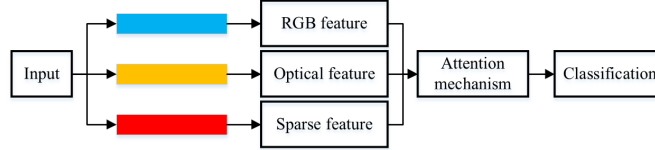


Fig. 4. Proposed temporal segment network based on feature fusion for motion recognition

3.1. Sparse feature

In many practical applications, the known data matrix D is often low-rank or approximately low-rank, but there are errors with arbitrarily large random amplitudes and sparse distribution, which destroy the low-rank of the original data. In order to restore the low-rank structure of matrix D , the matrix D can be decomposed into the sum of two matrices, that is, $D = A + E$, where matrices A and E are unknown, but A is low-rank and E is sparse.

When the elements of matrix E obey the independent Gaussian distribution, the classical principal component analysis method can be used to obtain the optimal matrix A , which is transformed into an optimization problem.

$$\min_{A,E} \|E\|_F, s.t. rank(A) \leq r, D = A + E. \quad (4)$$

Where $\|\cdot\|_F$ denotes the Frobenius norm of the matrix.

When E is a sparse large noise matrix, PCA cannot give ideal results. So the robust principal component analysis (RPCA) can be used to obtain the optimal matrix A , then the problem in equation (4) can be transformed into an optimization problem:

$$\min_{A,E} rank(A) + \lambda \|E\|_0, s.t. D = A + E. \quad (5)$$

Where the rank function $rank(\cdot)$ and 0-norm of matrix are non-convex. They become NP-hard problem, which needs to be relaxed. Since the kernel norm is the convex hull of the rank function, and the 1-norm is the convex hull of the 0-norm, the NP-hard problem of equation (5) can be transformed into a convex optimization problem after relaxation:

$$\min_{A,E} \|A\| + \lambda \|E\|_1, s.t. D = A + E. \quad (6)$$

Where A is the low-rank component and E is the corresponding sparse component. $\|\cdot\|_*$ represents the kernel norm of the matrix, it is the sum of the singular values of the matrix, and it is also the convex approximation of $rank(\cdot)$. $\|\cdot\|_1$ denotes the L1 norm.

$\|\cdot\|_1$ is a weighted parameter greater than zero to balance the two norms. Under certain conditions, it has been proved that as long as the error matrix E is sparse enough relative to matrix A , the low-rank component and the sparse component can be accurately recovered from the matrix D by solving the convex optimization problem (equation (4)), that is, the weighted combination of the above kernel norm and L1 norm can be minimized.

For the RPCA problem described in equation (6), the augmented Lagrange multiplier method can be used to optimize it. The Lagrange function is:

$$L(A, E, Y, \mu) = \|A\|_s + \lambda\|E\|_1 + \langle Y, D - A - E \rangle + \frac{\mu}{2}\|D - A - E\|_F^2. \quad (7)$$

Where Y is the Lagrange multiplier and μ is a smaller positive number.

RPCA is widely used in image and video processing such as image correction, denoising, video background modeling, foreground target extraction, image segmentation, saliency detection [32]. For foreground target segmentation in video, the background is approximated as a low-rank component due to the correlation between frames. However, the foreground target only occupies a small part of the pixels in the image, such as the human motion. The moving part can be regarded as the sparse component. Through the above augmented Lagrange multiplier method to solve the RPCA problem, the sparse features shown in figure 5 can be obtained for the motion video.

In Figure 5, the first row represents RGB image, the second row represents the motion optical flow image in the x-axis direction, the third row represents the motion optical flow image in the y-axis direction, and the fourth row represents the sparse image. As can be seen from figure 5, the RGB image represents the apparent features of the image, including both background and foreground targets. The optical flow image represents the movement direction and speed of the moving object in the image. For the x-axis direction, white indicates the movement to the right, and the higher gray value denotes the faster speed. black indicates the movement to the left, and the lower gray value denotes the faster movement. The rest of the gray area means that nothing is moving. It is the same as the y-axis, white means moving up, black means moving down. Unlike color and optical flow images, sparse feature images can focus on the behaviors of foreground targets and effectively extract motion objects. Meanwhile, removing background can effectively reduce data redundancy and significantly improve the speed of network training.

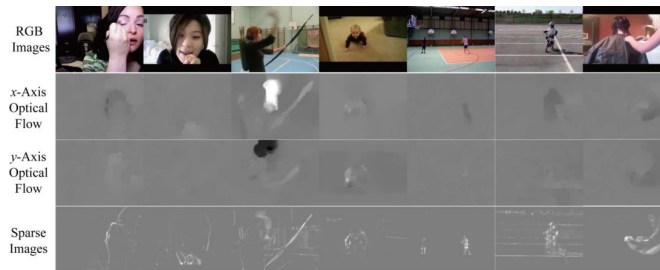


Fig. 5. Comparison of RGB, optical flow and low rank data

3.2. Network feature fusion with attention mechanism

In view of the lack of information interaction in the deep network, the deep network combines the high-level semantic information and low-level detail information to jointly identify the motion, so that the network performance has more advantages.

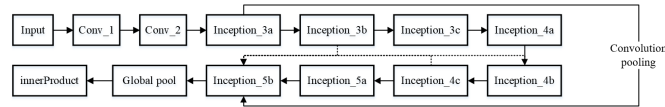


Fig. 6. Multilayer convolutional feature neural network

Multi-layer feature fusion is based on the low-level detail feature and high-level semantic feature of convolutional neural network, and the features of different deep convolutional layer features are used to achieve the fusion. We take InceptionV2 network as an example to illustrate the principle of the improved convolutional neural network as shown in figure 6. The network is composed of multi-stream convolutional neural networks. For the spatial-flow convolutional neural network, assuming that the input color image size is $224 \times 224 \times 3$, the convolution kernel of 7×7 and the step size of 2 are first selected. The convolution layer is used to extract the features of the input image, and $64 \times 112 \times 112$ feature maps are obtained. Then, the maximum pooling is performed to obtain the 5656 feature map. The convolution kernel of 3×3 and step size of 2 is selected, and the pooling features are extracted by reconvolution and pooling. The pooled feature size is $28 \times 28 \times 192$. Then, the obtained features are successively passed through 10 Inception structural units, from structural units Inception3a to Inception5b, and the size of the obtained features is $7 \times 7 \times 1024$. After one average pooling, it outputs the feature with $1 \times 1 \times 1024$, the 1D vector is expanded as the input of the fully connected layer. At the same time, the output features after shallow convolution are also expanded as 1D vectors and sent to the fully connected layer. Finally, the shallow convolution features and deep convolution features are input into the fully connected layer to form a 1×101 vector.

As shown in figure 6, the fusion process of multi-layer convolution features is illustrated by taking the output features of inception3a layer and Inception5b as an example. In order to clearly illustrate the fusion principle of features with high and low dimensions, table 1 lists the feature size output by each layer of the convolutional neural network.

Firstly, $28 \times 28 \times 192$ feature map is obtained after the input image goes through the first two convolution layer and pooling layer. The first 2-dimensional data represents the length and width of the feature map, and the third dimension data represents the number of channels. Then, the features are fed into the Inception3a layer, and four groups of features are obtained respectively through the four branches of the Inception structure unit. The four groups of features are connected in series as the input of the next layer. Meanwhile, the pooling operation is used for this feature. We select the average pooling, because it can reduce dimension and retain more image background information. It is beneficial to transfer the information to the next module for feature extraction, and make its size the same as the deep convolution feature size, which is convenient for feature fusion. In addition, because feature fusion will increase feature dimension and computational complexity, the

Table 1. Map size in each network layer

Network layers	Kernel size	Stride	Output size
Convolution-1	7×7	2	$112 \times 112 \times 64$
Pooling	3×3	2	$56 \times 56 \times 64$
Convolution-2	3×3	1	$56 \times 56 \times 192$
Pooling	3×3	2	$28 \times 28 \times 192$
Inception3a	—	—	$28 \times 28 \times 256$
Inception3b	—	—	$28 \times 28 \times 320$
Inception3c	2	2	$28 \times 28 \times 576$
Inception4a	—	—	$14 \times 14 \times 576$
Inception4b	—	—	$14 \times 14 \times 576$
Inception4c	—	—	$14 \times 14 \times 576$
Inception4d	—	—	$14 \times 14 \times 576$
Inception5a	—	—	$7 \times 7 \times 1024$
Inception5b	—	—	$7 \times 7 \times 1024$
Pooling	—	—	$1 \times 1 \times 1024$

shallow convolution features are obtained by reducing dimension with convolution kernel of 1×1 . The shallow convolution feature is connected in series with the output feature of inception5b layer and expanded into a 1-dimensional vector as the input of the fully connected layer. Time flow convolutional network and sparse convolutional neural network are similar to spatial flow convolutional neural network. Shallow convolutional features are obtained according to the above networks, and they are fused with the deep convolutional features output by the last layer of Inception structural unit to participate in the final classification. For the two feature maps $x_t^a \in R^{H \times M \times D}$ and $x_t^b \in R^{H' \times M' \times D'}$, they are will be utilized to generate the various feature map $y_t \in R^{H'' \times M'' \times D''}$. In here, t is the time. H, M and D represent the height, width and channel number of the three feature maps respectively. Because the cascade fusion is simple and efficient, this paper uses the cascade fusion method to fuse the low-level detail information and the high-level semantic information. The low-level detail information mainly extracts the color, texture and other detail features. While the high-level semantic information is more representative, it can extract the detailed feature for motion recognition. Through fusion process, features can be fully utilized to improve the recognition accuracy.

Series fusion means that the feature maps of the corresponding channels of the two features are connected in sequence, and the combined features are taken as new features, namely:

$$y_{i,j,2d} = x_{i,j,d}^a \tag{8}$$

$$y_{i,j,2d-1} = x_{i,j,d}^b \tag{9}$$

Where $1 \leq i \leq H, 1 \leq j \leq M, 1 \leq d \leq D$, and $x^a, x^b \in R^{H \times M \times D}, y \in R^{H \times M \times D}$.

Drawing on the signal processing mechanism of the human brain, the attention mechanism quickly scans all features to obtain the feature categories that need to be focused on, and assigns corresponding attention weights according to the critical degree of feature

categories, so that the brain can process huge information with limited resources. The application in CNN is reflected in the difference of importance of each generated feature map. The core objective of attention mechanism is to obtain the difference of importance between each feature map by calculation, allocate computing resources according to its importance, and use the execution effect to guide the feature in reverse map weights are updated, and the task is finally completed efficiently and accurately. The implementation principle of attention module is shown in Figure 7.

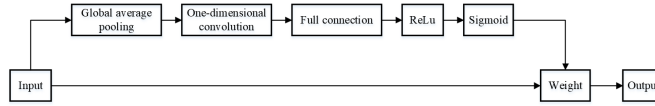


Fig. 7. Schematic of attention mechanism

Specific implementation methods are as follows. Firstly, each feature map obtained by convolution is global average pooling operation, and each feature map is extruded into a real number, as shown in Equation (10). The squeezed real numbers of each feature map are combined into a vector, namely the weight of each feature. After the weight vector is obtained, the full connection, ReLu activation function and sigmoid activation function are shown in Equation (11). The weighted feature map is obtained by assigning weights to each feature category. Finally, equation (12) is used to guide the feature map to update in a direction conducive to the recognition task.

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(ij). \quad (10)$$

$$s = \sigma(W_2 \delta(W_1 z)). \quad (11)$$

$$x_c = s_c u_c. \quad (12)$$

Where H and W represent the length and width of feature map. u_c stands for the result after convolution. z_c represents the importance of each feature map. s_c is the weight vector of all feature maps. σ stands for Relu activation function. δ represents Sigmoid activation function. W_1 and W_2 are two different fully connected operations.

4. Experiments and Analysis

In this section, experiments are conducted on two large motion datasets to verify the effectiveness of the feature fusion temporal segmentation network. The two datasets are UCF101 and HMDB51 respectively. UCF101 dataset contains 101 motion categories and 13320 video clips. The HMDB51 dataset is a large number of realistic videos from a variety of sources, such as movies and web videos. The dataset consists of 6849 video clips from 51 motion categories. The experiment followed the original evaluation scheme using three training/test groups, namely dataset group 1, dataset group 2, and dataset

group 3. It takes the average accuracy of these groups as the final motion recognition accuracy.

The experiment in this section uses the small-batch random gradient descent algorithm to learn network parameters. The batch size is set to 32 and the momentum is set to 0.9. In addition, the model trained in advance with the dataset is used to initialize the network weights, and a small learning rate is set in the experiment. For spatial network, the learning rate is initialized to 0.001 and reduced by 1/10 per 2000 iterations. The entire training process stops at 10000 iterations. For temporal network and sparse network, the initial learning rate is set to 0.005, which is reduced to 1/10 after 12000 and 18000 iterations. The maximum iteration number is set to 20000. In order to extract optical flow quickly, the TVL1 optical flow algorithm implemented by CUDA in OpenCV is selected. To speed up the training, multiple GPUs are adopted implemented by using Caffe and OpenMP12.

4.1. Experiment datasets

UCF101 has 13320 videos including 101 categories. There are great changes in the aspects of camera movement, object appearance and posture, object proportion, viewpoint, messy background, lighting conditions, etc.,. In addition, the motion videos are all edited rather than performed by actors, which is the most challenging dataset to a certain extent. Some motion categories in the dataset are shown in figure 8.

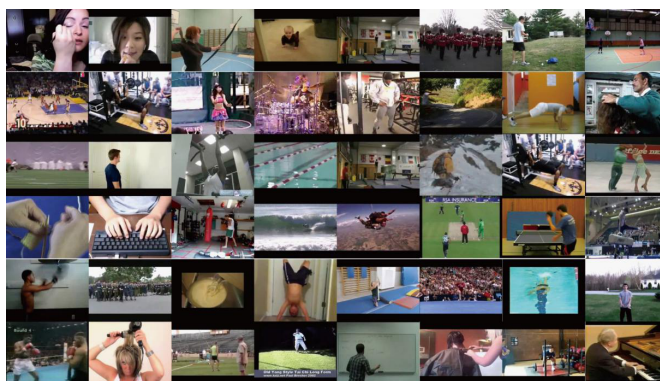


Fig. 8. Partial action categories in UCF101 dataset

The HMDB51 dataset contains 6766 video clips divided into 51 motion categories. Each action category contains at least 101 video clips. Some of the motion categories are shown in figure 9. Most samples of the HMDB51 dataset come from movies, while a small amount of samples come from public video sites such as Prelinger Archive, YouTube and Google Video.

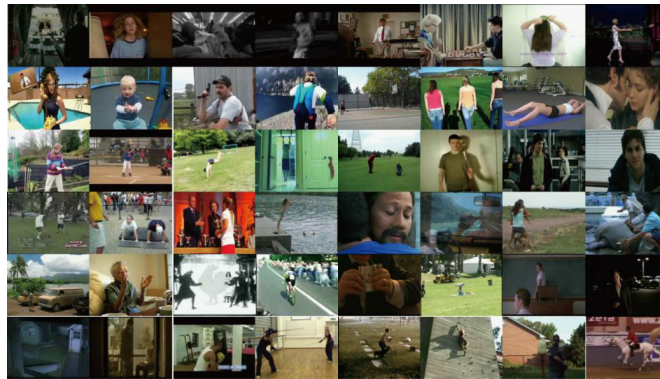


Fig. 9. Partial action categories in HMDB51 dataset

4.2. Experimental Environment

Deep learning hardware environment: CPU E5-2696V4, GPU, two GTX1080Tis, 256 GB SSD, 32 GB memory. Network learning and test environment: Ubuntu16.04, NVIDIA CUDA8.0, cudnn5, Caffe, opencv3.0, Python.

4.3. Effect of sparse feature

The experiment is conducted on the two public motion recognition datasets UCF101 and HMDB51, and we compare the proposed method with some classical algorithms and commonly used algorithms in recent years. The comparison results are shown in table 2.

As can be seen from table 2, the compared algorithms are divided into three categories. And figure 10 is the visual diagram for table 2. The first category is the traditional classical machine learning algorithm without deep learning. This algorithm manually extracts motion features and has high stability. The recognition rate can reach about 88% on UCF101 and exceed 61% on HMDB51. For example, the MoFAP method is a combinatorial motion feature way, which consists of three parts: local motion feature, motion atom, and motion statement. The motion atom refers to a certain sub-stage in the process of motion, and the motion statement is the combination of these sub-stages. For example, the high jump is divided into three sub-stages, the run-up, take-off and landing, namely the motion atom. The different combinations of the three become motion statement. In this way, the ability of features to represent motion is stronger, so as to improve the recognition accuracy.

The second category is the deep learning algorithm based on 3D convolution. This algorithm is fast, it can achieve real-time requirement, and the recognition rate is higher than that of the traditional algorithm. For example, reference [48] verified that different actions had different temporal and spatial patterns, and some motions could require a long time to be recognized. So LTC network structure was proposed to improve the recognition accuracy by increasing the duration of input video.

The third category is based on two-stream convolutional neural networks. This kind of algorithm has the highest accuracy, which can exceed 88%. As can be seen from table 2,

Table 2. Accuracy comparison with different methods on UCF101 and HMDB51/%

Method	UCF101	HMDB51
DT+MVSV[30]	83.7	56.1
IDT+FV[31]	86.1	57.4
IDT+HSV[32]	88.1	61.3
MoFAP[33]	88.5	61.9
C3D+IDT[12]	90.6	62.8
TDD+IDT[34]	91.7	66.1
LTC[35]	91.9	65.0
LTC+IDT[35]	92.8	67.4
P3D ResNet+IDT[36]	93.9	68.2
Two streams[37]	88.2	59.6
Two streams+LSTM[38]	88.8	63.4
Two streams fusion[10]	92.7	65.6
Transformations[39]	92.6	62.2
TSN(RGB+Optical Flow)[19]	94.2	69.4
RCA[40]	95.7	72.2
FFN[41]	96.9	74.3
Sparse+TSN	97.1	76.6

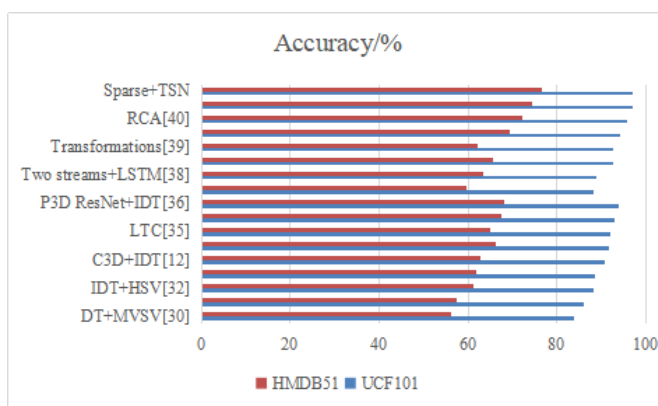


Fig. 10. Visual diagram of table 2

the recognition rate of the temporal segmentation network with sparse features are 97.1% and 76.6% on UCF101 and HMDB51 respectively.

4.4. Multi-layer feature fusion experiment

In order to verify the effectiveness of multi-layer convolution feature fusion convolution network, the experiment of UCF101 group 1 is taken as an example. The outputs from inception3a to inception5a are fused with the feature of inception5b. Table 3 lists the recognition rate of the temporal segmentation network trained by RGB, optical flow images and sparse images after adding the multi-layer feature fusion method. Similar to RGB image, the temporal segmentation network trained by optical flow image and sparse image is also fused with the convolution feature output by Inception5a layer and inception5b layer, and the highest recognition rate is obtained, which reaches 93.68% and 86.22% respectively. The optical flow basically remains unchanged. The recognition rate of sparse network is more than 0.6% higher than that of network without shallow convolution feature fusion, indicating that the addition of shallow convolution feature can improve the network performance.

Table 3. Comparison recognition rate with different convolution layers fusion on UCF101/%

Fusion layer	RGB	Optical flow	Sparse
Inception3a → Inception5b	87.82	92.71	85.14
Inception3b → Inception5b	87.97	93.15	84.92
Inception3c → Inception5b	87.95	93.10	85.06
Inception4a → Inception5b	87.97	92.98	85.48
Inception4b → Inception5b	87.25	93.03	85.98
Inception4c → Inception5b	87.22	92.89	85.57
Inception4d → Inception5b	87.83	93.12	85.83
Inception4e → Inception5b	88.21	92.89	86.09
Inception5a → Inception5b	88.34	93.68	86.22

In order to further verify the effect of the multi-layer feature fusion on temporal segmentation network, the experiment is conducted on UCF101 and HMDB51 data sets. Compared with some classical algorithms and commonly used algorithms, the comparison results are shown in table 4.

As can be seen from table 4, the motion recognition temporal segmentation network with multi-layer feature fusion has a certain improvement compared with the original temporal segmentation network with sparse feature fusion. The recognition rates of UCF101 and HMIDB51 are 97.3% and 76.9%, indicating that shallow convolutional layer and deep convolutional layer fusion have a certain effect on the improvement of network performance. The confusion matrix is shown in figure 11 and figure 12.

Table 4. Accuracy comparison with different methods on UCF101 and HMDB51/%

Method	UCF101	HMDB51
C3D+IDT	90.6	62.8
TDD+IDT	91.7	66.1
LTC	91.9	65.0
LTC+IDT	92.8	67.4
P3D ResNet+IDT	93.9	68.2
Two streams	88.2	59.6
Two streams+LSTM	88.8	63.4
Two streams fusion	92.7	65.6
Transformations	92.6	62.2
TSN(RGB+Optical Flow)	94.2	69.4
Sparse+TSN	97.3	76.9

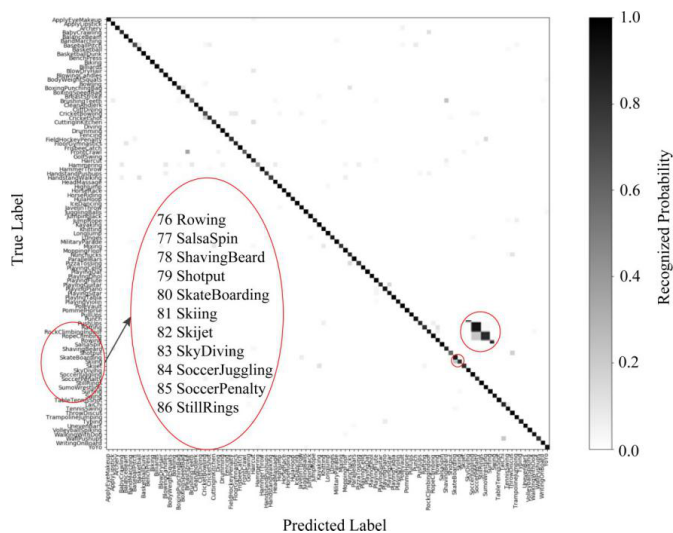


Fig. 11. Confusion matrix on UCF101 dataset

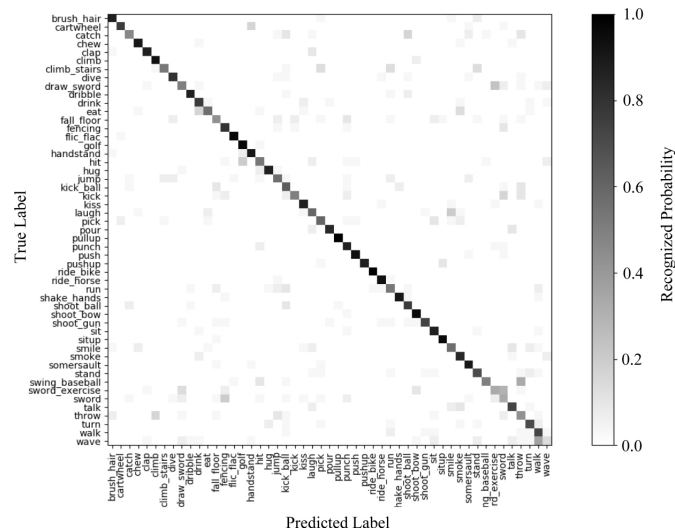


Fig. 12. Confusion matrix on HMDB51 dataset

5. Conclusion

In this paper, the two-stream convolutional neural network based on temporal segmentation network is studied and a temporal segmentation network with sparse features is proposed. Meanwhile, to solve the problem of low feature utilization, a multi-layer feature fusion temporal segmentation network for motion recognition is proposed. Based on the motion recognition network of sparse features and multi-layer feature fusion, the recognition effect of the new algorithm on the UCF101 and HMDB51 is better than other algorithms. In the future, more deep learning-based models will be utilized for action recognition.

References

1. Yao, G., Lei, T., Zhong, J. "A Review of Convolutional-Neural-Network-Based Action Recognition," *Pattern Recognition Letters*, vol. 118, pp. 14-22. (2018)
2. Li, H., Ding, Y., Li, C., et al., "Action recognition of temporal segment network based on feature fusion," *Journal of Computer Research and Development*, Vol. 57, No. 1, pp. 145-158. (2020)
3. Olivieri, D. N., Conde, I.G., Sobrino, X.A.V. "Eigenspace-based fall detection and activity recognition from motion templates and machine learning," *Expert Systems with Applications*, Vol. 39, No. 5, pp. 5935-5945. (2012)
4. Zheng, D., Li, H., Yin, S. "Action Recognition Based on the Modified Two-stream CNN," *International Journal of Mathematical Sciences and Computing (IJMSC)*, Vol. 6, No. 6, pp. 15-23. (2020)
5. J. Long, X. Wang, W. Zhou, J. Zhang, D. Dai and G. Zhu. "A Comprehensive Review of Signal Processing and Machine Learning Technologies for UHF PD Detection and Diagnosis (I): Preprocessing and Localization Approaches," *IEEE Access*, vol. 9, pp. 69876-69904, (2021).

6. Wang, P., Li, W., Ogunbona, P., et al. "RGB-D-based Human Motion Recognition with Deep Learning: A Survey," *Computer vision and image understanding*, Vol. 171, pp. 118-139. (2017)
7. Kim, K., Yong, K.C. "Effective inertial sensor quantity and locations on a body for deep learning-based worker's motion recognition," *Automation in Construction*, Vol. 113. (2020)
8. Yin, S., Li, H. "GSAPSO-MQC:medical image encryption based on genetic simulated annealing particle swarm optimization and modified quantum chaos system," *Evolutionary Intelligence*, vol. 14, pp. 1817-1829. (2021)
9. Ji, S., Xu, W., Yang, M., and Yu, K. "3D Convolutional Neural Networks for Human Action Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 1, pp. 221-231. (2013)
10. Feichtenhofer, C., Pinz, A., Zisserman, A. "Convolutional Two-Stream Network Fusion for Video Action Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1933-1941.
11. Wang, H., Schmid, C. "Action Recognition with Improved Trajectories," *2013 IEEE International Conference on Computer Vision, 2013*, pp. 3551-3558.
12. Tran, D., Bourdev, L., Fergus, R., et al. "Learning Spatiotemporal Features with 3D Convolutional Networks," *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 4489-4497.
13. Zhu, W., Hu, J., Sun, G., Cao X., et al. "A Key Volume Mining Deep Framework for Action Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1991-1999.
14. Kar, A. Rai, N. Sikka K. and Sharma, G. "AdaScan: Adaptive Scan Pooling in Deep Convolutional Neural Networks for Human Action Recognition in Videos," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5699-5708.
15. Yi, Z., Lan, Z., Newsam, S., et al. Hidden Two-Stream Convolutional Networks for Action Recognition. 2017. arXiv:1704.00389
16. Sevilla-Lara, L., Liao, Y., Güney, F., et al. "On the Integration of Optical Flow and Action Recognition," *Pattern Recognition. GCPR 2018. Lecture Notes in Computer Science*, vol. 11269, pp. 281-297, Springer, Cham. (2019)
17. Zhang, B., Wang, L., Wang, Z., et al. "Real-Time Action Recognition With Deeply Transferred Motion Vector CNNs," *IEEE Transactions on Image Processing*, Vol. 27, No. 5, pp. 2326-2339. (2018)
18. Choutas, V., Weinzaepfel, P., Revaud J. "PoTion: Pose MoTion Representation for Action Recognition," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7024-7033.
19. Wang, L., et al. "Temporal Segment Networks: Towards Good Practices for Deep Action Recognition," *Computer Vision-ECCV 2016. ECCV 2016. Lecture Notes in Computer Science*, vol. 9912, pp. 20-36, Springer, Cham. (2016)
20. Lan, Z., Zhu, Y., Hauptmann, A. G., and Newsam, S. "Deep Local Video Feature for Action Recognition," *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1219-1225. (2017)
21. Zhou, B., Andonian, A., Oliva, A., et al. "Temporal Relational Reasoning in Videos," *Computer Vision-ECCV 2018. ECCV 2018. Lecture Notes in Computer Science*, vol. 11205, pp. 831-846, Springer, Cham. (2018)
22. Xu, H., Das, A., and Saenko, K. "R-C3D: Region Convolutional 3D Network for Temporal Activity Detection," *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 5794-5803. (2017)
23. Yin, S., Li, H., Teng, L. "Airport Detection Based on Improved Faster RCNN in Large Scale Remote Sensing Images," *Sensing and Imaging*,?Vol. 21. (2020).
24. Chen, J., Kong, J., Sun, H. et al. "Spatiotemporal Interaction Residual Networks with Pseudo3D for Video Action Recognition," *Sensors*, Vol. 20, No. 11, 3126. (2020)

25. Jiang, D., Li, H., Yin, S. "Speech Emotion Recognition Method Based on Improved Long Short-term Memory Networks," *International Journal of Electronics and Information Engineering*, Vol. 12, No. 4, pp. 147-154. (2020)
26. Jiang, Y., Wu, Z., Tang, J., et al. "Modeling Multimodal Clues in a Hybrid Deep Learning Framework for Video Classification," *IEEE Transactions on Multimedia*, vol. 20, no. 11, pp. 3137-3147. (2018)
27. Du, W., Wang, Y., Qiao, Y. "RPAN: An End-to-End Recurrent Pose-Attention Network for Action Recognition in Videos," *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 3745-3754.
28. Duan, Z., Zhang, T., Tan, J. et al. "Non-Local Multi-Focus Image Fusion With Recurrent Neural Networks," *IEEE Access*, Vol. 8, pp. 135284-135295. (2020)
29. Byeon, Y.H., Kwak, K.C. "Facial Expression Recognition Using 3D Convolutional Neural Network," *International Journal of Advanced Computer Science & Applications*, Vol. 5, No. 12. (2014).
30. Cai, Z., Wang, L., Peng, X., Qiao, Y. "Multi-view Super Vector for Action Recognition," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 596-603.
31. Luong, V. D., Wang, L., Xiao, G. "Action Recognition Using Hierarchical Independent Sub-space Analysis with Trajectory," *Springer International Publishing*, 2015.
32. Peng, X., Wang, L., Wang, X., et al. "Bag of Visual Words and Fusion Methods for Action Recognition: Comprehensive Study and Good Practice," *Computer Vision & Image Understanding*, Vol. 150, pp. 109-125. (2016)
33. Wang, L., Qiao, Y., Tang, X. "MoFAP: A Multi-level Representation for Action Recognition," *International Journal of Computer Vision*, Vol. 119, No. 3, pp. 254-271. (2016)
34. Wang, L., Qiao, Y., Tang, X. "Action recognition with trajectory-pooled deep-convolutional descriptors," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 4305-4314.
35. Varol, G., Laptev, I., Schmid, C. "Long-Term Temporal Convolutions for Action Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 40, No. 6, pp. 1510-1517. (2018)
36. Qiu, Z., Yao, T., Mei, T. "Learning Spatio-Temporal Representation with Pseudo-3D Residual Networks," *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2017.
37. Simonyan, K., Zisserman, A. "Two-stream convolutional networks for action recognition in videos," *Neural Information Processing Systems*, Vol. 1, No. 4, 568576. (2014)
38. Joe Yue-Hei Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga and G. Toderici. "Beyond short snippets: Deep networks for video classification," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 4694-4702.
39. Wang, X., Farhadi A., and Gupta, A. "Actions Transformations," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2658-2667.
40. Dianhuai Shen, Xueying Jiang, Lin Teng. "Residual network based on convolution attention model and feature fusion for dance motion recognition," *EAI Endorsed Transactions on Scalable Information Systems*, 21(33), e8, 2021. <http://dx.doi.org/10.4108/eai.6-10-2021.171247>
41. Jisi A and Shoulin Yin. "A New Feature Fusion Network for Student Behavior Recognition in Education," *Journal of Applied Science and Engineering*, vol. 24, no. 2, pp. 133-140. (2021)

Chen Chen is a lecturer and doctor at the School of Physical Education of Henan University of Technology. Research direction: Social sports, humanities and Sociology of sports.

Received: May 01, 2022; Accepted: September 08, 2022.

A New Frog Leaping Algorithm-oriented Fully Convolutional Neural Network for Dance Motion Object Saliency Detection

Yin Lyu¹ and Chen Zhang²

¹ Music College, Huaiyin Normal University
Huaian City, 223001 China
8201711037@hytc.edu.cn

² College of Sports Art, Harbin Sport University
Harbin City, 150000 China
sarkozyteague@foxmail.com

Abstract. Image saliency detection is an important research topic in the field of computer vision. With the traditional saliency detection models, the texture details are not obvious and the edge contour is not complete. The accuracy and recall rate of object detection are low, which are mostly based on the manual features and prior information. With the rise of deep convolutional neural networks, saliency detection has been rapidly developed. However, the existing saliency methods still have some common shortcomings, and it is difficult to uniformly highlight the clear boundary and internal region of the whole object in complex images, mainly because of the lack of sufficient and rich features. In this paper, a new frog leaping algorithm-oriented fully convolutional neural network is proposed for dance motion object saliency detection. The VGG (Visual Geometry Group) model is improved. The final full connection layer is removed, and the jump connection layer is used for the saliency prediction, which can effectively combine the multi-scale information from different convolution layers in the convolutional neural network. Meanwhile, an improved frog leaping algorithm is used to optimize the selection of initial weights during network initialization. In the process of network iteration, the forward propagation loss of convolutional neural network is calculated, and the anomaly weight is corrected by using the improved frog leaping algorithm. When the network satisfies the terminal conditions, the final weight is optimized by one frog leaping to make the network weight further optimization. In addition, the new network can combine high-level semantic information and low-level detail information in a data-driven framework. In order to preserve the unity of the object boundary and inner region effectively, the fully connected conditional random field (CRF) model is used to adjust the obtained saliency feature map. In this paper, the precision recall (PR) curve, F-measure, maximum F-measure, weighted F-measure and mean absolute error (MAE) are tested on six widely used public data sets. Compared with other most advanced and representative methods, the results show that the proposed method achieves better performance and it is superior to most representative methods. The presented method reveals that it has strong robustness for image saliency detection with various scenes, and can make the boundary and inner region of the saliency object more uniform and the detection results more accurate.

Keywords: Image saliency detection, dance motion, deep convolutional neural network, frog leaping algorithm, fully connected conditional random field.

1. Introduction

Saliency object detection is the most eye-catching object or region in the image [1]. The result is usually represented by a grayscale image, and the grayscale value of each pixel in the image indicates the probability that the pixel belongs to a saliency object. Saliency object detection has become an important preprocessing step in many computer vision applications, including image and video compression [2], image relocation [3], video tracking [4] and robot navigation [5], etc.

Although the detection performance of the saliency object detection method has been significantly improved, there are still some bottlenecks to be broken through in the computer vision task. Traditional saliency object detection methods focus on the low-level features in the manually selected images, and use a variety of prior knowledge to calculate saliency, such as contrast prior [6], center prior [7], background prior and object prior [8]. However, the detection effect of these models is not satisfactory in practical problems. For example, it is difficult to detect foreground objects when background and foreground objects share some similar visual features (see line 1 in figure 1(c)(d)). In addition, detection may fail when multiple saliency objects partially or completely overlap each other (see line 2 of figure 1(c) (d)).

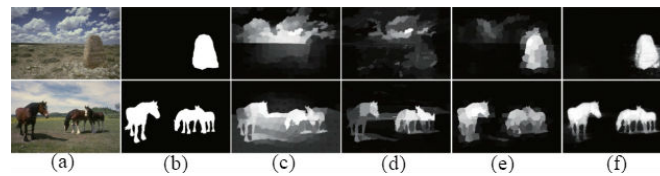


Fig. 1. Comparison with different methods. (a) input images; (b) ground truth; (c) non-deep learning1 method; (d) non-deep learning2 method; (e) deep learning method; (f) proposed

The convolutional neural network (CNN) based methods have successfully overcome the performance bottleneck of traditional manual feature selection methods in many computer vision tasks, such as image classification [9] and semantic segmentation [10], etc.. Similarly, the saliency detection methods based on CNN greatly improve the detection performance. The CNN-based models have demonstrated their advantages in feature extraction and can better capture the high-level semantic information of objects in the complex backgrounds, which can achieve better performance than traditional methods as shown in figure 1(e)(f).

Generally, each object can be represented by three different feature levels, namely low feature, intermediate feature, and high feature. Low-level features correspond to shallow features of deep convolutional networks, such as texture, color, and edge. Intermediate features are related to object shape and contour information, while high-level features are related to object semantic information. Though only using high-level semantic information can improve detection performance, other features levels are also important for detecting saliency objects. Therefore, it is a key and challenging problem to extract and fuse effective feature information of all levels in CNN model. A standard convolu-

tional neural network usually consists of repeated cascade convolutional layers. Deeper convolution layers encode semantic information at the expense of spatial resolution. The shallow convolution layer contains more detailed information about object structure but lacks global properties.

This paper proposes a simple but effective deep convolutional neural network model for saliency detection tasks. It can effectively combine multi-level features to capture unique high-level semantic information and shallow detail information simultaneously in complex images. Meanwhile, an improved frog leaping algorithm is used to optimize the selection of initial weights during network initialization. The new deep network consists of a feature extraction module and a feature fusion module. The feature extraction module can not only generate effective high-level semantic features at different scales, but also capture subtle visual contrast features between low-level and intermediate feature maps for accurate saliency detection. The main contributions of this paper are as follows:

1. A new deep convolutional network based on fully convolutional networks (FCN) is proposed for saliency object detection, which can effectively learn rich multi-scale and multi-level features from complex background images. The model can learn the global and local features of images and avoid the interference of irrelevant background information.
2. An improved frog leaping algorithm is used to optimize the selection of initial weights during network initialization. In the process of network iteration, the forward propagation loss of convolutional neural network is calculated, and the anomaly weight is corrected by using the improved frog leaping algorithm. When the network satisfies the terminal conditions, the final weight is optimized by one frog leaping to make the network weight further optimization.
3. A multi-scale feature fusion mechanism is introduced. The model combines the shallow and deep feature maps in the deep convolutional network, which significantly improves the detection performance and does not need to be supplemented by manually selected features.
4. According to the five commonly used evaluation indexes, the new method is tested on DUT-OMRON, ECSSD, SED2, HKU, PASCAL-S and SOD data sets to prove the effectiveness of the new method by quantitative and qualitative analysis.

Section 2 introduces the related works. In section 3, we give the detailed proposed model analysis. Experiments are conducted in section 4. There is a conclusion in section 5.

2. Related Works

With the continuous progress of science and technology, how to quickly search and locate the information that people are interested in from the massive data resources has become an important research content in the field of computer vision [11]. Visual saliency has been regarded as an important mechanism for processing information tasks in computer vision. Saliency object detection obtains the areas of interest in images by simulating visual saliency and ignores the areas of uninterest. Saliency object detection is widely used in image matching, image retrieval, image compression, image quality assessment, object recognition and object relocation.

According to feature selection methods, saliency detection can be divided into two categories: artificial feature selection and deep convolutional network-based feature extraction.

2.1. Artificial Feature Selection

Traditional saliency object detection methods usually use manually selected features at the pixel level. Most of these algorithms make computation based on local or global features, such as color, orientation and texture. Global-based methods estimate the saliency of each pixel or region by using global contrast and feature statistics. Li et al. [12] proposed an automated saliency object segmentation method based on context and shape prior, which obtained saliency images by continuously iterating update of multi-scale context information and shape prior. Shen et al. [13] proposed a saliency object detection method based on low-rank matrix recovery, and obtained saliency map by fusing low-level features and high-level prior information. Yang et al. [14] proposed a graph regularization saliency detection method based on convex hull center prior, and refined the primary saliency images calculated by contrast and center prior with the graph regularization method to obtain the final saliency images. Xie et al. [15] proposed a Bayesian saliency detection method based on low and medium level cues, it used Bayesian methods to fuse saliency information on medium and low level cues to obtain final saliency maps. Zhang et al. [16] proposed a new saliency detection method using prior information such as frequency, color and position to obtain saliency maps. Li et al. [17] proposed a saliency detection method based on background prior and foreground seed selection, using the fusion of background prior and foreground prior to obtain a saliency map. Piao et al. [18] proposed a saliency detection method based on cellular automata (Single-layer cellular automata method (SCA), Single-layer Cellular Automata Optimizing Background Map (BSCA), Multi-layer Cellular automata (MCA)). According to the updating criteria, the saliency value was repaired by the cellular automata mechanism to obtain the saliency map. Zhou et al. [19] proposed a detection method based on integral fusion compactness and local contrast, using the complementary properties of compactness and local contrast cues to obtain saliency images. Tang et al. [20] proposed a saliency object detection method based on weighted low-rank matrix recovery, using position, color and boundary connectivity to generate a high-level background prior map, which was integrated into the weighted matrix to obtain a saliency map. Although they are easy to implement, they lack geometric structure cues and semantic information, so contrast-based algorithms cannot uniformly detect complete saliency objects, nor can they effectively suppress messy backgrounds in complex images.

2.2. Deep Convolutional Network-based Feature Extraction

The traditional saliency object detection methods mainly rely on the low-level features of the image which are manually selected and cannot describe the deep semantic feature information. Therefore, saliency objects cannot be accurately detected in complex images. At present, deep neural network technology is widely used in computer vision tasks, which greatly improves the performance of models. For the saliency object detection task, the data-driven model aims to obtain the semantic information of saliency

objects directly from a set of training data with pixel-level tags in the supervised learning model. CNN-based saliency detection methods can be divided into two categories, based on superpixel segmentation and based on FCN. The former uses the superpixel as the basic unit to train the deep neural network to predict saliency. All the pixels located in the same superpixel enjoy the same saliency value in the final prediction map. Wang et al. [21] used CNN to calculate the saliency score for each pixel in the local context, and then fine-tuned the saliency score for each object region in the global content. Li et al. [22] predicted the saliency score of each superpixel by combining local context and global context simultaneously in multi-context CNN. Chen et al. [23] used global and local context information to integrate them into the backbone network based on deep convolutional network for saliency detection. However, these superpixel-based methods tend to deal with local regions alone and cannot effectively capture global information about saliency objects. In addition, they rely on the over-segmentation method, so the network must run many times to calculate the saliency value of all the superpixels in the image, which makes the algorithm very time-consuming. Finally, they ignore the important spatial context information because they simply assign saliency values to each superpixel. In practice, the context information of an image is very useful for saliency detection.

To overcome these shortcomings, researchers tend to use the FCN model to detect saliency targets in a pixel-to-pixel manner.

Wang et al. [24] used end-to-end convolutional neural networks to compute visual contrast information within an image. Liu et al. [25] designed a two-step deep network to obtain rough global predictions by autonomously learning globally saliency cues, and then used another network to further fine-tune the details of the prediction map by integrating local context information. On this basis, Li et al. [26] proposed saliency detection network with sharing features, and utilized a Traplus-regularized nonlinear regression model for saliency adjustment. Although these deep learning-based methods have made significant progress, the CNN-based model still can be improved, so that it can uniformly highlight the whole saliency object and retain accurate boundary information in complex images with messy backgrounds.

3. Proposed Saliency Object Model

The proposed saliency object detection algorithm in this paper mainly includes two steps: 1) multi-scale fully convolutional network. The weight values of the network are updated by the modified frog leaping algorithm, and the rich features from each convolution layers are extracted and fused. 2) Saliency update method. A fully connected conditional random field (CRF) model is used to update the prediction results to produce more refined saliency detection results.

3.1. Multi-scale Full Convolutional Network

In order to design a FCN network that can learn the pixel-to-pixel saliency detection task, a multi-scale deep convolutional neural network is proposed to extract more multi-scale and multi-level feature information that is beneficial to saliency detection. The proposed neural network model is shown in figure 2 with three parts: weight update module, multi-scale feature extraction module and feature fusion module.

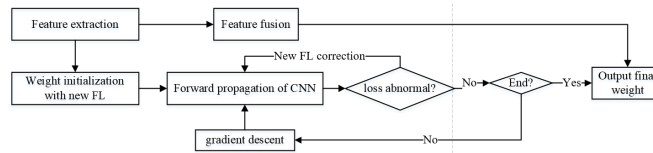


Fig. 2. The proposed saliency detection model

Weight update by IFL in multi-scale deep convolutional neural network Aiming at the complexity of saliency images, the improved frog leaping (IFL) algorithm is introduced into the weight initialization and weight update in convolutional neural networks [27]. By constantly updating the position of the worst frog, the global optimal frog is found as the initial weight of the network. In the process of network iteration, the calculated loss value each time is collected, and the network weight which produces abnormal loss value is corrected by the IFL.

When the generated weight by the network is too poor, the network needs to spend more time to perform gradient descent operation to correct the weight. So it is easy to make the network fall into local optimum and affect the final result. To solve this problem, the IFL is proposed to initialize the weight of convolutional neural network.

The traditional FL algorithm is a kind of sub-heuristic population evolutionary algorithm, which has excellent global search ability and can effectively calculate the global optimal solution. However, if we directly apply it to the weight initialization in this paper, it will generate a large number of calculations, and can increase the time cost, affect the efficiency. Therefore, an improved FL algorithm is proposed in this study. Different from the traditional FL, we no longer divide the initial samples into several populations, but regard all frogs as a population and directly conduct global optimization. And some improvements in the FL rule are made. The details of weight initialization algorithm and weight updating algorithm are shown in **Algorithm 1** and **Algorithm 2**.

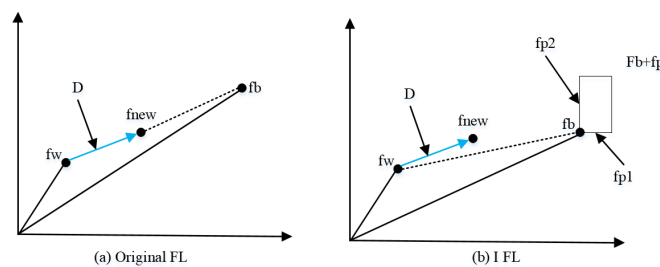


Fig. 3. Improved frog leaping algorithm

The back propagation of traditional CNN is essentially the gradient descent method, which updates network parameters through the loss value calculated by forward propagation, so it can find the optimal solution. However, the saliency image background is complex and the targets overlap each other, so it is difficult to detect the saliency of the tar-

Algorithm 1 Weight initialization algorithm

- 1: Step 1. Parameter initialization. Determine the frog population according to the Gaussian distribution formula:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right). \tag{1}$$

Where, we set $\mu = 0, \sigma = 1$.

- 2: Step 2. Ranking. All the frogs whose loss values are not calculated are brought into the CNN model. n images are randomly selected from the training image library as the reference images for forward propagation, and the loss value of each frog is calculated. Here, the loss value calculation function is the fitness function of the FL algorithm, and the loss calculation formula is:

$$loss = \sum_{k=1}^b \left(-\sum_{i=1}^n \sum_{j=1}^s t_{ij} \ln(p_{ij})\right). \tag{2}$$

Where p represents the output value of the network. t represents the real value. s represents the dimension of the saliency object label, and b represents the number of saliency targets to be detected at the same time. It ranks all frogs in ascending order according to their fitness functions.

- 3: Step 3. Searching and updating the location. The optimal frog f_b and the worst frog f_w can be obtained from Step 2. The position of the worst frog is updated by the position update function. For frog position updating, this study adds an offset to the traditional FL formula, and appropriately increases the random interval of $rand()$ function. The formula is as follows:

$$D = (f_b + f_p - f_w) \times rand(0, 1, 2). \tag{3}$$

$$f_{new} = f_w + D. \tag{4}$$

$$f_{p_i} = \frac{rand(-1, 1) \times 0.0001}{\exp(f_b + f_a) + 1}. \tag{5}$$

Here, f_p represents the offset, whose dimension is the same as that of each frog. f_{p_i} represents the value on the i -th dimension of f_p . f_{new} represents the updated frog. By adding offset, the performance of FL can be improved effectively. Increasing the random interval makes it easy to find the optimal solution. The leaping way mentioned in this step is mapped to two-dimensional coordinate as shown in figure 3:

- 4: Step 4. Judging whether the algorithm meets the convergence condition. If YES, stop the algorithm and take the optimal frog value as the initial weight of the convolutional neural network. Otherwise, return to Step 3.
-

get. When gradient descent algorithm is used, it is easy to occur abnormal situation with a large range of loss value, which affects the efficiency and even causes the algorithm falling into the local optimum. In view of the above problems, the improved frog leaping algorithm is used to correct the poor gradient of the back propagation of convolutional neural network.

In the process of network iteration, the loss value in each forward propagation is calculated. If the absolute value of the difference between the loss value in the i time and $i - 1$ time is greater than the threshold o , that is, $|l_i - l_{i-1}| > o$, it is considered that the weight in the i -th forward propagation is invalid. The IFL algorithm is used to re-find the optimal weight. After the convolutional neural network satisfies the end conditions, the trained weight FA is obtained, and the final algorithm performs the improved FL calculation again. The local optimization can be directly and effectively avoided by the last improved FL optimization.

Feature extraction module The multi-scale feature extraction module outputs feature maps with different resolutions from the sides of different convolution groups of the backbone network. The proposed model uses VGGNet-16 (Visual Geometry Group), which has been pre-trained for image classification in ImageNet data set as the backbone network. And it is modified to meet the requirements. It retains its 13 convolution layers and removes the fifth pooling layer and the fully connection layer. The modified VGGNet consists of five groups of convolution layers. For brevity, the third sub-layer in the fifth group of convolution layers is represented as conv5_3, and other convolution layers in VGGNet are also represented by this method. For the input image 256×256 pixels, the modified VGGNet-16 produces five feature maps $f_1^a, f_2^a, \dots, f_5^a$, the spatial resolution decreases according to stride 2. These feature maps are generated from (Conv1_2, Conv2_2, \dots , Conv5_3) respectively. The feature map from Conv1_2 has the maximum spatial resolution, while the feature map f_5^a from Conv5_3 has the minimum spatial resolution.

Feature fusion module Different convolution layers usually produce different feature representations, ranging from low-level structural features to high-level semantic features. The shallow convolution layer contains rich details while the deep convolution layer contains rich semantic information but lacks spatial context information. Feature fusion module involves multi-scale convolution feature fusion.

For each feature graph $f_i^a (i \in 1, 2, 3, 4, 5)$, the feature graph f_i^b is obtained by a 3×3 convolution layer and a 5×5 convolution layer. It sets its channel number as the channel number output by the i -th side of VGGNet-16. Then, by using a single-channel 1×1 convolution layer for dimension reduction, five feature graphs $f_i^c (i \in 1, 2, 3, 4, 5)$ with sizes of 256×256 pixels, 128×128 pixels, 64×64 pixels, 32×32 pixels and 16×16 pixels are obtained. In order to make these feature maps f_i^c have the same size as the input images, deconvolution operation and bilinear interpolation are used to up-sample the feature maps. The steps of the deconvolution layer in the five output layers are set to 1, 2, 4, 8 and 16 respectively. Then, these feature maps f_i^d with the same resolution are spliced together. Finally, a saliency prediction map S is generated through a 1×1 convolution layer. In the training stage, stochastic Gradient Descent (SGD) method is used to minimize all training samples.

Algorithm 2 Weight update algorithm

- 1: Step 1. Input a sample. Inputting the target image and the corresponding label, carry out one-hot coding for the label.
- 2: Step 2. Convolution, pooling. Multi-layer convolution and pooling operations are carried out on the input target image. After each pooling calculation, ReLU function is used to activate the output results.
- 3: Step 3. Fully connection calculation. The predicted value is obtained through multi-layer fully connection calculation. The predicted value is substituted into the Softmax() function for calculation. After obtaining the result, the loss value is calculated by formula (2).
- 4: Step 4. Error value comparison. Judging whether the absolute value difference between the loss value of this iteration and the previous loss value is greater than the threshold value. If YES, return to Step 5. Otherwise, go to Step 10.
- 5: Step 5. Parameter initialization. Recording the weight w_b of the $(i - 1) - th$ forward propagation. Setting the frog number c .
- 6: Step 6. Generating frog. Different from the way of generating frogs by Gaussian distribution mentioned above, here the frogs are mainly generated based on , and the generation formula is as follows:

$$w_{ij} = w_{bj} + 0.01 \times rand(-1, 1). \quad (6)$$

Where, $1 \leq i \leq c - 1$, w_{ij} represents the value in the $j - th$ dimension of the generated $i - th$ frog. n is the total number of frogs. In the frog colony, there is a frog w_g , which meets the conditions: $l_g \leq l_b$. l_g and l_b are the loss values of w_g and w_b , respectively.

- 7: Step 7. Ranking. A small number of training images are randomly selected as input values, and all frogs in c are brought into the CNN model. The loss value is calculated according to formula (2). All frogs are sorted in descending order by loss value.
 - 8: Step 8. Searching and location update. This step is the same as step 3 in Algorithm 1.
 - 9: Step 9. Checking the frog leaping stop conditions. Whether the algorithm meets the convergence condition. If so, stop the algorithm and update the network weight with the value of the optimal frog; otherwise, return to Step 8.
 - 10: Step 10. Check the network stop conditions. Whether the network meets the convergence condition. If Yes, stop the iteration; otherwise, go to Step 1.
 - 11: Step 11. Optimizing the final weights. After trained the algorithm, the weights in the network are the final weights, and the final weights are taken as the initial frog w_b . The frog swarm is generated by formula (6), and then steps 7, 8 and 9 are successively executed to finally obtain the global optimal frog w_{qb} , which is the final trained weight value of the algorithm.
-

In order to simulate the spatial correlation of the whole image and reduce the amount of computation, a network structure based on fully convolution is used. The fully convolution operation has the ability to share the convolution features through the whole image, thus reducing feature redundancy and making the full convolution network model simple and effective.

3.2. Optimization of Spatial Continuity

The proposed saliency detection algorithm using multi-level features can accurately locate the saliency objects and effectively suppress the complex background. However, the saliency prediction map of the proposed multi-scale fully convolutional network is relatively rough, and the contour information of saliency objects is not well retained. In order to improve the spatial continuity of the detection results, the proposed algorithm additionally uses the fully connected CRF method [28] to update the saliency map of the network pixel by pixel during the test phase. This method solves the problem of binary pixel label allocation and uses the following energy function for calculation, i.e.

$$E(L) = - \sum_i \log P(l_i) + \sum_{i,j} \theta_{ij}(l_i, l_j). \quad (7)$$

Where, i and j represent the horizontal and vertical coordinates of pixels in the image respectively. L represents the binary labels of all pixels. $P(l_i)$ is the probability of pixel x_i with label l_i , which indicates the possibility that pixel x_i belongs to the saliency object.

Initially, $P(1) = S_i$, $P(0) = 1 - S_i$, where S_i is the saliency score at pixel x_i of the fused saliency graph S , that is, the binary potential function $\theta_{ij}(l_i, l_j)$ is defined as:

$$\theta_{ij} = \mu(l_i, l_j) \left[\omega_1 \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_\alpha^2}\right) - \left(-\frac{\|I_i - I_j\|^2}{2\sigma_\beta^2}\right) + \omega_2 \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_\gamma^2}\right) \right]. \quad (8)$$

In the formula, p_i and I_i represent the position of pixel x_i and pixel value respectively. ω_1 , ω_2 , σ_α , σ_β and σ_γ are the weights. If $l_i \neq l_j$, then $\mu(l_i, l_j) = 1$. θ_{ij} contains two convolution kernels. The first convolution kernel depends on pixel position p and pixel intensity I . This convolution kernel enables adjacent pixels with similar colors to have similar saliency fractions. The second convolution kernel is used to remove small isolated regions.

As shown in figure 4, the fused saliency graph of the multi-scale fully convolutional network without CRF is rough and cannot uniformly display the internal region of the saliency object, while the saliency graph updated by CRF well retains the contour of the saliency object and uniformly highlights the whole saliency object.

4. Experiments and Analysis

4.1. Data Set

In order to evaluate the performance of the proposed algorithm, a series of subjective and objective experiments are carried out on six benchmark data sets. These datasets have

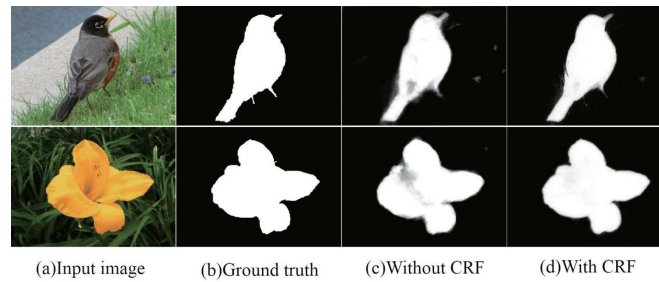


Fig. 4. Comparison of saliency detection results with and without CRF

pixel-level labels, including DUT-OMRON, ECSSD (Extended Complex Scene Saliency Dataset), SED2, HKU, PASCAL-S and SOD (Saliency objects dataset) [29]. The HKU is a large data set with over 4000 challenging images, most of which have low contrast and multiple saliency objects. The DUT-OMRON includes 5168 images with one or more saliency objects and relatively complex backgrounds. ECSSD contains 1000 semantically meaningful and complex images. PASCAL-S contains 850 real-world images selected from a PASCAL-VOC data set with 20 object classes. SED2 is a multi-object data set that typically contains two saliency objects per image. SOD consists of 850 images containing one or more objects with a cluttered background. In contrast, the HKU, PASCAL-S and SOD datasets are more challenging due to the presence of multiple saliency objects in their images and the complex background.

4.2. Evaluation Index

In this section, five commonly evaluation indicators are used to measure the performance of the proposed algorithm, including precision recall (PR) curve, F-measure, Max F-measure (maxF), weighted F-measure (wF) and mean absolute error (MAE) [30].

PR curve: Precision refers to the percentage of positive samples in all data predicted to be positive samples. Recall rate refers to the proportion of the data predicted as true samples to all positive samples. The saliency feature map is segmented with a fixed threshold value ranging from 0 to 255. A pair of accuracy-recall fractions are calculated to form PR curves to describe the performance of the algorithm under different conditions.

F-measure and maximum F-measure: F-measure is a comprehensive quantitative index of PR, which is calculated as,

$$F_{\beta} = \frac{(1 + \beta^2) \cdot P \cdot R}{\beta^2 \cdot P + R}. \quad (9)$$

Where, β is the balance parameter. P is precision, and R is the recall rate. In this paper, β^2 is set to 0.3 to improve the proportion of important precision. The threshold is set to twice the average saliency of the entire image. The $maxF$ is defined as the maximum F-measure calculated using the PR curve.

Weighted F measure (wF). This index is a weighted version of the F measure, which corrects the interpolation, dependence, and equal-importance defects of the F measure.

Similar to the F measure, the weighted F measure is calculated by the weighted harmonic average of the weighted precision P^w and the weighted recall rate R^w , i.e.,

$$F^w = \frac{(1 + \beta^2) \cdot P^w \cdot R^w}{\beta^2 \cdot P^w + R^w}. \quad (10)$$

Mean Absolute Error (MAE) is used to measure the mean error, which is defined as the absolute error of the average pixel between the truth graph and the predicted saliency graph.

$$M = \frac{1}{h \times w} \sum_{i=1}^h \sum_{j=1}^w |S_{ij} - G_{ij}|. \quad (11)$$

Where S represents saliency graph, G represents truth graph. h and w represent the height and width of the image.

4.3. Implementation Details

The experiment environment in this paper is: Windows7 64-bit operating system, Intel(R) Core(TM) i5-4210U, CPU 1.7GHz processor, Memory 4 GB, 32 RAM, MATLAB R2017a, NVIDIA 1060T GPU.

The proposed network is implemented on the an open source framework (Caffe). More specifically, the pre-trained VGGNet-16 network model is used and modified in the feature extraction module, and the parameters of the convolution layer are randomly initialized in the feature fusion module. The entire network is fine-tuned on the MSRA-B dataset to achieve the task of pixel-to-pixel saliency detection. MSRA-B is a public data set containing 5000 test images. The resolution of all test images and truth images is adjusted to 256×256 pixels for training, and only one image is loaded each time. The learning rate is set to 10^{-9} . Weight attenuation is 0.0005. The momentum is 0.9. Loss weight of each side output is 1. In addition, the weights of fusion layer are all initialized to 0.2 in the training stage. The stochastic gradient descent is adopted for network learning.

The CRF parameters in this paper are determined by using the cross validation method on the verification data set ECSSD. In the experiments, $\omega_1 = 3.0$, $\omega_2 = 1.0$, $\sigma_\alpha = 8.0$, $\sigma_\beta = 60.0$, $\sigma_\gamma = 5.0$.

4.4. Comparison of Performance

The proposed algorithm in this paper is compared with 14 saliency object detection methods, including RFCN (Recurrent Fully convolutional network) [31], PAGR (Progressive Attention Guided Recurrent Network) [32], UCF (Uncertain convolutional Features) [33], SF (Supervision by Fusion) [34], DCL (Deep Contrast Learning) [35], MC (Multi-context Deep Learning) [36], MTDS (Multi-Task Deep Neural Network) [26], ELD (Encoded Low Level Distance Map and High Level Features) [37], LEGS (Local Estimation and Global Search) [21], MDF (Multi-scale deep CNN Features) [38], KSR (Kernelized Subspace Ranking) [39], DRFI (Discriminative Regional Feature Integration) [40], SMD (Structured Matrix Decomposition) [41] and RR (Regularized Random Walks ranking) [42]. For equitable comparison, the saliency results of compared model are provided by the authors.

Here, RFCN, UCF, MTDS, ELD, DCL, SF, MC, LEGS, MDF and KSR are based on deep learning.

A. Subjective comparative analysis

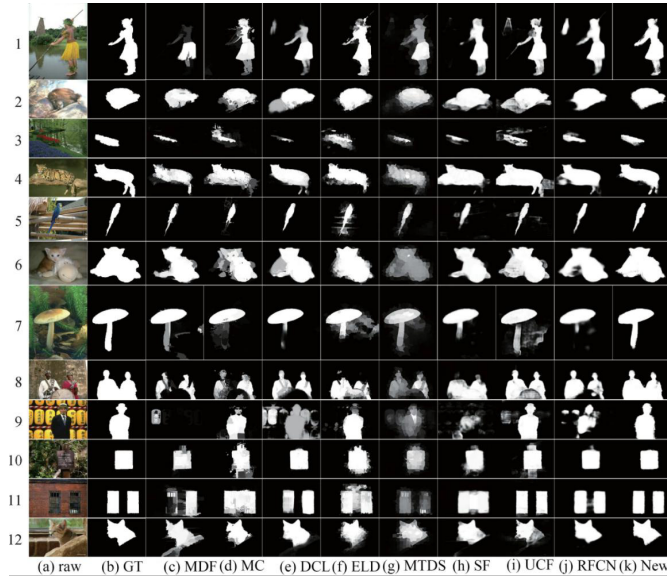


Fig. 5. Visual comparison results with different models on ECSSD dataset

Figure 5 shows a visual comparison of saliency images generated by different algorithms on six data sets. Experimental results show that the proposed method can deal with all kinds of complex images better, it not only can display the whole saliency object evenly, but also can retain the contour of saliency object in various scenes. For example, foreground objects and background low contrast (2 row, 4 row and 12 row in figure 5), the image boundary of saliency objects (88 row and 97 row), several saliency objects (6 row and 8 row), with complex texture and structure of the saliency objects (1 row, 3 row, 8 row and 11 row), and background clutter (5 row, 6 row, 7 row, and 10 row), etc.

B. Objective comparative analysis

For quantitative evaluation, figure 6 shows the PR curves of the proposed method and 14 representative algorithms on six benchmark data sets. It can be seen that: 1) the saliency object detection method based on FCN is superior to other methods; 2) The new method in this paper is competitive in ECSSD, DUT-OMRON, HKU, PASCAL-S and SOD data sets, but slightly inferior to MTDS, DCL, UCF and PAGR algorithms.

In addition, the F and wF scores of the proposed model are compared with these existing methods on six benchmark data sets. The results are shown in figures 7 and 8. The results of MAE and the maxF are shown in table 1.

It can be seen that the performance of the proposed method is worse than that of UCF on SED2, because most of the images in SED2 contain two separate small-size objects, while the new method does not introduce corresponding modules to deal with this situa-

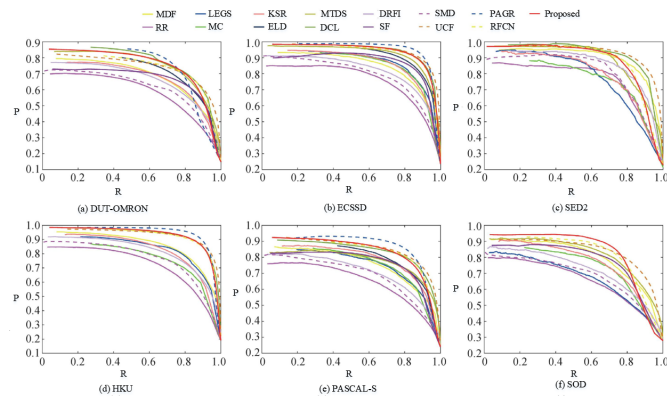


Fig. 6. PR curves of saliency maps produced by different approaches on six datasets

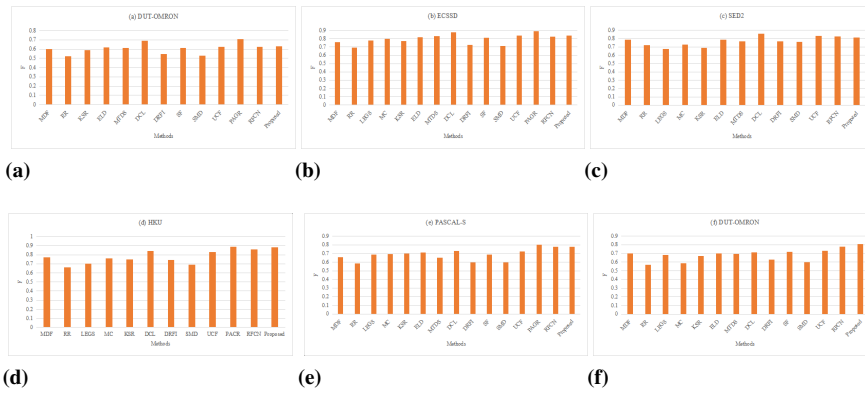


Fig. 7. F-measure scores of saliency maps with different approaches on six datasets

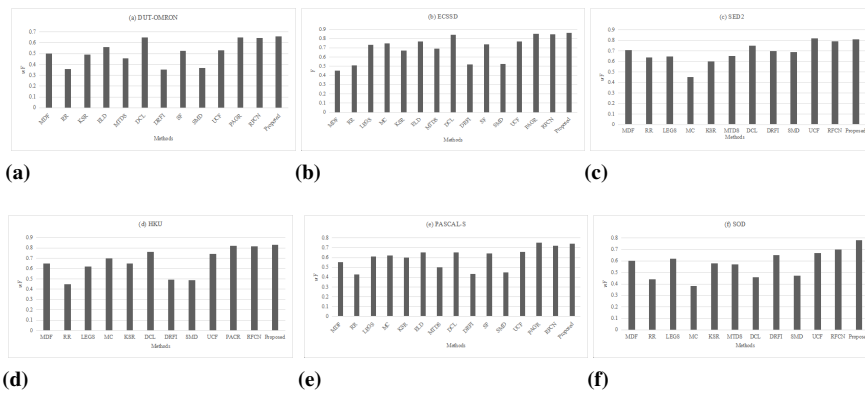


Fig. 8. Weighted F-measure scores of saliency maps with different approaches on six datasets

Table 1. MAE and maxF scores of saliency maps with different approaches on six datasets

Data	DUT-OMRON	DUT-OMRON	ECSSD	ECSSD	HKU	HKU
Index	MAE	maxF	MAE	maxF	MAE	maxF
RR	0.195	0.628	0.194	0.755	0.183	0.723
SMD	0.177	0.635	0.184	0.770	0.166	0.753
DRFI	0.160	0.674	0.181	0.793	0.155	0.788
LEGS	0.132	0.637	0.129	0.838	0.141	0.758
MDF	0.102	0.705	0.116	0.842	0.123	0.871
MC	0.114	0.805	0.111	0.847	0.102	0.818
DCL	0.090	0.767	0.078	0.911	0.074	0.901
ELD	0.101	0.716	0.089	0.878	0.093	0.852
MTDS	0.131	0.756	0.132	0.893	0.104	0.893
KSR	0.141	0.689	0.143	0.840	0.131	0.803
SF	0.118	0.695	0.098	0.863	0.099	0.876
UCF	0.131	0.740	0.080	0.914	0.073	0.898
PAGR	0.082	0.782	0.072	0.938	0.059	0.929
RFCN	0.088	0.744	0.076	0.793	0.065	0.894
Proposed	0.073	0.758	0.061	0.897	0.049	0.887
Data	SED2	SED2	PASCAL-S	PASCAL-S	SOD	SOD
Index	MAE	maxF	MAE	maxF	MAE	maxF
RR	0.184	0.795	0.237	0.663	0.270	0.657
SMD	0.169	0.831	0.217	0.699	0.245	0.690
DRFI	0.156	0.846	0.218	0.704	0.235	0.714
LEGS	0.152	0.763	0.170	0.756	0.206	0.747
MDF	0.134	0.670	0.153	0.769	0.235	0.713
MC	0.126	0.818	0.153	0.754	0.272	0.670
DCL	0.115	0.887	0.125	0.822	0.233	0.698
ELD	0.129	0.796	0.131	0.789	0.229	0.704
MTDS	0.145	0.884	0.186	0.773	0.201	0.795
KSR	0.163	0.792	0.165	0.778	0.208	0.755
SF	0.125	0.804	0.141	0.770	0.167	0.780
UCF	0.085	0.898	0.126	0.827	0.158	0.817
PAGR	0.096	0.813	0.101	0.870	0.156	0.801
RFCN	0.119	0.876	0.141	0.827	0.155	0.819
Proposed	0.091	0.857	0.098	0.823	0.123	0.834

tion. The new method is slightly worse than PAGR algorithm on six data sets. The possible reasons are as follows: 1) PAGR algorithm uses VGG-19 model with higher accuracy as the backbone network; 2) The PAGR algorithm introduces the recurrent network module, which can better detect the saliency from coarse to fine. Comprehensive analysis for the experimental results shows that the proposed method has certain advantages in processing complex scene images, and its performance is close to the true value.

4.5. Algorithm Analysis

A. Feature graph analysis

In the proposed multi-scale detection network, the proposed method extracts five feature graphs f_i^d from the modified VGGNet-16 and fuses them in the connection layer. On the ECSSD benchmark data set, each feature graph f_i^d is compared with the fused feature graph, and the results are shown in table 2. It can be seen from table 2 that: 1) The performance of the feature graph f_5^d from the deepest convolution layer edge is closer to that of the fused feature graph; 2) The saliency feature map obtained by combining multi-layer features is better than the single feature map.

Table 2. Comparison of feature maps from different side output

Feature map	maxF	F	wF	MAE
f_1^d	0.345	0.097	0.223	0.491
f_2^d	0.317	0.174	0.172	0.319
f_3^d	0.501	0.469	0.298	0.318
f_4^d	0.427	0.302	0.235	0.243
f_5^d	0.871	0.835	0.724	0.105

B. Feature map fusion analysis

In order to verify the effectiveness of the proposed combination scheme, these features are combined in different ways and expressed as: $S_1 = f_5^d$, $S_2 = \sum_{i=4}^5 f_i^d$, $S_3 = \sum_{i=3}^5 f_i^d$, $S_4 = \sum_{i=2}^5 f_i^d$. The training set and the used hyperparameters are consistent with the model in this paper. The evaluation results on the ECSSD data set are shown in table 3. It can be seen that the proposed method in this article obtains a better performance.

Table 3. Comparison with different fusion ways

Method	maxF	F	wF	MAE
S1	0.876	0.814	0.758	0.081
S2	0.862	0.797	0.743	0.088
S3	0.855	0.795	0.724	0.093
S4	0.879	0.812	0.758	0.079
Proposed	0.877	0.834	0.779	0.077

C. Effect of IFL on the detection result

IFL is used for searching the optimization weight in the network. So we test the effect of IFL on the proposed method, the results are shown in table 4. The results show that the IFL can greatly improve the saliency detection.

Table 4. Effect of IFL on the detection result

Data set	Method	maxF	F	wF	MAE
DUT-OMPON	Without IFL	0.763	0.802	0.711	0.082
DUT-OMPON	With IFL	0.879	0.851	0.792	0.071
HKU	Without IFL	0.792	0.784	0.697	0.079
HKU	With IFL	0.881	0.845	0.785	0.068

D. Effect of CRF on the detection result

As a post-processing step, the CRF method updates the saliency map obtained from the network to further highlight the consistency of the region inside the saliency object and retain the accurate contour information of the saliency object.

To verify its effectiveness, maxF, F, wF and MAE scores are used to evaluate the performance of the saliency method with/without CRF, and the results are shown in table 5. It can be seen from table 5 that the accuracy of the proposed model can be further improved by using the CRF method in the test phase. We extend this work by replacing VGGNet-16 with ResNet-101. The conv1, res2c, res3b3, res4b22 and res5c of ResNet-101 are used as the side output, and other settings are kept unchanged. In table 5, the VGG network in proposed method as the backbone network is denoted as V and ResNet-101 as the backbone network is denoted as R. As can be seen from table 5, with the same training sets, the saliency graph generated by ResNet-101 without CRF improves the performance of the algorithm by 2% averagely, which indicates that the overall performance of the algorithm can be further improved by using the backbone network with better performance.

Note: V/R+CRF=VGGNet/ResNet with CRF, V/R-CRF=VGGNet/ResNet without CRF

5. Conclusions

The existing saliency object detection methods are difficult to highlight the clear boundary of the whole object and uniformly highlight the entire internal region in complex images. Therefore, we propose a new frog leaping algorithm-oriented fully convolutional neural network for motion object saliency detection. It extracts multi-scale and multi-level features from different convolution layers in the VGG network. Meanwhile, an improved frog leaping algorithm is used to optimize the selection of initial weights during network initialization. The shallow convolution layer produces detailed information, while the deep convolution layer produces global information. Then, the connection layer is used to combine these rich image saliency features to generate a saliency map. In the test phase, in order to further obtain the saliency detection results with accurate contour and uniform internal region, it introduces the fully connected CRF for saliency update. The

Table 5. Comparisons with/without CRF

Data set	Method	maxF	F	wF	MAE
DUT-OMPON	V+CRF	0.758	0.707	0.667	0.085
DUT-OMPON	V-CRF	0.728	0.645	0.573	0.101
DUT-OMPON	R+CRF	0.797	0.750	0.722	0.072
DUT-OMPON	R-CRF	0.769	0.682	0.612	0.087
ECSSD	V+CRF	0.899	0.887	0.865	0.072
ECSSD	V-CRF	0.887	0.845	0.789	0.087
ECSSD	R+CRF	0.906	0.859	0.797	0.081
ECSSD	R-CRF	0.926	0.906	0.889	0.064
SED2	V+CRF	0.870	0.808	0.783	0.104
SED2	V-CRF	0.837	0.790	0.727	0.113
SED2	R+CRF	0.868	0.813	0.792	0.104
SED2	R-CRF	0.831	0.784	0.710	0.115
HKU	V+CRF	0.901	0.879	0.855	0.060
HKU	V-CRF	0.872	0.822	0.760	0.076
HKU	R+CRF	0.912	0.893	0.874	0.057
HKU	R-CRF	0.883	0.829	0.759	0.076
PASCAL-S	V+CRF	0.825	0.783	0.743	0.099
PASCAL-S	V-CRF	0.809	0.747	0.670	0.124
PASCAL-S	R+CRF	0.835	0.795	0.759	0.096
PASCAL-S	R-CRF	0.819	0.757	0.667	0.122
SOD V+CRF		0.844	0.796	0.773	0.135
SOD V-CRF		0.826	0.770	0.706	0.826
SOD R+CRF		0.849	0.798	0.784	0.132
SOD R-CRF		0.830	0.796	0.693	0.146

experimental results show that the proposed method performs better than the 14 representative methods in terms of five performance evaluation indexes on six public available benchmark data sets. For subjective vision, the saliency image obtained by the proposed method can better deal with various complex images, it not only can display the whole saliency object uniformly, but also can well retain the contour of saliency object in various scenes.

In future research, the recurrent network module and boundary detection module will be used to improve the detection performance on small object images. And we will apply them to practical engineering.

Acknowledgments. This work was supported by: 1) Jiangsu Art Fund: "The Stage Art Funding Project of "Cao. Ji"" (No.: 2020-14-002); 2) National Educational Information Technology Research Project Youth Project: Traditional Dance Movement Collection Based on Motion Capture and Construction of Teaching Resource Database (No. 186140095).

Availability of data and materials. The data used to support the findings of this study are available from the corresponding author upon request.

Competing interests. The authors declare that they have no conflicts of interest.

References

1. Song H, Deng B, Pound M, et al. "A fusion spatial attention approach for few-shot learning," *Information Fusion*, vol. 81, pp. 187-202, 2022.
2. C. Guo and L. Zhang. "A Novel Multiresolution Spatiotemporal Saliency Detection Model and Its Applications in Image and Video Compression," *2013 IEEE Transactions on Image Processing*, vol. 19, no. 1, pp. 185-198, Jan. 2010, doi: 10.1109/TIP.2009.2030969.
3. Radojičić, D., Radojičić, N., Kredatus, S. "A multicriteria optimization approach for the stock market feature selection," *Computer Science and Information Systems*, Vol. 18, No. 3, pp. 749-769, 2021. <https://doi.org/doi.org/10.2298/CSIS200326044R>
4. Chen Y, Yang X, Zhong B, et al. "CNNTracker: Online discriminative object tracking via deep convolutional neural network," *Applied Soft Computing*, vol. 38, pp. 1088-1098, 2016.
5. Guo, Z., Han, D., Li, K. "Double-Layer Affective Visual Question Answering Network," *Computer Science and Information Systems*, Vol. 18, No. 1, pp. 155-168, 2021. <https://doi.org/10.2298/CSIS200515038G>
6. Li, H., Han, D. "Multimodal Encoders and Decoders with Gate Attention for Visual Question Answering," *Computer Science and Information Systems*, Vol. 18, No. 3, pp. 1023-1040, 2021. <https://doi.org/10.2298/CSIS201120032L>
7. N. Tong, H. Lu, L. Zhang and X. Ruan. "Saliency Detection with Multi-Scale Superpixels," *IEEE Signal Processing Letters*, vol. 21, no. 9, pp. 1035-1039, 2014. doi: 10.1109/LSP.2014.2323407.
8. Gao S. "A Two-channel Attention Mechanism-based MobileNetV2 And Bidirectional Long Short Memory Network For Multi-modal Dimension Dance Emotion Recognition," *Journal of Applied Science and Engineering*, 26(4): 455-464, 2022.
9. Lamsiyah S, Mahdaouy A E, Ouatik S, et al. "Unsupervised extractive multi-document summarization method based on transfer learning from BERT multi-task fine-tuning," *Journal of Information Science*, 2021:016555152199061.

10. L. Jing, Y. Chen and Y. Tian, "Coarse-to-Fine Semantic Segmentation From Image-Level Labels," *IEEE Transactions on Image Processing*, vol. 29, pp. 225-236, 2020. doi: 10.1109/TIP.2019.2926748.
11. Wang G, Wang Z, Jiang K, et al. "Silicone Mask Face Anti-spoofing Detection based on Visual Saliency and Facial Motion," *Neurocomputing*, vol. 458, pp. 416-427, 2021.
12. Zheng X, Chen W. "An Attention-based Bi-LSTM Method for Visual Object Classification via EEG," *Biomedical Signal Processing and Control*, vol. 63:102174, 2021.
13. X. Shen and Y. Wu. "A unified approach to salient object detection via low rank matrix recovery," *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 853-860, 2012. doi: 10.1109/CVPR.2012.6247758.
14. C. Yang, L. Zhang and H. Lu, "Graph-Regularized Saliency Detection With Convex-Hull-Based Center Prior," *IEEE Signal Processing Letters*, vol. 20, no. 7, pp. 637-640, July 2013. doi: 10.1109/LSP.2013.2260737.
15. Y. Xie, H. Lu and M. Yang, "Bayesian Saliency via Low and Mid Level Cues," *IEEE Transactions on Image Processing*, vol. 22, no. 5, pp. 1689-1698, May 2013. doi: 10.1109/TIP.2012.2216276.
16. L. Zhang, Z. Gu and H. Li, "SDSP: A novel saliency detection method by combining simple priors," *2013 IEEE International Conference on Image Processing*, pp. 171-175, 2013. doi: 10.1109/ICIP.2013.6738036.
17. Li L, Zhou F, Zheng Y, et al. "Saliency detection based on foreground appearance and background-prior," *Neurocomputing*, vol. 301(AUG.2), pp. 46-61, 2018.
18. Y. Piao, X. Li, M. Zhang, J. Yu and H. Lu, "Saliency Detection via Depth-Induced Cellular Automata on Light Field," *IEEE Transactions on Image Processing*, vol. 29, pp. 1879-1889, 2020, doi: 10.1109/TIP.2019.2942434.
19. L. Zhou, Z. Yang, Q. Yuan, Z. Zhou and D. Hu, "Salient Region Detection via Integrating Diffusion-Based Compactness and Local Contrast," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3308-3320, Nov. 2015, doi: 10.1109/TIP.2015.2438546.
20. C. Tang, P. Wang, C. Zhang and W. Li, "Salient Object Detection via Weighted Low Rank Matrix Recovery," *IEEE Signal Processing Letters*, vol. 24, no. 4, pp. 490-494, April 2017, doi: 10.1109/LSP.2016.2620162.
21. L. Wang, H. Lu, X. Ruan and M. Yang, "Deep networks for saliency detection via local estimation and global search," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3183-3192, doi: 10.1109/CVPR.2015.7298938.
22. Guanbin Li and Y. Yu, "Visual saliency based on multiscale deep features," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 5455-5463, doi: 10.1109/CVPR.2015.7299184.
23. Chen H., Li Y., Su D. "RGB-D Saliency Detection by Multi-stream Late Fusion Network," *ICVS 2017. Lecture Notes in Computer Science*, vol. 10528, 2017. Springer, Cham.
24. S. Wang, R. Clark, H. Wen and N. Trigoni, "DeepVO: Towards end-to-end visual odometry with deep Recurrent Convolutional Neural Networks," *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 2043-2050, doi: 10.1109/ICRA.2017.7989236.
25. N. Liu and J. Han, "DHSNet: Deep Hierarchical Saliency Network for Salient Object Detection," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 678-686, doi: 10.1109/CVPR.2016.80.
26. X. Li et al., "DeepSaliency: Multi-Task Deep Neural Network Model for Salient Object Detection," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3919-3930, Aug. 2016, doi: 10.1109/TIP.2016.2579306.
27. Zou L. "An Intelligent Improvement Method Of Classroom Cognitive Efficiency Based On Multidimensional Interactive Devices," *Journal of Applied Science and Engineering*, 2022, 26(3): 445-454.

28. I. Batatia, "A Deep Learning Method with CRF for Instance Segmentation of Metal-Organic Frameworks in Scanning Electron Microscopy Images," *2020 28th European Signal Processing Conference (EUSIPCO)*, 2021, pp. 625-629, doi: 10.23919/Eusipco47968.2020.9287366.
29. Zhang Q, Zuo B C, Shi Y J and Dai M. "A multi-scale convolutional neural network for salient object detection," *Journal of Image and Graphics*, vol. 25, no. 06, pp. 116-129, 2020. doi: 10.11834/jig.190395.
30. S. Yin and H. Li. "Hot Region Selection Based on Selective Search and Modified Fuzzy C-Means in Remote Sensing Images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 5862-5871, 2020, doi: 10.1109/JSTARS.2020.3025582.
31. L. Wang, L. Wang, H. Lu, P. Zhang and X. Ruan, "Salient Object Detection with Recurrent Fully Convolutional Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 7, pp. 1734-1746, 1 July 2019, doi: 10.1109/TPAMI.2018.2846598.
32. X. Zhang, T. Wang, J. Qi, H. Lu and G. Wang, "Progressive Attention Guided Recurrent Network for Salient Object Detection," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 714-722, doi: 10.1109/CVPR.2018.00081.
33. P. Zhang, D. Wang, H. Lu, H. Wang and B. Yin, "Learning Uncertain Convolutional Features for Accurate Saliency Detection," *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 212-221, doi: 10.1109/ICCV.2017.32.
34. D. Zhang, J. Han and Y. Zhang, "Supervision by Fusion: Towards Unsupervised Learning of Deep Salient Object Detector," *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 4068-4076, doi: 10.1109/ICCV.2017.436.
35. G. Li and Y. Yu, "Deep Contrast Learning for Salient Object Detection," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 478-487, doi: 10.1109/CVPR.2016.58.
36. L. Huang, K. Song, J. Wang, M. Niu and Y. Yan, "Multi-graph Fusion and Learning for RGBT Image Saliency Detection," *IEEE Transactions on Circuits and Systems for Video Technology*, doi: 10.1109/TCSVT.2021.3069812.
37. G. Lee, Y. Tai and J. Kim, "Deep Saliency with Encoded Low Level Distance Map and High Level Features," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 660-668, doi: 10.1109/CVPR.2016.78.
38. J Su, Yi H, Ling L, et al. A surface roughness grade recognition model for milled workpieces based on deep transfer learning," *Measurement Science and Technology*, vol. 33, no. 4, 045014, 2022 (11pp).
39. L. Zhang, J. Sun, T. Wang, Y. Min and H. Lu, "Visual Saliency Detection via Kernelized Subspace Ranking With Active Learning," *IEEE Transactions on Image Processing*, vol. 29, pp. 2258-2270, 2020, doi: 10.1109/TIP.2019.2945679.
40. Wang J, Jiang H, Yuan Z, et al. Salient Object Detection: A Discriminative Regional Feature Integration Approach," *International Journal of Computer Vision*, vol. 123, pp. 251-268, 2017. <https://doi.org/10.1007/s11263-016-0977-3>
41. J. Li, Z. Wang and Z. Pan, "Double Structured Nuclear Norm-Based Matrix Decomposition for Saliency Detection," *IEEE Access*, vol. 8, pp. 159816-159827, 2020. doi: 10.1109/ACCESS.2020.3020966.
42. Y. Yuan, C. Li, J. Kim, W. Cai and D. D. Feng, "Reversion Correction and Regularized Random Walk Ranking for Saliency Detection," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1311-1322, March 2018, doi: 10.1109/TIP.2017.2762422.

Yin Lyu was born in Shenyang, Liaoning, in 1981 in China. He received a postgraduate master's degree from the Belarusian State University of Culture and Arts. Now, studying for a PhD in Dance at Taipei University of the Arts, Taiwan. Associate Professor of the

Dance Department of the Music College of Huaiyin Normal University. His research interests include: new media dance and digital dance research. His research interests include image processing, dance big data and cloud computing.

Chen Zhang was born in Harbin, P.R. China ,in 1965. Professor of Harbin sport university, master's tutor, dean of the College of Sports Art, concurrently vice chairman of the Heilongjiang Dancers Association, and chairman of the Heilongjiang Dance Sports Association. Main research direction: physical education training (sports dance); sports art.

Received: March 20, 2022; Accepted: September 01, 2022.

A Novel Art Gesture Recognition Model Based on Two Channel Region-Based Convolution Neural Network for Explainable Human-computer Interaction Understanding

Pingping Li¹ and Lu Zhao²

¹ School of Fine Arts, Zhengzhou Normal University
450000 Zhengzhou, China
910675024@qq.com

² School of Fine Arts, Yulin Normal University
537000 Yulin, China
zhaoluvip@163.com

Abstract. The application development of hot technology is both an opportunity and a challenge. The vision-based gesture recognition rate is low and real-time performance is poor, so various algorithms need to be studied to improve the accuracy and speed of recognition. In this paper, we propose a novel gesture recognition based on two channel region-based convolution neural network for explainable human-computer interaction understanding. The input gesture image is extracted through two mutually independent channels. The two channels have convolution kernel with different scales, which can extract the features of different scales in the input image, and then carry out feature fusion at the fully connection layer. Finally, it is classified by the softmax classifier. The two-channel convolutional neural network model is proposed to solve the problem of insufficient feature extraction by the convolution kernel. Experimental results of gesture recognition on public data sets NTU and VIVA show that the proposed algorithm can effectively avoid the over-fitting problem of training models, and has higher recognition accuracy and stronger robustness than traditional algorithms.

Keywords: explainable human-computer interaction understanding, two channel region-based convolution neural network, gesture recognition, softmax classifier, feature fusion.

1. Introduction

The application potential of hot technology in the field of explainable human-computer interaction (EHCI) has begun to show such as geo-spatial tracking technology on smartphones [1-3]. And the motion recognition technology is for wearable computers, stealth technology, immersive games, etc. Tactile interaction technology is for virtual reality, remote robotics, and telemedicine. Speech recognition technology is for call routing, home automation and voice dialing. Silent speech recognition is used for people with speech impairments and eye-tracking technology is used for advertising, websites, product catalogs, and magazine utility tests. The human-machine interface technology based on brainwave is used in the "mind wheelchair" developed for people with speech and mobility disorders [4,5].

Gestures are an integral part of interpersonal communication. Gesture recognition opens up new ways for humans to interact with machines, devices or computers. With the development of science and technology, gesture recognition technology has developed from the era of data gloves with the help of external auxiliary equipment to the stage of pattern classification based on computer vision.

The gesture is a natural and intuitive means of HCI. It has become a trend to use gestures as computer input. In recent years, gesture recognition has gradually become an important research direction in the field of computer vision, especially in the application prospect of human-robot Interaction (HRI) technology, which greatly promotes the research development of gesture recognition. Gesture recognition refers to the use of certain algorithms to make the computer recognize the gesture of the human body in the picture or lens, and then understand the meaning of the gesture, to achieve mutual communication between the user and the computer. In the process of human-computer interaction to make the computer accurately understand people's intentions, the gesture recognition algorithm must have a highly accurate recognition effect, excellent processing speed, and recognition ability under different light, Angle, background, and other complex environments [6-8].

Currently, the popular visual gesture recognition can be divided into three stages: segmentation, feature extraction and recognition. Gesture segmentation is the basis of gesture recognition. Due to the clustering characteristics of skin color in color space, the majority of gesture segmentation methods at present use the color features of skin color (YUV, HSV [9], YCbCr1 [10], etc.) or geometric features (such as elliptic model and graph model [11]). Traditional gesture recognition algorithms mainly include two categories: (1) gesture recognition based on hidden Markov model (HMM) [12], which can be used to express a Markov process with hidden unknown parameters, and gesture recognition process can be regarded as a Markov chain with time series. Therefore, this model is widely used in gesture recognition. (2) Gesture recognition based on set features [13]. This method uses gesture edge, contour, regional distribution and other features to recognize gestures. Both of these two gesture recognition methods require manual feature extraction, which is highly complex and requires high professional knowledge and experience of personnel. At the same time, it also has the problem of poor adaptability to unfamiliar scenes.

Feature extraction based on hand range is the key stage of gesture recognition. At present, influential studies are as follows. Lin et al. [14] extracted low-frequency coefficient features of fuzzy palmprint through Laplace smoothing transform and fused them with geometric features of hand to represent gestures. However, such feature extraction steps were complex and time-consuming. Asaari et al. [15] expressed gestures by integrating geometric features of hands with features such as palmlines, knuckle lines and veins of hands. Due to the complex background of the acquired images, the accuracy of gesture recognition based on texture features was low. Liu et al. [16] proposed a gesture recognition method that combined finger contour features with geometric features. Although it improved the robustness of gesture recognition, it required fingers to be separated from each other, which had certain limitations in practical application. Zhu et al. [17] proposed a two-level detection model, which could detect gestures, and the provided gesture border information could also be used for further gesture recognition analysis. However, the two stages of this method were trained separately, and the second stage required a large amount of label data.

Convolutional Neural Network (CNN) is one of the most widely used models in the field of machine vision and image processing, and has attracted great attention in industry and academia [18,19]. Convolutional neural networks can learn the local and global features of the input image through training, which solves the problem of insufficient feature extraction caused by manual feature extraction. With strong feature extraction and classification capabilities, CNN has been widely applied in many fields of pattern recognition, and has made remarkable achievements in image classification, face recognition, voice recognition and other fields.

In the field of image processing, the application direction of convolutional neural network is mainly image classification, target recognition, image segmentation and so on. In the field of gesture recognition, some scholars have tried. Nguyen et al. [20] combined maximum pooling with convolutional neural network for gesture recognition and obtained a recognition rate of 96.77%. Peng et al. [21] discussed the integration of gesture image pre-processing and gesture recognition process before the input of convolutional network, thus realizing end-to-end gesture recognition and improving the accuracy of recognition. Fang et al. [22] changed the first convolution layer of convolutional neural network into 3d convolution, so that dynamic gesture could be input into the model in the form of stereogram, which successfully solved the problem of regularization of input of dynamic gesture recognition. Singh et al. [23] creatively used the stereoconvolution kernel for gesture recognition and obtained a good gesture recognition accuracy.

Most of the existing works use pattern classification algorithms to realize gesture recognition. Rahman et al. [24] proposed an artificial neural network method for Bengali sign language static gesture letter recognition, but this method had high requirements on the number of samples and extracted data characteristic values. It also had low mean average precision (mAP) recognition accuracy. Panwar et al. [25] proposed a bit-coding sequence based on shape parameter features to achieve gesture classification, but this method had limitations for gesture direction placement and low robustness. Dominio et al. [26] proposed a gesture recognition method based on depth information with the help of Kinect equipment, but this method had high requirements on equipment, complex algorithm and low accuracy of experimental results. Yang et al. [27] proposed a gesture recognition method based on gesture main direction and Hausdorff-like distance template matching. The 2D Cartesian coordinate system was constructed to extract the gesture feature vector through the main direction of the gesture. However, this method had a high limitation on the main direction of the gesture. When the main direction of the gesture obtained was inconsistent with the main direction of similar gestures in the training library, it was prone to error recognition. In view of the complexity of current gesture feature extraction methods and the low gesture recognition rate in complex background, the research and practice of references [28,29] found that convolutional neural network (CNN) had scale invariance to flip, pan and scale. It was better than other machine vision methods in gesture detection and recognition applications. In references [30,31], R-CNN was used for target detection framework, which reached the world's leading level in face detection and pedestrian detection. Le et al. [32] proposed a multi-scale R-CNN method, which used the RPN network to detect a single object first, then extract geometric information. Finally, it determined whether a driver used a mobile phone or how many hands he had on the steering wheel. This method achieved the highest accuracy in the famous VIVA data

set. But when detecting the number of hands on the steering wheel, the detection rate was only 65%, which still had a large missed detection rate.

It can be seen from the above analysis that the current research work is to separate gesture detection and recognition. However, since CNN method can reach the world's leading level in gesture detection and recognition, can CNN be used to realize both detection and recognition? In this paper, a new gesture recognition method is proposed by extending R-CNN algorithm, which can detect and recognize gestures simultaneously. This paper uses the strategy of modifying key parameters to improve the existing R-CNN framework. In order to achieve higher recognition accuracy, a two-channel network is proposed to avoid the over-fitting problem of the training model. For NTU-Microsoft-Kinect-Hand Posture Dataset (NTU) and the Vision for Intelligent vehicles and Experiments on Applications (VIVA) gesture datasets, experiment results show that the proposed algorithm has higher accuracy and stronger robustness than traditional algorithms.

2. Convolutional neural network (CNN)

Convolutional neural network has a unique advantage in the field of image processing, its understanding of image is from local to global. Generally speaking, the pixels in the part of the image are closely related, while the parts that are far apart are weakly related. Convolutional neural network firstly perceives local features, and then integrates these local features at a high level to obtain the global features and topological structure of the image, and then judge the attributes and categories of the image [33,34]. Therefore, convolutional neural networks are highly invariant to shape translation, scale scaling, tilt or other forms of deformation.

Convolutional neural networks have two typical characteristics: one is that the two layers of neurons are locally connected rather than fully connected through the convolution kernel. Therefore, the convolution layer connected with the input image is a local link established for pixel blocks, rather than the traditional full link based on pixel points. Second, in the same layer, the weight parameters of the convolution kernel in each convolution layer are shared. These two features greatly reduce the number of parameters of the deep network, reduce the complexity of the model, and speed up the training speed of the model, so that the convolutional neural network has a great advantage in the image processing with pixel values as the processing unit. The main components of convolutional neural network include convolutional layer, pooling layer, activation function, full connection layer and classifier. Generally, the first layer directly connected to the input image is the convolution layer. This layer is responsible for connecting images directly. The input is transformed into a form that can be understood by the convolutional network through the processing of pixel values, and then propagated forward. The pooling layer and activation function are usually connected behind the convolutional layer, alternating with the convolutional layer.

2.1. Convolutional layer

The convolutional layer is the core component of convolutional neural network. Its main function is to extract local features of input through the fixed step movement of the convolutional kernel. The output of the convolution layer is the feature graph, and each element

in the feature graph is the output of a neuron. The input of this neuron connection is a local region of the previous layer's output feature map, also known as the local receptive field. The input in the sensory field is calculated by a set of synaptic weights and the output of the neuron is obtained by the activation function. By sharing this set of synaptic weights, also known as the convolution kernel, the number of parameters can be greatly reduced during feature graph generation. For a network, the size of the convolution kernel is fixed. However, the weight parameters of the convolution kernel, namely the convolution template, are obtained through the training of training samples.

The convolution kernel is the core of the convolution layer. Convolution kernel is a mapping relation of image features extracted by local receptive fields. The convolution kernel can also be viewed as an eigenmatrix. During the convolution operation, the convolution kernel moves in turn on the input, and the product accumulation operation is carried out between the convolution kernel and the elements at the corresponding positions on the receptive field to obtain the convolution value of the receptive field. After the moving, the eigenmatrix of the input is obtained, which is also called the eigengraph. A single convolution kernel can only extract a certain type of image features. Therefore, multi-convolutional kernels are generally used in practical convolutional neural networks. The mathematical expression of convolution operation is as follows:

$$x_j^n = f\left(\sum_{i \in M_i} x_i^{n-1} \cdot k_{ij} + b_j^n\right). \quad (1)$$

Where x_j^n is the j-th feature graph of the n-th convolution layer. x_i^{n-1} is the i-th output feature graph of layer n-1. $f(\cdot)$ represents the activation function. M_i represents the set of input graphs. k_{ij} is the convolution kernel between the i-th feature graph of the previous layer and the j-th feature graph of the current layer. b_j^n is the bias of the n-th layer, i.e. the current layer. Convolutional neural networks generally contain multiple convolutional layers to globalize extracted features.

2.2. Pooling layer

The pooling layer performs function transformation on the non-overlapping rectangular areas in the output feature map of the previous layer to obtain higher level invariant features [13]. Its function is to aggregate the feature graph obtained from the convolution layer, reduce the dimension of the feature graph, and reduce the sensitivity of the output to tilt, displacement and other forms of deformation, thus enhancing the generalization ability of the model. In the process of forward calculation of input image and feature image, the local features of the image are gradually expanded and integrated into global features through pooling operation. The commonly used pooling methods include mean pooling and max pooling. Pooling layer can keep original feature information while reducing feature dimension.

$$x_j^n = f(\beta_j^{n-1} \text{down}(x_j^{n-1}) + b_j^n). \quad (2)$$

Where x_j^n is the j-th feature map of the n-th pooling layer. $f(\cdot)$ represents the activation function. $\text{Down}()$ indicates the pooling process. β_j^n is multiplicative bias. b_j is additive bias.

2.3. Fully connection layer

Because the convolution layer usually uses multiple convolution kernel templates. Therefore, the output is also juxtaposed feature maps of the same size. In order to fuse these feature maps together for classification, one or more fully connected layers are required. The full connection layer is generally connected between the pooling layer and the classifier, which is used to fuse different features expressed by multiple feature graphs. The powerful feature extraction capability of convolutional neural network comes from the convolution operation of multi-convolutional kernel template. The output of each convolution kernel template represents a feature expression from different Angles. Therefore, the integration of these features becomes an important process. Each neuron of the full connection layer is connected with all neurons of the output feature graph of the upper layer, which can be expressed as:

$$h(x) = f(W^T x + b). \quad (3)$$

Where x is the input of the feature graph of the previous layer. $h(x)$ is the output of the full connection layer. W is the connection weight and b is the bias. $F()$ is the activation function. The full connection layer combines all the features of the previous layer's feature map and then inputs them into the softmax classifier. The common activation functions include Sigmoid, Tanh and Relu (Rectified Linear Unit).

3. Proposed Two Channel R-CNN

R-CNN algorithm consists of two networks: (1) a region proposal network (RPN) candidate box extraction network, which can extract the candidate regions of interest (RoIs) that may contain the target; (2) R-CNN network is used to classify RoIs (target or background) and refine the bounding box (BBox) of target area. Using RPN network to generate BBox is the main innovation of R-CNN compared with other detection algorithms. During the training, the method of alternating training RPN and R-CNN is adopted.

Step1. Train RPN.

Step2. Train R-CNN with candidate region extracted by RPN.

Step3. R-CNN is used to initialize the convolution layer common to RPN network.

Step4. Select the generation to perform Step1-Step 3 until the end of the training.

This is the training method used in reference [35]. In the first generation selection, the model obtained by ImageNet is used to initialize the parameters of the convolution layer in RPN and R-CNN. Starting from the second generation selection, when training RPN, the shared convolutional layer parameters of R-CNN are used to initialize the shared convolutional layer parameters in RPN, and then only the convolutional layer not shared by fine-tuning and corresponding parameters of other layers. When R-CNN is trained, its convolution layer parameters shared with RPN are kept unchanged. Only the parameters corresponding to the layer not shared by fine-tuning can be realized, so that the feature sharing training of two network convolution layers can be realized.

3.1. RPN

R-CNN uses the RPN network to extract gesture candidate regions, which is essentially a sliding window. RPN obtains a series of target candidate regions with target scores from

images of any size. The specific process is as follows. A small network is used for sliding scanning on the feature graph obtained in the last convolutional layer. This network is fully connected with the $N \times N$ window on the feature graph every sliding. Then it maps to a lower-dimensional vector, such as 512 dimensions. Finally, the low-dimensional vector is sent into two fully connected layers, namely, box-regression Layer (REG) and box-classification Layer (CLS). For each position, CLS outputs the probability belonging to the foreground and background from the 512-dimensional features. REG outputs four panning scaling parameters from the 512-dimensional feature.

RPN anchors consider k possible reference windows for each slide location, which means that each slide location will predict up to nine candidate areas at once. For a $W \times H$ feature map, $W \times H \times k$ candidate regions are generated. RPN's anchor has translation invariance. Its principle is to sample the slide-to-height ratio of multi-scale anchor points located in the area of $N \times N$ with the window as the center. The base area size is 16×16 , and the width to height ratio is 2:1, 1:1 and 1:2 respectively. The window scale of the center point [8,16,32] is sampled so that nine anchors are created in each sliding window location.

The choices for anchors are as follows. We have two kinds of anchors that are categorized as detection accuracy, that is, the intersection-over-union (IoU) between the object box generated by the model and any one of the tagged boxes. Positive sample calibration rules are as follows.

Rule 1. If the IoU value of anchor box and Ground Truth corresponding to anchor has the maximum value, they will be marked as positive samples.

Rule 2. If the IoU of the candidate frame and marker frame corresponding to anchor is greater than 0.7, it is marked as a positive sample. In fact, rule 2 can basically find enough positive samples, but for some extreme cases, for example, the IoU of candidate boxes and marker boxes corresponding to all anchors is not greater than 0.7, so rule 1 can be used to generate them. Negative sample calibration rules are as follows:

Rule 3. If the IoU of candidate box corresponding to anchor and marker box is less than 0.3, it is marked as negative sample. The rests are neither positive nor negative samples for final training. Candidate boxes that cross image boundaries are also discarded. IoU calculation formula is:

$$IoU = \frac{S_{AnchorBox \cap S_{GroundTruth}}}{S_{AnchorBox \cup S_{GroundTruth}}}. \tag{4}$$

For each anchor, a binary classifier is first attached behind, and two score outputs are used to represent the probability that it is an object and the probability that it is not an object. It then attaches a REG output representing the four coordinate positions of this anchor. The loss function of RPN is defined as:

$$L(p_i, t_i) = s \frac{1}{N_{cls}} \sum_i L_{cls}(p_i + p_i^*) + \lambda \frac{1}{N_{reg}} p_i^* L_{reg}(t_i + t_i^*). \tag{5}$$

Where i represents the index of each RoI. p_i^* is the label representing the category (positive sample=1, negative sample=0). $t_i = t_x, t_y, t_w, t_h$ indicates the offset of the suggestion box relative to the candidate box. t_i^* represents the offset of the marker box with respect to the candidate box. The goal of learning is to make the former close to the value of the latter, and the calculation formula is:

$$t_x = \frac{x - x_\alpha}{w_\alpha}. \tag{6}$$

$$t_w = \log(w/w_\alpha). \tag{7}$$

$$t_h = \log(h/h_\alpha). \tag{8}$$

$$t_h = \log(h/h_\alpha). \tag{9}$$

Where x , y , w and h represent the central coordinates of the proposed area and its width and height respectively. The suggestion box is generated by the candidate box fine-tuning through the regression process until it approaches the marker box.

The classification loss L_{cls} represents the logarithmic loss of the candidate box predicted as the target and background respectively. Regression loss $L_{reg}(t, t'_i) = R(t_i - t_i^*)$. Where, the loss function is:

$$R(x) = 0.5x^2, |x| < 1. \tag{10}$$

N_{cls} represents the number of mini-batches extracted from an image. A mini-batch is made up of 256 candidate regions randomly selected from an image. Among them, the ratio of positive and negative samples is 1:1. If the positive samples are less than 128, more negative samples should be used to meet the requirement that there are 256 candidate regions for training. N_{reg} stands for the total number of anchors. λ is the balance factor of loss in the CLS layer and REG layer, usually $\lambda = 1$. In the detection process, the setting rule is to combine the prediction boxes whose probability is greater than a certain threshold and IoU is greater than a certain threshold by the non-maximum suppression method.

3.2. Two channel R-CNN

The network structure of gesture recognition used in this paper is shown in Figure 1. Images of any size are input to CNN and propagated forward to the last shared convolution layer through CNN. On the one hand, RPN is used to output candidate regions; on the other hand, R-CNN is used to detect and identify targets in candidate regions extracted by RPN.

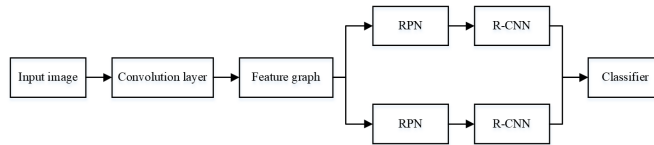


Fig. 1. Proposed network structure diagram of gesture recognition

1. Input parameters of CNN. The gesture data size of the dataset used is 640×480 . To improve the recognition rate, we set the input size as 640×480 for both the training and test phases. The purpose of this paper is to identify 10 gestures in the NTU dataset and detect hands in the VIVA dataset. The number of categories in the NTU and VIVA datasets is set to 11 and 2 (including background), respectively.
2. Anchors parameters. In this paper, the size of gesture area in VIVA and NTU data sets is analyzed. In order to improve the convergence speed during training and the recognition rate in test stage, the benchmark area is set as 8×8 . The ratio of width to height is 2:1, 1:1 and 1:2 respectively. The scale used in this paper is [8,12,16]. The nine anchors examples generated at each slide window location are shown in figure 2.

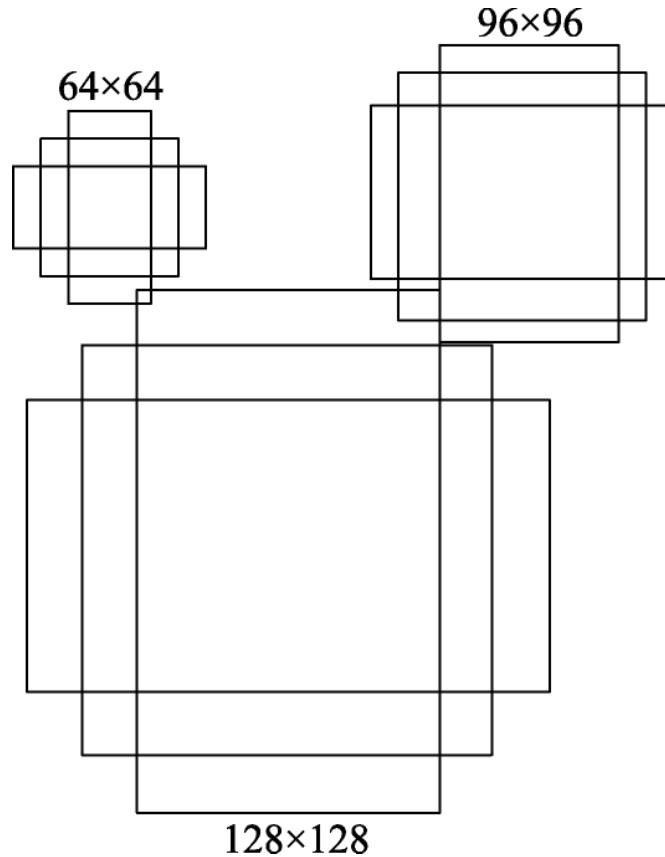


Fig. 2. Anchors scale example diagram

3. Hyperparameters of training. During training, the publicly trained ImageNet classification model is used to initialize the network layer shared by RPN and R-CNN. The remaining layers are randomly initialized with a Gaussian distribution with a

mean value of 0 and a standard deviation of 0.01. After some experimental analysis and comparisons, the learning rate of the first 60K selection is set at 0.001, and the learning rate of the subsequent 40K selection is set at 0.0001. After 104 times of selection training, good results can be achieved. Momentum and weight decay factors use empirical values of 0.9 and 0.0005.

3.3. Algorithm implementation

Due to the large amount of data required for training with CNN, it is easy to overfit when there are few images in the data set. It is a major challenge to obtain a robust CNN model to avoid the problem of over-fitting during training. At present, it is mainly through strengthening regularization of loss layer to avoid over-fitting, which means that the degree of model fitting to training data is much higher than test data. In this paper, a variety of over-fitting methods (data enhancement, weight attenuation, k-fold cross validation, etc.) are tried and found that the improvement effect is not obvious. Reference [36] used noise data to train the performance of CNN, in which some free labels might be correct or wrong for each training image. However, this method was not applicable to training R-CNN network. The Newlabling algorithm proposed in this paper adds disturbance labels to training data to reduce the degree of fitting. During the training, 10 images are randomly selected for each generation. Since positive and negative samples are not specified in the production of data sets, but are determined in training according to the IoU value and the label of the target real box, this paper randomly selects part of the IoU in every 1000 generations according to probability and set it as 0.5. The rest of the IoU is set as 0.7. In essence, when the IoU is set very low, the original positive label may become a negative label, and the negative label may also become a positive label, thus generating disturbance labels. Noise is added to the loss layer by the disturbance tag, and this noise gradient is propagated in the RPN back propagation stage.

The steps of *Newlabling* algorithm are as follows:

In the RPN training stage, the label data sent into RPN is $D = (p_n^*, t_n^*)_{n=1}^L$. Among them, the $p_n^* \in R^{C+1}$, C is category number, $p_n^* = [0, 1, 2, \dots, C]$, 0 is background. $1, 2, \dots, C$ represents C target category markers to be identified. Data labels are 4-dimensional vectors $t_n^* = [x^*, y^*, w^*, h^*]$, they represent the center coordinates of the target on the original image and the width and height of the BBox respectively. L represents the number of images used in each RPN training network. In this paper, $L = 10$. Its aim is to train a model $M : f(p, t, \theta) \in R^2$, θ is a model parameter, which is usually initialized with white noise θ_0 , and then the stochastic Gradient Descent (SGD) algorithm is used for updating. At the m-th iteration,

$$\theta_{m+1} = \theta_m + \gamma_m \cdot \frac{1}{|D_m|} \sum (p, t) \in D_m \cdot \nabla_{\theta_t} [L(p_i, t_i)]. \quad (11)$$

Wherein, $L(p_i, t_i)$ is calculated by formula (2). $\nabla_{\theta_t} [L(p_i, t_i)]$ is used to back-propagate the gradient. γ_m stands for learning rate. D_m randomly selects images from the total data set. In the training (test) stage, RPN firstly outputs the categories, positions and probability scores of 6K candidate regions, and finally selects the top 300 candidate regions with probability scores from these 6K candidate regions, and then feeds the information to R-CNN network.

Newlabling generates disturbance which primarily affects the labeling of 12K candidate area categories. For each candidate region, the disturbance labeling is expressed as $p = p[p_0, p_1]$. Where, p is generated by the input data according to the positive and negative sample calibration rules (p_0 represents the probability that the candidate box is the background, p_1 represents the probability that the candidate box is the target). The size of the IoU is decisive.

$$IoU = \begin{cases} 0.7 & \tilde{I}_j = 1 \\ 0.5 & \tilde{I}_j = 0 \end{cases} \quad (12)$$

$$\tilde{I} = [\tilde{I}_1, \dots, \tilde{I}_N]. \quad (13)$$

N is the number of selected generations each time. In this paper, $N = 1000$. \tilde{I} follows a Bernoulli distribution.

$$j \sim \phi_j(\alpha), \tilde{I}_j = 1, \tilde{I}_i = 0. \quad (14)$$

Where, the $\phi_j = \frac{1}{N} \cdot \alpha$. α is the noise rate.

The role of the noise rate in the *Newlabling* algorithm is as follows. The noise rate α determines the number of possible error labels in each training set. When $\alpha = 0$, there is no noise, and the detection algorithm will be proceeded according to the normal situation. When $\alpha = 100\%$, the accuracy rate of all labels is 50%, that is, random labels at this time are completely unreliable. This *Newlabling* algorithm is applied to NTU and VIVA data sets in this paper, and loss rate results and mAP of VGG_M model are shown in figure 3, figure 4, table 1, and table 2. It can be seen that when $\alpha = 10\%$, higher accuracy can be achieved, and the convergence speed is also improved to a certain extent. The *Newlabling* algorithm can improve the generalization ability of CNN model. When α is higher (up to 15%), the convergence speed and accuracy rate of the network will be reduced. The reason is that the label of training data is not reliable at this time. By comparing figure 3 and figure 4, it can be seen that *Newlabling* algorithm has a more obvious effect on NTU dataset than VIVA dataset. The reason is that the hand of VIVA data set is in a complex background, and its background light intensity changes greatly and the occlusion is serious. However, the gestures in the NTU dataset are in a simple background, and the light intensity basically does not change. There is no gesture occlusion. Therefore, the loss rate jitter of VIVA data set is larger and the convergence rate is slower in each iteration.

Table 1. mAP with different α on NTU data sets/%

α	mAP
0	98.2
5	98.4
10	99.0
15	97.9

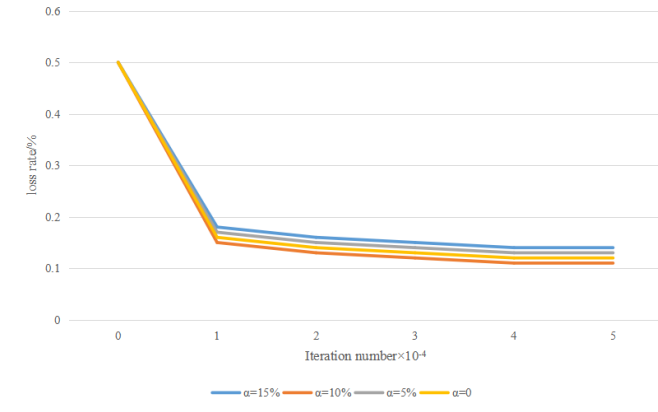


Fig. 3. Loss rate of gesture recognition using different α on NTU datasets

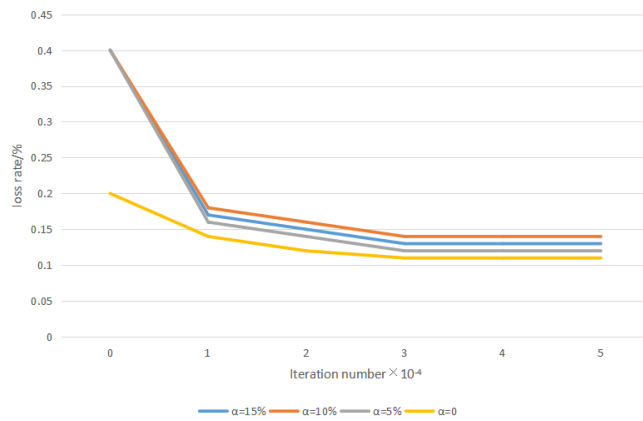


Fig. 4. Loss rate of gesture recognition using different α on VIVA datasets

Table 2. mAP with different α on VIVA data sets/%

α	mAP
0	83.8
5	84.6
10	84.7
15	83.2

3.4. Experiments and analysis

Experimental environment is Ubuntu14.04 with the py-RCNN (<https://github.com/rbgirshick/py-rcnn>). NTU and VIVA gesture data sets are adopted. In this paper, only color images are used, and the depth images in the dataset are not used. The original gesture images in the database are 248-256 or 128-128 pixels. The image contains a large amount of data, and gestures occupy a small area in the whole image with a lot of background redundancy. If the original image is directly used as the input of the convolutional network, the amount of data to be processed will be very large and the classification results will be easily affected by the complex background. Therefore, the image is preprocessed and then used as the input of the model. The image information content after preprocessing is reduced to 1/64 of the original image.

In order to verify the practical application effect of two-channel convolutional neural network in gesture recognition, two groups of experiments are designed here. The first group compares the recognition effects of single channel and two-channel convolutional neural networks. At the same time, the recognition accuracy of two-channel convolutional neural networks with different convolution kernel sizes is compared. The second experiment compares the recognition accuracy of the proposed algorithm with that of previous gesture recognition algorithms to verify the improvement of feature extraction ability and recognition accuracy of two-channel convolutional neural network.

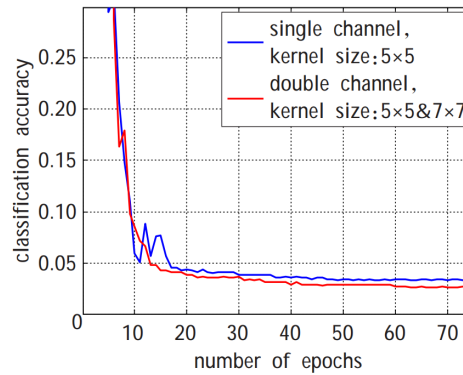
According to the principle of the proposed algorithm in this paper, the improvement of feature extraction ability of two-channel convolutional neural network mainly comes from convolution kernels with different scales. However, for different processing objects and different application scenarios, the selection of the convolution kernel size is not fixed. Only with the most appropriate convolution kernel, the best classification effect can be achieved. Therefore, the networks selected for this experiment include single-channel convolutional neural network with convolution kernel size of 5-5 and two-channel convolutional neural network with convolution kernel size of 3×3 , 5×5 , which are matched with four sizes. Experiments are conducted on two static gesture databases respectively.

Table 3 shows the recognition results of different convolution kernel sizes in the data sets. After using two-channel convolutional neural network, gesture recognition accuracy is significantly improved compared with single-channel convolutional neural network. At the same time, the convolution kernels of different sizes match each other, and the recognition effect is also different. According to the image sizes of the data sets and the network input after pre-processing in this experiment, the best recognition effect can be obtained by the collocation of convolution kernels with the size of 5×5 and 7×7 . The experimental results show that the proposed model uses two independent convolutional channels to extract features from input images, and can obtain richer feature information than single-channel networks. The feature information is reflected in local features with different sizes. Combining these features to classify images can effectively improve the accuracy of static gesture recognition of convolutional neural network.

Figure 5 compares the classification errors of single-channel and double-channel convolutional neural networks with the increase of iteration number during training process. The experimental results show that the classification accuracy of the model tends to be stable after more than 20 iterations. The error rate of two-channel convolutional neural network is obviously better than that of single-channel convolutional neural network and has a more stable convergence process.

Table 3. Gesture recognition accuracy of two-channel model and single-channel model with different sizes of convolution kernel/%

Model	NTU	VIVA
Single channel RCNN(5×5)	96.99	96.05
Two-channel RCNN($3 \times 3, 5 \times 5$)	97.76	96.76
Two-channel RCNN($5 \times 5, 5 \times 5$)	97.75	97.23
Two-channel RCNN($7 \times 7, 5 \times 5$)	98.21	97.67
Two-channel RCNN($3 \times 3, 7 \times 7$)	96.71	96.48

**Fig. 5.** The curve of recognition rate changes with the number of iterations

In this experiment, the representative traditional gesture recognition algorithm and the recognition method based on convolutional neural network are selected and verified on the static gesture database successively. Based on the above experimental results, the two-channel convolutional neural network model chooses the convolution kernel of 7×7 , 5×5 .

The results of comparative experiments are listed in Table 4. Big and Deep MPCNN methods combine maximum pooling with convolutional neural network to form deep convolutional neural network, achieving a recognition rate of 96.88%. Bottom-up structured DCNN is an end-to-end deep convolutional neural network with a gesture recognition accuracy of 88.89%. The recognition accuracy of the proposed algorithm reaches 98.21%, which is higher than the traditional convolutional neural network model.

Table 4. Recognition rate comparison with different gesture recognition algorithms

Method	Recognition rate/%
Spatial Pyramid [37]	85.43
bottom-up structured DCNN [38]	88.89
Tiled CNN [39]	90.59
Big and Deep MPCNN [40]	96.88
Proposed Method	98.21

Based on the above results, the following conclusions can be drawn:

1. Two-channel convolutional neural network uses two convolutional channels with convolution kernels of different sizes to process the input image, so that the network can learn more features. The adequacy of feature extraction is higher than that of traditional single-channel convolutional neural network, so the accuracy of gesture recognition is better.
2. The proposed algorithm in this paper extends the traditional convolutional neural network, but also adopts the supervised learning method for network training. The process of feature extraction does not need human participation, which reflects the excellent scalability of convolutional neural network. At the same time, it also reflects the great potential of the structural expansion of convolutional neural network to improve performance.

4. Conclusion

In this paper, a two-channel convolutional neural network model is proposed, which uses different convolution sizes to check the original gesture images for feature extraction, so as to obtain richer local information and overall topology of gestures. After pooling, features are fused at the full connection layer to extract deeper classification information. The experimental results show that the two-channel convolutional neural network can classify 24 kinds of gestures and adapt to various background forms such as simple and complex, bright and dark, with strong generalization ability. At the same time, two-channel convolutional neural network has a lot of research and development space, mainly including the following three aspects: (1) try to introduce more hierarchical and scale features to further improve the adaptability of the model to complex background; (2) At present, the accuracy of dynamic gesture recognition still has a lot of room for improvement, and the model can be applied to the field of dynamic gesture recognition; (3) The convolutional neural network model for gesture recognition needs a large number of labeled image data for training. In the future, network training can be carried out through unsupervised or semi-supervised learning to reduce the model's dependence on a large number of labeled data.

Availability of data and materials. The data used to support the findings of this study are available from the corresponding author upon request.

Competing interests. The authors declare that they have no conflicts of interest.

References

1. Nguyen K A. "Utilizing a Human-Computer Interaction Approach to Evaluate the Design of Current Pharmacogenomics Clinical Decision Support," *Journal of Personalized Medicine*, vol. 11, 2021.
2. Zhong Q, Yang Q. "Analyzing the Mental States of the Sports Student Based on Augmentative Communication with Human-Computer Interaction," *Journal of Interconnection Networks*, 2021.

3. Jing Yu, Hang Li, Shoulin Yin. "Dynamic Gesture Recognition Based on Deep Learning in Human-to-Computer Interfaces," *Journal of Applied Science and Engineering*, vol. 23, no. 1, pp. 31-38, 2020.
4. Chaaba Ne S, Etien Ne A M, Schyns M, et al. "The Impact of Virtual Reality Exposure on Stress Level and Sense of Competence in Ambulance Workers," *Journal of Traumatic Stress*, 2021.
5. X. Zhang, F. Zhang and C. Xu. "Joint Expression Synthesis and Representation Learning for Facial Expression Recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 3, pp. 1681-1695, 2022.
6. Feng T. "Mask RCNN-based Single Shot Multibox Detector For Gesture Recognition In Physical Education," *Journal of Applied Science and Engineering*, vol. 26, no. 3, pp. 377-385, 2022.
7. Gasteiger N, Hellou M, Ahn H S. "Factors for Personalization and Localization to Optimize Human-Robot Interaction: A Literature Review," *International Journal of Social Robotics*, 2021:1-13.
8. Eskofier B M. "A Smart Capacitive Sensor Skin with Embedded Data Quality Indication for Enhanced Safety in Human-Robot Interaction," *Sensors*, vol. 21, 2021.
9. Hamuda E, Ginley B M, Glavin M, et al. "Automatic crop detection under field conditions using the HSV colour space and morphological operations," *Computers & Electronics in Agriculture*, vol. 133(Complete), pp. 97-107, 2017.
10. Udoh N, Ekpenyong M. "A Knowledge-Based Framework for Cost Implication Modeling of Mechanically Repairable Systems with Imperfect Preventive Maintenance and Replacement Schedule," *Journal of Applied Science and Engineering*, vol. 26, no. 2, pp. 221-234, 2022.
11. Bhattacharjee H, Anesiadis N, Vlachos D G. "Regularized machine learning on molecular graph model explains systematic error in DFT enthalpies," *Scientific Reports*, vol. 11, no. 1, 2021.
12. J. Wan, Q. Ruan, G. An and W. Li, "Gesture recognition based on Hidden Markov Model from sparse representative observations," *2012 IEEE 11th International Conference on Signal Processing*, 2012, pp. 1180-1183, doi: 10.1109/ICoSP.2012.6491787.
13. Q. Chen, N. D. Georganas and E. M. Petriu, "Real-time Vision-based Hand Gesture Recognition Using Haar-like Features," *2007 IEEE Instrumentation & Measurement Technology Conference IMTC 2007*, 2007, pp. 1-6, doi: 10.1109/IMTC.2007.379068.
14. Lin S, Yuan W, Jing L, et al. Blurred palm-print recognition based on fusion of Laplacian smoothing transform and geometric features of hand," *Chinese Journal of Scientific Instrument*, vol. 34, no. 2, pp. 415-422, 2013.
15. Asaari M S M, Suandi S A, Rosdi B A. Fusion of Band Limited Phase Only Correlation and Width Centroid Contour Distance for finger based biometrics," *Expert Systems with Applications*, vol. 41, no. 7, pp. 3367-3382, 2014.
16. Liu F, Liu H Y, Gao L, et al. Hand shape recognition based on fusion features of fingers and particle swarm optimization," *Optics & Precision Engineering*, vol. 23, no. 6, pp. 1774-1782, 2016.
17. X. Zhu, W. Liu, X. Jia and K. -Y. K. Wong, "A two-stage detector for hand detection in ego-centric videos," *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1-8, 2016, doi: 10.1109/WACV.2016.7477665.
18. Qingwu Shi, Shoulin Yin, Kun Wang, Lin Teng and Hang Li. Multichannel convolutional neural network-based fuzzy active contour model for medical image segmentation," *Evolving Systems*, 2021. <https://doi.org/10.1007/s12530-021-09392-3>
19. Wu S. "Simulation of classroom student behavior recognition based on PSO-kNN algorithm and emotional image processing," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 4, pp. 7273-7283, 2021.
20. Nguyen-Trong K, Vu H N, Trung N N, et al. "Gesture Recognition Using Wearable Sensors With Bi-Long Short-Term Memory Convolutional Neural Networks," *IEEE Sensors Journal*, vol. 21, no. 13, pp. 15065-15079, , 2021.

21. Y Peng, Wang J, Pang K, et al. "A Physiology-Based Flexible Strap Sensor for Gesture Recognition by Sensing Tendon Deformation," *IEEE Sensors Journal*, vol. 21, no. 7, pp. 9449-9456, 2021.
22. Fang Y, Zhang X, Zhou D, et al. "Improve Inter-day Hand Gesture Recognition Via Convolutional Neural Network-based Feature Fusion," *International Journal of Humanoid Robotics*, 2021.
23. Singh D K. "3D-CNN based Dynamic Gesture Recognition for Indian Sign Language Modeling," *Procedia Computer Science*, vol. 189, pp. 76-83, 2021.
24. Rahman M A. "Recognition of Static Hand Gestures of Alphabet in Bangla Sign Language," *IOSR Journal of Computer Engineering*, vol. 8, no. 1, pp. 07-13, 2012.
25. M. Panwar, "Hand gesture recognition based on shape parameters," *2012 International Conference on Computing, Communication and Applications*, pp. 1-6, 2012, doi: 10.1109/IC-CCA.2012.6179213.
26. Dominio F, Donadeo M, Zanuttigh P. "Combining multiple depth-based descriptors for hand gesture recognition," *Pattern Recognition Letters*, vol. 50, pp. 101-111, 2014.
27. Yang X, Feng Z, Huang Z, et al. "Gesture Recognition Based on Combining Main Direction of Gesture and Hausdorff-like Distance," *Journal of Computer-Aided Design & Computer Graphics*, 2016.
28. X. Zhang, F. Zhang and C. Xu, "Joint Expression Synthesis and Representation Learning for Facial Expression Recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 3, pp. 1681-1695, March 2022, doi: 10.1109/TCSVT.2021.3056098.
29. Y. Xia, W. Zheng, Y. Wang, H. Yu, J. Dong and F. -Y. Wang, "Local and Global Perception Generative Adversarial Network for Facial Expression Synthesis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 3, pp. 1443-1452, March 2022, doi: 10.1109/TCSVT.2021.3074032.
30. H. Zhang, W. Su, J. Yu and Z. Wang, "Identity-Expression Dual Branch Network for Facial Expression Recognition," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 13, no. 4, pp. 898-911, Dec. 2021, doi: 10.1109/TCDS.2020.3034807.
31. Chen C., Liu MY., Tuzel O., Xiao J. "R-CNN for Small Object Detection," *Computer Vision - ACCV 2016. ACCV 2016. Lecture Notes in Computer Science*, vol. 10115, 2017. Springer, Cham.
32. T. H. N. Le, Y. Zheng, C. Zhu, K. Luu and M. Savvides, "Multiple Scale Faster-RCNN Approach to Driver's Cell-Phone Usage and Hands on Steering Wheel Detection," *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 46-53, 2016, doi: 10.1109/CVPRW.2016.13.
33. D. Poux, B. Allaert, N. Ihaddadene, I. M. Bilasco, C. Djeraba and M. Bennamoun, "Dynamic Facial Expression Recognition Under Partial Occlusion With Optical Flow Reconstruction," *IEEE Transactions on Image Processing*, vol. 31, pp. 446-457, 2022, doi: 10.1109/TIP.2021.3129120.
34. F. Zhang, T. Zhang, Q. Mao and C. Xu, "Geometry Guided Pose-Invariant Facial Expression Recognition," *IEEE Transactions on Image Processing*, vol. 29, pp. 4445-4460, 2020, doi: 10.1109/TIP.2020.2972114.
35. Ahmad M, Ahmed I, Jeon G. "An IoT-enabled real-time overhead view person detection system based on Cascade-RCNN and transfer learning," *Journal of Real-Time Image Processing*, vol. 6, 2021.
36. Sukhbaatar S, Bruna J, Paluri M, et al. Training convolutional networks with noisy labels[OL]. [2017-06-01]. <https://arxiv.org/abs/1406.2080>
37. Wang J, Lv P, Wang H, et al. "SAR-U-Net: squeeze-and-excitation block and atrous spatial pyramid pooling based residual U-Net for automatic liver CT segmentation," *Computer Methods and Programs in Biomedicine*, vol. 208, 2021.
38. Jin Y, Bhatia A, Wanvarie D. Seed Word Selection for Weakly-Supervised Text Classification with Unsupervised Error Estimation. 2021. <https://doi.org/10.48550/arXiv.2104.09765>

39. M Trusca, Spanakis G. "Hybrid Tiled Convolutional Neural Networks (HTCNN) Text Sentiment Classification." 2020. <https://doi.org/10.48550/arXiv.2001.11857>
40. Zhang C., He D., Li Z., Wang Z. "Parallel Connecting Deep and Shallow CNNs for Simultaneous Detection of Big and Small Objects," *Pattern Recognition and Computer Vision. PRCV 2018. Lecture Notes in Computer Science*, vol. 11259, 2018. Springer, Cham.

Pingping Li is with College of Fine Arts, Zhengzhou Normal University. Her research interests include: art image processing and digital art research.

Lu Zhao is with School of Fine Arts, Yulin Normal University. Main research direction: sports art.

Received: March 22, 2022; Accepted: September 12, 2022.

Adaptive Wavelet Transform Based on Artificial Fish Swarm Optimization and Fuzzy C-means Method for Noisy Image Segmentation

Rui Yang^{1,2} and Dahai Li^{1,2}

¹ School of Electronic and Electrical Engineering,
Zhengzhou University of Science and Technology
Zhengzhou, 450015 China
snowycry@qq.com

² Henan Intelligent Information Processing and
Control Engineering Technology Research Center
Zhengzhou, 450015 China
haihaideyou@163.com

Abstract. Aiming at the problem that traditional fuzzy C-means (FCM) clustering algorithm is susceptible to noise in processing noisy images, a noisy image segmentation method based on FCM wavelet domain feature enhancement is proposed. Firstly, the noise image is decomposed by two-dimensional wavelet. Secondly, the edge enhancement of the approximate coefficient is carried out, and the artificial fish swarm (AFS) optimization algorithm is used to process the threshold value of the detail coefficient, and the processed coefficient is reconstructed by wavelet transform. Finally, the reconstructed image is segmented by FCM algorithm. Five typical gray-scale images are selected by adding Gaussian noise and Salt& pepper noise, respectively, and segmented by various methods. The peak signal-to-noise ratio (PSNR) and error rate (MR) of segmented images are used as performance indexes. Experimental results show that compared with traditional FCM clustering algorithm segmentation method, particle swarm optimization (PSO) segmentation method and other methods, the indexes of image segmentation by the proposed method is greatly improved. It can be seen that the proposed segmentation method retains the texture information of image edge well, and its anti-noise performance and segmentation performance are improved.

Keywords: FCM, artificial fish swarm optimization, wavelet transform, noisy image segmentation.

1. Introduction

Image segmentation is a basic pre-processing step to deal with subsequent practical problems, which builds a bridge between initial image processing and later recognition. Image segmentation is to segment an image into object and background. The segmentation process is the grouping process of pixels, which occurs between pixels with similar attributes in the neighborhood, such as intensity, color or texture [1-3].

Noise is an inevitable part of computer vision processing. How to avoid the influence of noise is a significant research direction at present. Fuzzy C-means (FCM) algorithm is a popular image segmentation method at present. A large number of researchers have

carried out researches on noisy image segmentation based on FCM algorithm [4,5]. Compared with the hard segmentation method, FCM algorithm can retain more details of image texture. In view of the weak anti-noise performance of traditional FCM algorithm, relevant scholars propose the methods of combining FCM algorithm with other algorithms to remove noise. Reference [6] proposed a gray image segmentation method based on FCM and artificial bee colony (ABC) optimization. Reference [7-9] proposed an image segmentation method combining particle swarm optimization (PSO) algorithm and FCM. The introduction of PSO algorithm [10], genetic algorithm [11] and gray wolf algorithm [12] improves the robustness of image segmentation to a certain extent. PSO algorithm is easy to fall into local extremum in the search process, and has low convergence accuracy and slow speed in the later stage of evolution. At present, some researchers improve the FCM algorithm to enhance the anti-noise performance of the algorithm. A dynamic parameter FCM image segmentation algorithm based on edge subdivision was proposed in reference [13]. In reference [14], local and non-local information were introduced into the objective function, and the weight between pixel information and non-local spatial information was adjusted by information entropy to enhance the anti-noise performance of the algorithm. The above segmentation methods retain less edge texture information and only improve the anti-noise performance and segmentation performance.

In this paper, based on the traditional FCM image segmentation, a noisy image segmentation method based on wavelet transform is proposed. Firstly, the noisy image is decomposed by two-dimensional wavelet. Secondly, the AFC algorithm is used to process the threshold value of the image detail coefficient, and the edge enhancement of the approximate coefficient is carried out. Then, FCM algorithm is used for image segmentation, so that it can retain more image edge texture information during segmentation.

The rest of the paper is organized as follows. Section 2 and section 3 introduce the related works and preliminaries, including WTF, FCM. Section 4 displays the modified artificial fish swarm algorithm. The proposed noise image segmentation model is proposed in Section 5. The experiments are conducted on section 6, with the conclusion of our manuscript outlined in Section 7.

2. Related Works

Image segmentation is an important process of dividing an image into several regions with similar or identical features (including brightness, color, texture, etc.). In recent years, a variety of image segmentation algorithms have emerged for different application occasions [15]. Clustering method has been widely used in the field of image segmentation [16]. Fuzzy C-Means clustering (FCM) is a soft clustering algorithm based on fuzzy set theory. Different from hard clustering algorithm, each data point has a certain degree of membership for all cluster clusters. Through several iterations, the minimum value of the objective function is found and the cluster cluster with the maximum membership degree of each data point is output. Although FCM clustering algorithm has good segmentation performance for noiseless images, it does not consider information other than pixels, so its segmentation effect for noisy images needs to be improved. Reference [17] proposed a suppression FCM algorithm (S-FCM). Through competitive learning mechanism, the clustering cluster with the largest membership degree was rewarded and other clustering clusters were punished, so as to accelerate the convergence speed of the objective func-

tion and maintain the clustering effect. Reference [18] proposed a Bias Corrected FCM (BCFCM) with the introduction of spatial neighborhood restriction term for the segmentation of medical brain images. With the neighborhood restriction term, it had certain robustness to noise.

Reference [19] proposed a FCM with Generalized Improved Fuzzy Partitions (GIFP-FCM), a membership limit item was added to the objective function of FCM, which improved the classification effect of cluster clusters and the convergence speed. Reference [20] put forward a new Local information restriction term and added it into the FCM objective function, and put forward a Fuzzy Local information C-means (FLICM), which had a good segmentation effect on noise images. Reference [21] proposed a FCM with non-local spatial information by using the image non-local information and the objective function proposed in reference [22] to solve the problem that only considering the local image information was not enough to obtain good segmentation effect, so as to make more effective use of image information. Reference [23] proposed a self-tuning non-local spatial-information FCM algorithm, which could automatically obtain the most appropriate filtering parameters for different pixels and improve the flexibility and robustness of the algorithm. Reference [24] combined the suppressed FCM algorithm with the intuitive fuzzy set for membership degree, removed the suppressed FCM algorithm parameters, and applied non-local spatial information to propose a suppressed non-local spatial intuitive FCM algorithm (SNLS-IFCM). Reference [25] proposed the attribute similarity of 2-element topological subspace, and presented a new FCM based on similarity of attribute space (FCM-SAS). The accuracy of clustering was improved by using membership degree and sample attribute information of clustering center. FCM algorithm based on kernel method is an important method. The kernel method maps the data that is difficult to classify linearly from low dimension to high dimension so as to achieve linear classification of data in high dimension.

Based on the FLICM algorithm, the Kernel method was substituted for Euclidean distance in reference [26], a new fuzzy factor was given, and the Kernel Weighted FLICM (KWFLICM) algorithm was proposed. Based on the constraint factor in the fuzzy factor of KWFLICM algorithm, reference [27] proposed a new weighted image to be used in the constraint term, realized fuzzy clustering by using the Kernel method instead of Euclidean distance, and presented the adaptive constrained Kernel FCM algorithm (Kernel FCM) Fuzzy C-Means. On the basis of KWFLICM, reference [28] extended the clustering object to multidimensional data, sorted and considered the data and neighborhood of each dimension, realized the clustering of multidimensional data by the kernel method, and gave the Generalized KWFLICM algorithm. However, the above methods do not reduce the number of iterations required for convergence of objective function, and the improvement of segmentation efficiency is not obvious.

3. Preliminaries

3.1. Wavelet Transform

Mallat algorithm proposed the concept of multi-resolution analysis. Since digital images are usually represented by two-dimensional array $f(x, y)$, it is assumed that two-dimensional signal is $f(x, y)$, and two-dimensional Mallat algorithm is adopted to carry out wavelet changes [20,30]. The two-dimensional wavelet transform is defined as:

$$WT(a, b_1, b_2) = \frac{1}{a} \int \int f(x, y) \varphi(x, y) dx dy. \quad (1)$$

Where, a is the introduced normalized factor, which ensures the invariable energy before and after wavelet contraction. φ is the Fourier transform. Its inverse transformation is:

$$f(x, y) = \frac{1}{c_\phi} \int_0^{+\infty} \frac{1}{a^3} da \int \int (a, b_1, b_2) \phi(x, y) db_1 db_2. \quad (2)$$

where

$$c_\phi = \frac{1}{4\pi^2} \int \int \frac{|\phi(w_1 + w_2)|^2}{|w_1^2 + w_2^2|} dw_1 dw_2. \quad (3)$$

Digital images are broken down by two-dimensional wave and it gets four components namely, the diagonal coefficient D , horizontal coefficient H , vertical coefficient V and approximate coefficient A respectively. Where D , H , and V are also called detail coefficients.

3.2. Wavelet threshold function

In this paper, soft threshold function [31] is used for value processing of wavelet coefficients, removing or attenuating the coefficients easily damaged by noise and reserving the useful ones so as to suppress noise. The threshold function is as follows:

$$W = \text{sign}(w)(|w| - \lambda), |w| > \lambda. \quad (4)$$

when $W = 0$, $|w| \leq \lambda$. Artificial fish swarm algorithm performs threshold processing on images through Equation (4). Coefficient A does not carry out threshold processing, because coefficient A contains a lot of details useful to the image.

3.3. Adaptive evaluation

The performance of the fish should be evaluated at each generation selection so that the optimal solution can be obtained. The image is reconstructed by inverse wavelet transform with the detail coefficient and approximate coefficient after the threshold processing. Then, the FCM objective function is calculated according to the reconstructed image, as shown in Equation (5), where the data point is the gray value of each pixel point. This fitness evaluation is directly aimed at the segmentation result, and thus indirectly at the image noise.

$$J = \sum_{i=1}^C \sum_{j=1}^L u_{i,j}^m d^2(h, v_i). \quad (5)$$

Although this objective function needs to be calculated in each generation selection of the artificial fish swarm algorithm, the overall calculation cost is within an acceptable range due to the use of histogram based FCM.

3.4. Edge enhancement

Image noise exists in detail coefficient after wavelet decomposition. In order to retain the texture features of image edge, this paper uses Canny edge detection algorithm [32] to detect image edge information by acting on approximate coefficients. Formula (6) is used to realize image enhancement. The two coefficients are reconstructed by wavelet transform to preserve the edge texture features of the image and improve the image segmentation quality effectively.

$$A_f = k \times A + (1 - k) \times A_e. \quad (6)$$

Where, $k \in [0, 1]$ is a constant. A_e is the coefficient processed by Canny edge detection algorithm.

3.5. FCM

FCM algorithm is a fuzzy clustering algorithm based on objective function. In order to make FCM segmentation faster, this paper uses FCM to cluster gray level instead of pixel points. Because the number of gray levels is generally much smaller than the number of pixels. The essence of image segmentation by clustering method is to divide the gray level set into C class. Each class contains unique clustering center, which can be updated in the continuous generation selection, and the clustering result can be optimized by minimizing the objective function. Membership matrix is used to describe the generic properties of each pixel, and the degree of membership of a single pixel in different categories can be judged by its similarity to the cluster center.

The gray level set in the image is $X = x_1, x_2, \dots, x_c$, it divides these data into C categories, then there will be C cluster centers. x_i is a pixel-related feature vector in one-dimensional vector space, and the objective function can be changed to:

$$J = \sum_{i=1}^C \sum_{j=1}^L s_j u_{i,j}^m d^2(j, v_i). \quad (7)$$

Where s_j is the number of pixels with gray level j . L is gray level quantity. m is the membership factor, $m \in (1, +\infty)$. Lagrange multipliers are used to obtain the clustering center renewal equation and the membership equation that minimizes the objective function.

$$u_{i,j}^{k+1} = \frac{1}{\left(\sum_{h=1}^C \left(\frac{d(h, v_i^k)}{d(h, v_m^k)} \right)^{2/(m-1)} \right)}. \quad (8)$$

$$v_i^k = \sum_{l=1}^L (u_{j,i}^k)^m s_j j / \sum_{l=1}^L (u_{j,i}^k)^m s_j. \quad (9)$$

4. Modified artificial fish swarm algorithm

The artificial fish swarm algorithm imitates the characteristics of fish gathering and abstracts the real fish as an artificial fish in the fish swarm algorithm [33,34] to encapsulate

its own state and behavior. By receiving the stimulus information of the external environment, it selects the corresponding activities and affects the external environment through the change of its own state information. The artificial fish interact with each other to find the highest concentrations of food in the environment. During the state change of the artificial fish, the artificial fish may gather in the center of the range with the highest local food concentration. If the state center with the highest food concentration in the environment of all artificial fish is required, other influencing factors are applied to help the artificial fish jump out of the local optimal state center. The local optimal solution of artificial fish swarm algorithm is mainly due to the constant crowding factor which makes the optimal solution deviate from the reality. The moving step size of the artificial fish is fixed, so the artificial fish cannot continue to search for the optimal solution. In this paper, random number is introduced to add random value to step size to help artificial fish jump out of local optimal solution. Constant crowding factor lengthens the optimization time of the algorithm. In this paper, a fitness function is used to reduce the constant crowding factor adaptively so as to shorten the search time of the algorithm and reduce the error between the optimal solution and the actual value.

This paper improves the movement strategy. First, each artificial fish is rear-ended to determine the center with the highest food concentration and update the bulletin board status and the optimal artificial fish status until the artificial fish stops searching forward.

1. AF-follow behavior

If the current moving artificial fish is X_i , it searches f_i with the highest concentration of food near the artificial fish corresponding to another artificial fish X_j with the largest concentration. If $f_i/n_f = \sigma f_i$, it indicates that there is a high food concentration in the position of X and the surrounding environment of the artificial fish is not crowded, then the artificial fish moves forward in the direction of X_j , otherwise the foraging behavior is performed. This behavior is used to speed up the movement of the artificial fish to a better state.

Let $f_{max} = f_i$, $X_{max} = X_j$, it can obtain the artificial fish X_i forward position.

$$X_{next} = \frac{X_i + (X_{max} - X_i)}{\|X_i + (X_{max} - X_i)\|} \cdot step \cdot Rand(). \quad (10)$$

If the artificial fish stops searching, the foraging behavior of the artificial fish is performed, and the bulletin board and the best artificial fish are selected and updated.

2. AF-prey behaviour

Setting the current artificial fish state as X_i , selecting a state X_j in its visual range,

$$X_j = X_i + visual \cdot Rand(). \quad (11)$$

Judging whether the conditions of artificial fish moving forward are satisfied. Until the artificial fish meets the search times. If the forward condition is still not met, moving one step randomly according to equation (12).

$$X_i^{t+1} = X_i^t + step \cdot Rand(). \quad (12)$$

If the artificial fish stops searching in the above two steps, the clustering behavior will be performed until the iteration termination condition is satisfied.

3. AF-swarm behaviour.

If the current artificial fish status is X_i , it searches for the number of fish n_f near artificial fish X_i and the central location X_c with the highest food concentration. If there is $f_c/n_f = \sigma f_i$, it indicates that the food concentration in the center of the shoal is high and not crowded, and the artificial fish X_i moves forward to the center of the shoal. This behavior is used to make a few artificial fish trapped in the local optimal solution tend to the direction of the global optimal solution.

$$X_{next} = X_i + \frac{X_c - X_i}{\|X_c - X_i\|} \cdot step \cdot Rand(). \tag{13}$$

Otherwise, it performs foraging behavior.

The process of improving artificial fish swarm algorithm is shown in figure 1. Firstly, the total number of fish N and the number of generations selected by the artificial fish swarm algorithm are set to calculate the food concentration of each artificial fish in the current fish swarm. The artificial fish with the highest food concentration is selected as the optimal artificial fish in the current fish swarm. It initializes the state of the best artificial fish as the value of the bulletin board. Combined with the selection times of fish swarm optimization, the moving step size of artificial fish is improved, and the weight is introduced to the step size of artificial fish to solve the problem of convergence speed of the algorithm. According to the characteristics of fish movement, the algorithm chooses to carry out tail-chasing and foraging behaviors first, aiming to enable fish to quickly determine the center position with high food concentration, and then carries out the movement strategy of fish clustering behavior, aiming to avoid overcrowding between fish and adjacent artificial fish and improve the ability of artificial fish to get rid of local optimal solution. According to the above moving strategy, each artificial fish should perform the following operations, such as chasing behavior, clustering behavior and foraging behavior, and update the content of the optimal artificial fish and bulletin board iteratively until the average value obtained for several consecutive times does not exceed the sought extreme value or reaches the maximum number of generations. According to the optimal value saved on the bulletin board, the optimal weight matrix is calculated as the basis for setting the initial parameters of the wavelet function.

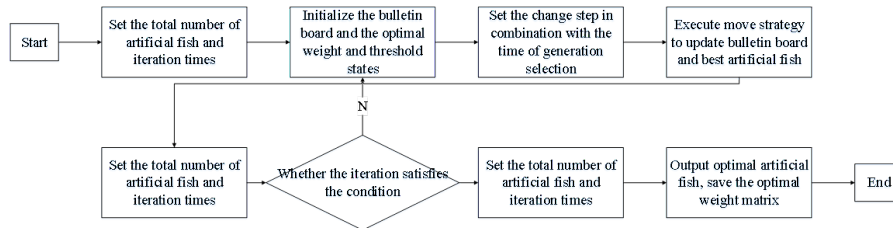


Fig. 1. Optimization process of AF

5. Proposed noise image segmentation

If FCM algorithm segments noise image directly, the segmentation effect is seriously affected by the image noise. Therefore, this paper first uses wavelet transform and AF algorithm fusion to reduce image noise and enhance image edge texture information. Then the reconstructed image is segmented by FCM to make the segmented image more robust. The flow of the proposed method is shown in figure 2.

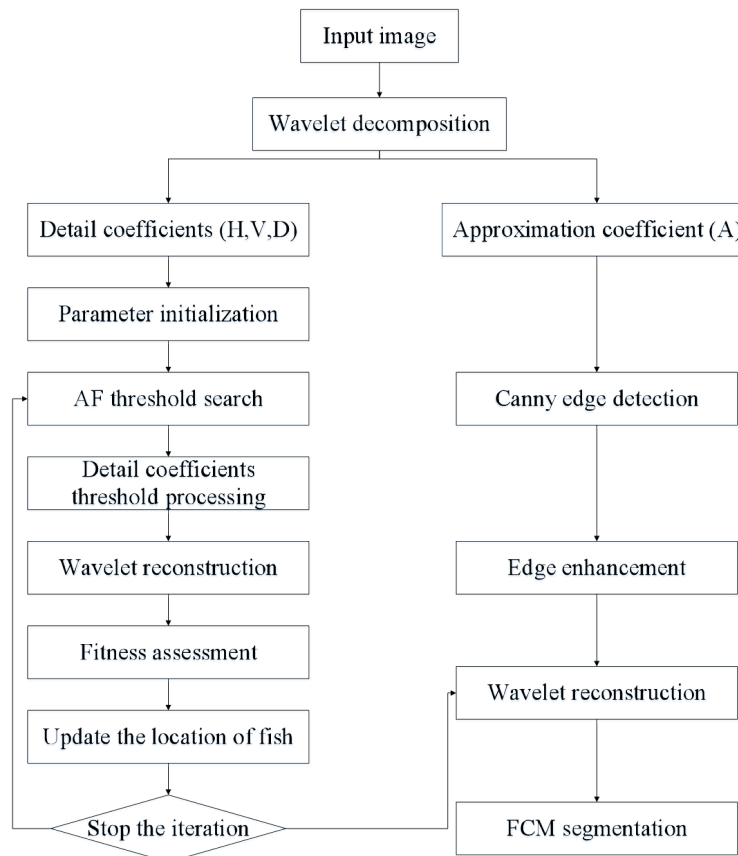


Fig. 2. Flow chart of proposed method

The method in this paper can be divided into four steps as follows.

Step 1. Input the image, perform wavelet decomposition for the image and obtain 4 coefficients:

$$W(X) \rightarrow (A, H, V, D). \quad (14)$$

Where W denotes the wavelet decomposition.

Step 2. AF algorithm is used to search the threshold values of detail coefficients D , H and V respectively. When the fitness value does not exceed the given wide value, the generation selection is stopped after 20 times. After each generation selection, the fuzzy coefficient is processed according to the current threshold, and then the fuzzy coefficient and the unprocessed detail coefficient are reconstructed by wavelet. The reconstructed image is evaluated according to equation (5), which ensures the optimal threshold.

Step 3. Edge enhancement is performed on the fuzzy coefficient A , and the specific expression is shown in Equation (6).

Step 4. The coefficients are reconstructed by wavelet transform, and the reconstructed image is segmented by FCM.

$$W^{-1}(H', V', D', A_f) \rightarrow \hat{X}. \quad (15)$$

$$FCM(X) \rightarrow X_n. \quad (16)$$

Where W^{-1} is the inverse wavelet transform, used to reconstruct the image.

6. Experiments and analysis

In order to verify the segmentation performance and noise suppression ability of the proposed method, five typical grayscale images are selected and named img1, img2, img3, img4, img5 respectively, as shown in figure 3. Gaussian noise and salt and pepper noise are added to the five images respectively. Firstly, the coefficient k in image edge enhancement are set to 0.5, 0.7 and 0.9, respectively. Then, img1 is reconstructed and segmented. The appropriate k value is selected by directly observing the reconstructed and segmented images. The clustering number C is set to 3, 5 and 6 for img4 segmentation experiment. The classification effect and convergence rate of objective function are analyzed, and the appropriate C value is selected. Finally, the proposed method, traditional FCM method and particle swarm optimization algorithm are used for segmentation, and three different results are obtained. The peak signal-to-noise ratio (PSNR) and misclassification rate are used to quantitatively evaluate the anti-noise performance and segmentation performance of the proposed algorithm, and the running time is used to evaluate the time complexity of the algorithm.

The experiment is carried out under Windows10 system, with AMD Ryzen5 2600 CPU, main frequency 3.40 GHz, and 16 GB operating memory. The experimental environment is Matlab a2017 version.

Table 1 shows the image reconstruction and segmentation effects under different k values. It shows that when $k = 0.9$, the reconstructed image texture is the clearest and the segmentation effect is the best.

Figure 4 and figure 5 are the segmentation results of the five images with different noises by using the method in this paper, FCM algorithm and artificial fish swarm algorithm. By directly observing the segmentation results, it can be seen that the image segmented by the proposed algorithm is less affected by noise, especially the segmentation effect of the image containing salt and pepper noise is significantly improved compared with the other two methods, and the proposed algorithm maintains a certain stability under these two kinds of noise.

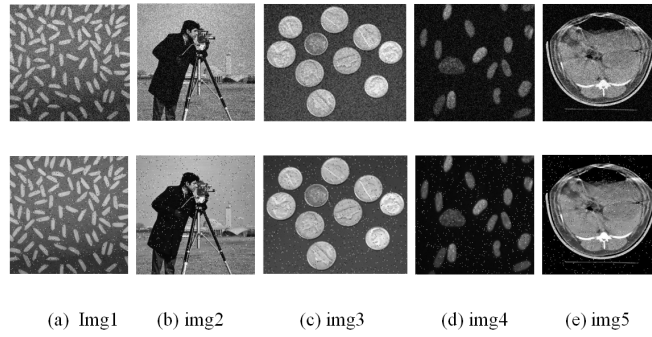


Fig. 3. Original and noise added images. The first row is added Gaussian noise and the second row is added salt & pepper noise.

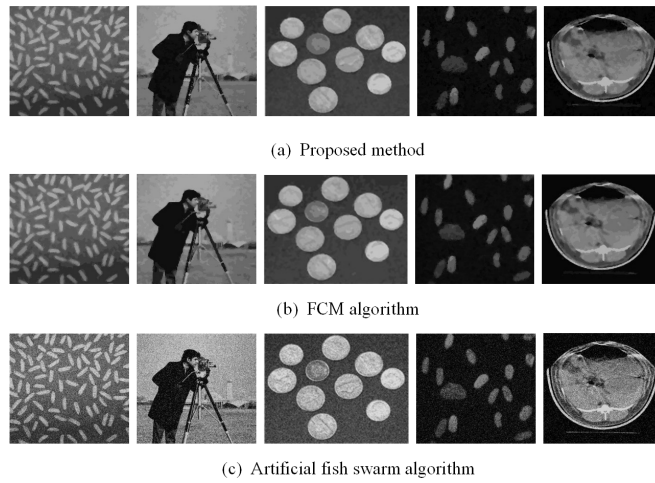


Fig. 4. segmentation results with Gaussian noise

Table 1. Comparison of image reconstruction effect and segmentation effect with different k values

k value	Reconstructed image	Segmented image	Reconstructed image	Segmented image
k=0.5				
k=0.7				
k=0.9				
k value	Gaussian noise	Gaussian noise	salt and pepper noise	salt and pepper noise

In order to further verify the anti-noise performance of the proposed method, the paper adopts the Peak signal-to-noise Ratio (PSNR) for quantitative evaluation [35]. The larger PSNR denotes the better the anti-noise performance of the proposed method, which is defined as:

$$PSNR = 10lg(MAX^2/MSE). \tag{17}$$

Where MAX is the maximum value of image pixels. MSE is the mean square error of image pixels and is defined as:

$$MSE = \frac{1}{mn} \sum_{i=1}^n \sum_{j=1}^m ||K(i, j) - I(i, j)||^2. \tag{18}$$

Where mn is the size of the image. K and I are the original noisy image and the segmented image respectively.

In order to further verify the segmentation performance of the proposed method, Misclassification Error (ME) is used as an indicator to evaluate the segmentation performance of the proposed method. The smaller ME value denotes the better segmentation performance of the proposed method, which is defined as:

$$ME = (1 - \sum_{i=1}^C A_i \cap B_j (\sum_{j=1}^C B_j)^{-1}) \times 100\%. \tag{19}$$

Where A_i represents the pixel points divided into class i in the segmentation algorithm. B_j represents the pixels divided into class j in an ideal image without noise. C is the number of categories.

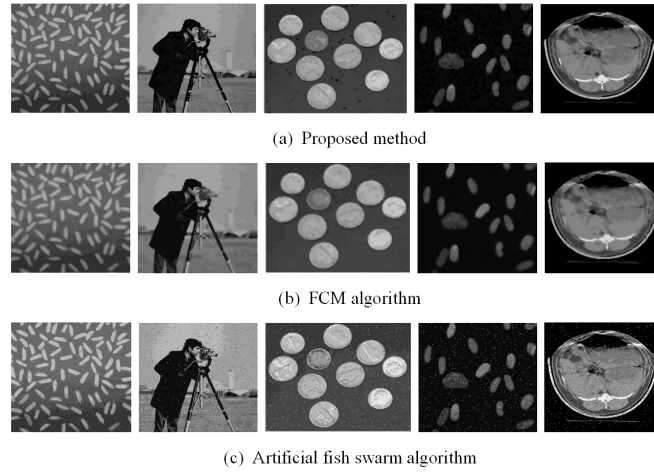


Fig. 5. Segmentation results with salt and pepper noise

In order to objectively evaluate the anti-noise performance and segmentation performance of the segmentation method in this paper, The PSNR value and ME value of segmented images by three methods are given in Table 2. The PSNR value and ME value of segmented images by these three methods are compared and analyzed. The PSNR value of the segmented images by this method is 16% and 20% higher than that of traditional FCM segmentation method and AF algorithm segmentation method on average. ME values decreases by 28% and 13% on average, respectively. The proposed method has better segmentation performance for noisy images and better anti-noise performance. However, compared with the traditional FCM method, the running time of the proposed algorithm is longer than that of the traditional FCM method, which has a slight advantage over the AF algorithm segmentation method.

The following contents are the comparison results with state-of-the-arts image segmentation methods including ACMAWF [36], ACMFR [37], RFRBSFCM [38], PDB-SCAN [39]. Here, mean Intersection over Union (mIoU) and Normalized Mutual Information (NMI) are used to evaluate the effectiveness of the proposed method.

$$mIoU = \frac{1}{K} \sum_{i=1}^K (A_i \cap C_i) / (A_i \cup C_i). \quad (20)$$

Where A_i is the pixel set of the i -th cluster in the segmentation result. C_i is the pixel set of the i -th cluster in the reference image. Larger mIoU indicates better segmentation effect. For gray image I_1 and gray image I_2 ,

$$NMI = 2MI(I_1, I_2) / [H(I_1) + H(I_2)]. \quad (21)$$

Where, I_1 and I_2 have the same size, $MI(I_1, I_2)$ represents the mutual information of I_1 and I_2 . $H(I_1)$ and $H(I_2)$ represent the entropy of I_1 and I_2 respectively. The larger NMI denotes the better segmentation result.

Table 2. PSNR, ME and time consumption of noisy image segmentation by three methods

Method	Images	Noisy image	PSNR/dB	ME	Time/s
FCM	Img1	Gaussian noise	4.2751	0.4744	1.102
FCM	Img1	salt and pepper noise	6.6731	0.4032	1.068
FCM	Img2	Gaussian noise	3.3621	0.4891	1.081
FCM	Img2	salt and pepper noise	7.8971	0.4161	1.094
FCM	Img3	Gaussian noise	5.3441	0.3871	1.242
FCM	Img3	salt and pepper noise	7.6271	0.4251	1.227
FCM	Img4	Gaussian noise	6.3481	0.3541	0.091
FCM	Img4	salt and pepper noise	7.2451	0.3961	1.103
FCM	Img5	Gaussian noise	4.4531	0.3922	1.104
FCM	Img5	salt and pepper noise	5.9599	0.3631	1.730
Proposed method	Img1	Gaussian noise	13.7831	0.2161	1.752
Proposed method	Img1	salt and pepper noise	14.8771	0.2241	1.711
Proposed method	Img2	Gaussian noise	12.4991	0.3341	1.793
Proposed method	Img2	salt and pepper noise	16.2786	0.2998	1.724
Proposed method	Img3	Gaussian noise	15.7391	0.3151	1.778
Proposed method	Img3	salt and pepper noise	16.5051	0.2847	1.786
Proposed method	Img4	Gaussian noise	16.5051	0.2847	1.786
Proposed method	Img4	salt and pepper noise	14.1941	0.3271	1.862
Proposed method	Img5	Gaussian noise	13.5921	0.3281	1.925
Proposed method	Img5	salt and pepper noise	15.9681	0.2971	1.833
AF	Img1	Gaussian noise	10.6561	0.3251	1.756
AF	Img1	salt and pepper noise	9.6499	0.3781	1.732
AF	Img2	Gaussian noise	9.9531	0.3531	1.735
AF	Img2	salt and pepper noise	13.7861	0.3261	1.717
AF	Img3	Gaussian noise	12.4921	0.3471	1.748
AF	Img3	salt and pepper noise	14.7591	0.3541	2.107
AF	Img4	Gaussian noise	14.4941	0.2971	1.686
AF	Img4	salt and pepper noise	14.1981	0.4291	1.982
AF	Img5	Gaussian noise	12.9951	0.3351	1.916
AF	Img5	salt and pepper noise	12.6971	0.3261	1.891

First, we segment the artificial grayscale image, as shown in Figure 6(a). The image size is 256×256 pixels with 5%, 10%, 15% and 20% noise. Figure 6 shows the segmentation results of artificial images containing 5% mixed noise by five algorithms. The quantitative index results of artificial images are shown in Table 3.

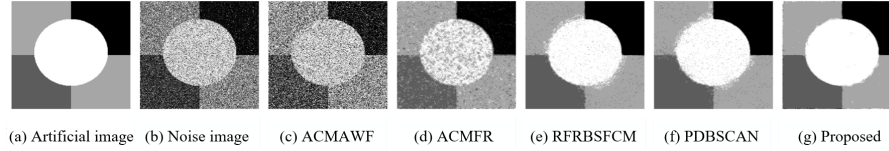


Fig. 6. Segmentation results of artificial images containing 5% mixed noise by five algorithms

Table 3. Segmentation results of artificial images with different mixed noises by five algorithms%

Index	mIoU	mIoU	mIoU	mIoU	NMI	NMI	NMI	NMI
Noise	5	10	15	20	5	10	15	20
ACMAWF	53.76	49.59	36.52	31.52	26.45	15.53	9.20	7.24
ACMFR	86.06	65.36	56.27	50.62	74.43	51.21	39.82	32.01
RFRBSFCM	97.46	88.51	69.57	49.91	87.04	78.08	54.74	39.21
PDBSCAN	97.78	89.86	71.23	55.70	87.11	80.29	58.94	41.24
Proposed	97.81	96.65	93.51	87.82	93.34	90.99	85.46	75.95

The experimental results show that the ACMAWF algorithm has the fastest segmentation speed and poor segmentation effect, and the ACMFR algorithm has improved the segmentation result compared with ACMAWF due to the consideration of local spatial information, but it is not easy to converge. RFRBSFCM and PDBSCAN algorithms use original non-local spatial information, and the latter is superior to the former due to the consideration of intuitive fuzzy sets and membership competitive punishment. In the case of 5% mixed noise, the number of generation selection of PDBSCAN algorithm is less than that of RFRBSFCM algorithm, but when the mixed noise increases to more than 10%, the convergence speed of PDBSCAN algorithm is slower than that of RFRBSFCM algorithm. The segmentation results of the proposed algorithm are better than those of other comparison algorithms, which shows that the proposed algorithm has good segmentation ability and detail retention ability.

The gray scale natural images are segmented with noise. The original images are gear images (263×264 pixels), 42049, 86016 and 118035, respectively. The last three images come from Berkeley image segmentation data set with the size of 481×321 pixels. We add 5%, 10%, 15% and 20% mixed noise to four images respectively. Figures 7 9 and Table 4 compare the segmentation effects and quantitative indexes by five algorithms for four natural images. In Table 4, for each segmentation algorithm, from top to bottom are the

segmentation quantitative index results of gear image, 42049, 86016 and 118035 under different mixed noises.

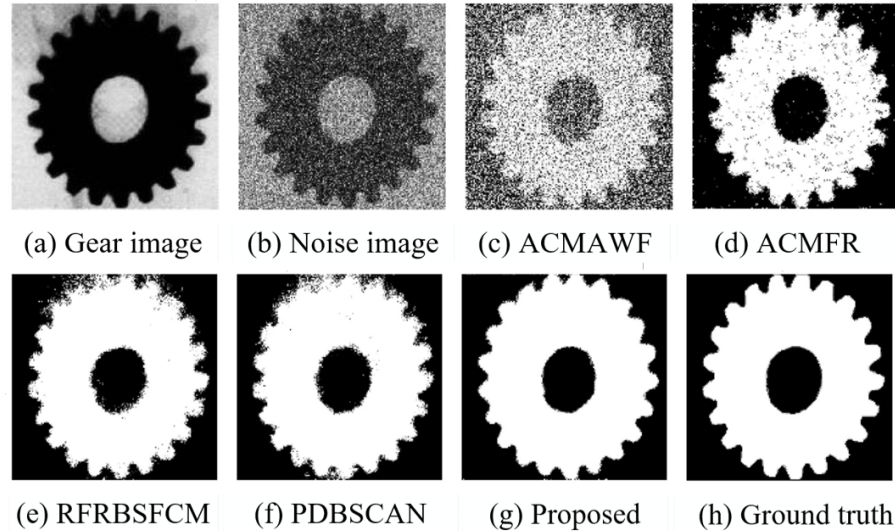


Fig. 7. Segmentation results with five algorithms containing 10% mixed noise for gear image

Experimental results show that because ACMAWF does not consider any image spatial information, the computational complexity is low, the segmentation effect is poor, and the segmentation speed is fastest. In the case of 5% mixed noise and excluding ACMAWF, the segmentation results of binary image by each algorithm are similar. When the intensity of mixed noise increases, ACMFR algorithm degrades the segmentation result of three classification images more than the other four algorithms. The segmentation result of RFRBSFCM is similar to that of PDBSCAN. Due to the original non-local spatial information calculation method, the segmentation time of both is longer. The segmentation result of the proposed algorithm has a small advantage over other algorithms under 5% mixed noise. With the addition of large mixed noise, the segmentation results of the proposed algorithm are better than those of other algorithms.

We also analyze the time complexity of these algorithms. Firstly, the calculation step expression E of the algorithm objective function is calculated. Secondly, all variables in E are unified as variable n , and the calculation step function $E(n)$ is obtained. Finally, let n approaches infinity, find an auxiliary function $f(n)$, so that $f(n)/E(n) = a$, then $E(n)$ and $f(n)$ are the same order of magnitude. $O[f(n)]$ is the time complexity of the algorithm, where a is a constant greater than 0.

In Table 4, H and W are the height and width of the image respectively. K is the number of clusters, $iter$ is the number of generation selection. k is the neighborhood side length of ACMAWF algorithm. S and s are the side length of the search area and the neighborhood side length of the non-local mean filter respectively. ACMFR needs to con-

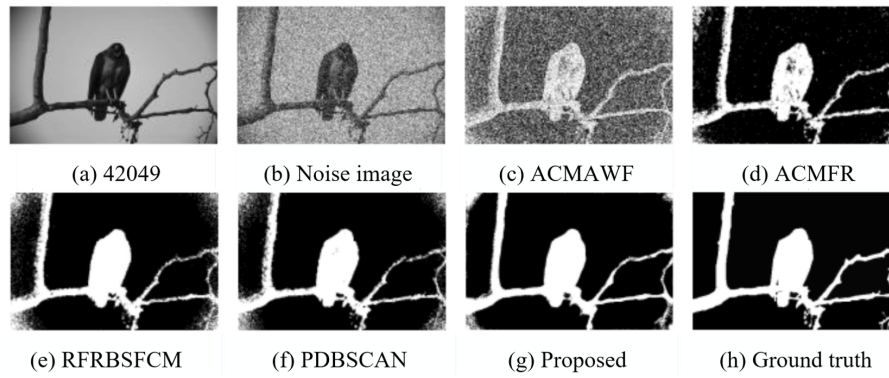


Fig. 8. Segmentation results with five algorithms containing 10% mixed noise for 42049

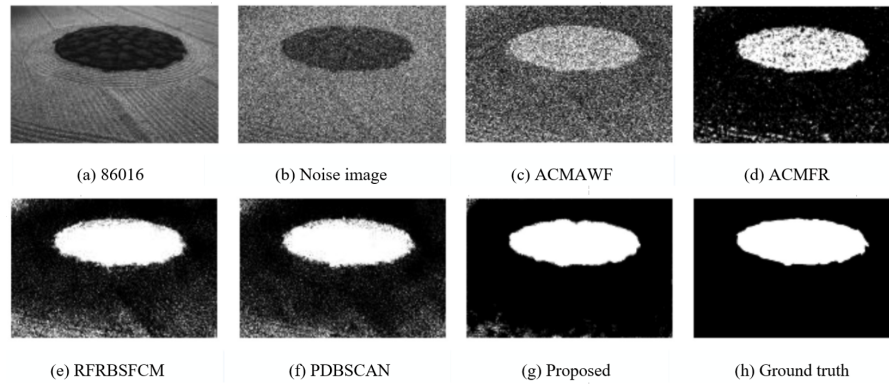


Fig. 9. Segmentation results with five algorithms containing 10% mixed noise for 86016

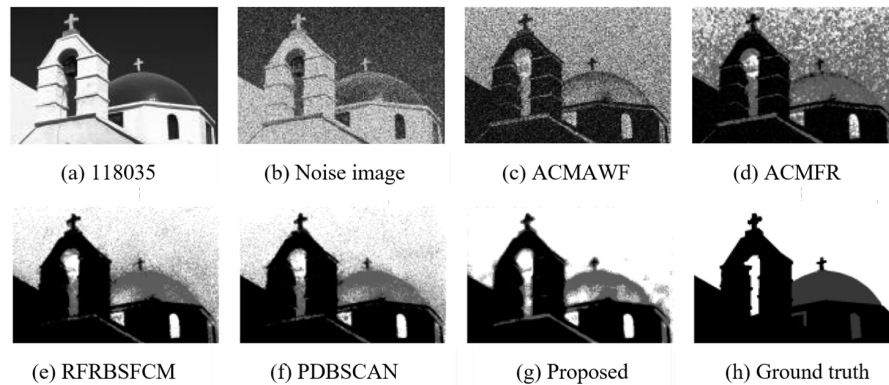


Fig. 10. Segmentation results with five algorithms containing 10% mixed noise for 118035

Table 4. Segmentation results of images with different mixed noises by five algorithms (%)

Index	mIoU	mIoU	mIoU	mIoU	NMI	NMI	NMI	NMI
Noise	5	10	15	20	5	10	15	20
ACMAWF (gear)	83.71	69.99	59.93	55.52	56.10	32.53	18.91	13.43
ACMAWF (42049)	52.27	43.85	39.85	37.37	16.85	8.67	5.61	4.13
ACMAWF (86016)	42.07	37.63	35.14	33.15	9.30	4.38	3.33	2.44
ACMAWF (118035)	37.61	31.84	29.33	13.46	20.56	11.07	7.22	5.23
ACMFR (gear)	98.79	97.38	94.63	90.47	91.91	87.55	80.01	70.14
ACMFR (42049)	87.42	82.03	76.50	72.01	64.92	53.97	43.57	35.74
ACMFR (86016)	79.70	79.21	66.56	62.19	54.02	47.12	29.13	21.09
ACMFR (118035)	47.46	40.31	31.96	26.10	50.12	38.92	28.32	21.58
RFRBSFCM (gear)	98.79	97.46	95.95	93.09	92.08	88.05	84.01	77.73
RFRBSFCM (42049)	87.37	83.50	66.48	52.60	64.80	56.53	32.28	18.38
RFRBSFCM (86016)	89.11	66.23	51.99	43.84	71.96	38.45	22.26	22.88
RFRBSFCM (118035)	68.59	61.19	53.42	42.95	63.93	56.47	45.74	37.54
PDBSCAN (gear)	98.94	97.25	95.81	92.40	92.54	87.51	83.61	76.62
PDBSCAN (42049)	90.75	83.97	68.12	53.67	74.68	57.78	34.14	19.59
PDBSCAN (86016)	90.87	66.62	52.32	43.60	80.92	42.66	23.10	13.34
PDBSCAN (118035)	71.91	60.98	52.83	44.97	66.89	55.26	45.83	36.27
Proposed (gear)	98.00	99.23	97.66	96.30	95.01	93.18	88.81	84.65
Proposed (42049)	97.15	85.66	84.56	83.26	88.12	62.17	59.38	56.99
Proposed (86016)	97.42	93.45	92.49	79.21	88.90	79.43	76.55	54.26
Proposed (118035)	73.86	67.38	68.96	61.14	70.95	66.69	64.24	55.57

sider the $k \times k$ neighborhood of each pixel and membership degree in each generation selection, so the time complexity is $O(n^6)$. Both RFRBSFCM and PDBSCAN algorithms use the original NLM filter, so the time complexity is mainly affected by the time complexity of NLM filter, which is $O(n^6)$. Table 4 shows that the time complexity of the algorithm in this paper is lower, which is $O(n^4)$.

7. Conclusion

The noise image is decomposed by two-dimensional wavelet, then the AF algorithm is used to process the threshold value of the detail coefficient. The approximate coefficients are enhanced by edge enhancement. Finally, the FCM algorithm is used to segment the image. In this paper, Gaussian noise and salt-and-pepper noise are added to five different images, and these images with noise are segmented by the proposed method, traditional FCM method and AF algorithm, respectively. The peak signal-to-noise ratio and misclassification rate of the segmented images are taken as performance indicators. Experimental results show that the proposed method can effectively preserve the edge features of images, and has good anti-noise performance and segmentation performance.

Table 5. Time complexity analysis

Method	E	$E(n)$	Time complexity
ACMAWF	$H \times W \times K \times iter$	n^4	$O(n^4)$
ACMFR	$H \times W \times K \times k^2 \times iter$	n^6	$O(n^6)$
RFRBSFCM	$H \times W \times (2S + 1)^2 \times (2s + 1)^2$ $+ H \times W \times K \times iter$	$n^2(n + 1)^4 + n^4$	$O(n^6)$
PDBSCAN	$H \times W \times (2S + 1)^2 \times (2s + 1)^2$ $+ H \times W \times (K - 1) \times iter$	$n^2(n + 1)^4 + n^4(n - 1)$	$O(n^6)$
Proposed	$H \times W \times [(2S + 1)^2 - 1]$ $+ (2HW) \times K \times iter$	$n^2[(n + 1)^2 - 1] + 2n^4$	$O(n^4)$

References

1. D. Fung, Q. Liu, J. Zammit, K. S. Leung and P. Hu. "Self-supervised deep learning model for COVID-19 lung CT image segmentation highlighting putative causal relationship among age, underlying disease and COVID-19," *Journal of Translational Medicine*, vol. 19, no. 1. (2021)
2. Q. W. Shi, S. L. Yin, K. Wang, L. Teng and H. Li. "Multichannel convolutional neural network-based fuzzy active contour model for medical image segmentation," *Evolving Systems*, (2021) <https://doi.org/10.1007/s12530-021-09392-3>
3. L. Duan, S. Yang, D. Zhang. "Multilevel thresholding using an improved cuckoo search algorithm for image segmentation," *The Journal of Supercomputing*, vol. 77, no. 7, pp. 6734-6753. (2021)
4. X. Xie , Q. Zhang. "An edge-cloud-aided incremental tensor-based fuzzy c-means approach with big data fusion for exploring smart data," *Information Fusion*, vol. 76, no. 5. (2021)
5. S. L. Yin, Y. Zhang, S. Karim. Large Scale Remote Sensing Image Segmentation Based on Fuzzy Region Competition and Gaussian Mixture Model," *IEEE Access*, vol. 6, pp. 26069-26080. (2018)
6. X. Lei and H. Ouyang. "Kernel-Based Intuitionistic Fuzzy Clustering Image Segmentation Based on Grey Wolf Optimizer With Differential Mutation," *IEEE Access*, vol. 9, pp. 85455-85463. (2021)
7. L. Frigau, C. Conversano, F. Mola. "Consistent validation of gray-level thresholding image segmentation algorithms based on machine learning classifiers," *Statistical Papers*, vol. 62. (2021)
8. Z. Zhao, Z. Zeng, K. Xu, C. Chen and C. Guan. "DSAL: Deeply Supervised Active Learning From Strong and Weak Labelers for Biomedical Image Segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 10, pp. 3744-3751. (2021)
9. X. Zheng, T. Chen. "High spatial resolution remote sensing image segmentation based on the multiclassification model and the binary classification model," *Neural Computing and Applications*, (2021).
10. G. An et al. "Short-Term Wind Power Prediction Based On Particle Swarm Optimization-Extreme Learning Machine Model Combined With Adaboost Algorithm," *IEEE Access*, vol. 9, pp. 94040-94052. (2021)
11. B. Chen, Y. Niu and H. Liu. "Input-to-State Stabilization of Stochastic Markovian Jump Systems Under Communication Constraints: Genetic Algorithm-Based Performance Optimization," *IEEE Transactions on Cybernetics*, (2021) doi: 10.1109/TCYB.2021.3066509.
12. A. Nd, B. Mk, C. Xi, et al. "Towards robust partially supervised multi-structure medical image segmentation on small-scale data," *Applied Soft Computing*, vol. 114. (2021)

13. H. Abdellahoum, N. Mokhtari, A. Br Ahimi, et al. "CSFCM: An improved fuzzy C-Means image segmentation algorithm using a cooperative approach," *Expert Systems with Applications*, 166:114063. (2021)
14. D. Chen, S. Yongchareon, E. Lai, J. Yu and Q. Z. Sheng. "Hybrid Fuzzy C-Means CPD-Based Segmentation for Improving Sensor-Based Multiresident Activity Recognition," *IEEE Internet of Things Journal*, vol. 8, no. 14, pp. 11193-11207. (2021)
15. J. Fan and B. Lei. "Image thresholding segmentation method based on reciprocal rough entropy," *Journal of Electronics & Information Technology*, vol. 42, no. 1, pp. 214-221. (2020)
16. W. Seema, B. Keshavamurthy. "A survey on image data analysis through clustering techniques for real world applications," *Journal of Visual Communication and Image Representation*, vol. 55, pp. 596-626. (2018)
17. Z. Zhu, Y. Liu, Z. Zhao, et al. "Improved suppressed fuzzy c-means clustering algorithm for segmenting the non-destructive testing image," *Chinese Journal of Scientific Instrument*, vol. 40, no. 8, pp. 110-118. (2019)
18. D. Wei, Z. Wang, L. Si, et al. "An image segmentation method based on a modified local-information weighted intuitionistic Fuzzy C-means clustering and Gold-panning Algorithm," *Engineering Applications of Artificial Intelligence*, vol. 101, no. 3, pp. 104209. (2021)
19. M. Kaushal, QMD Lohani. "Generalized intuitionistic fuzzy c-means clustering algorithm using an adaptive intuitionistic fuzzification technique," *Granular Computing*, vol. 1, pp. 183-195. (2021)
20. Z. Yang, P. Xu, Y. Yang, et al. "Noise robust intuitionistic fuzzy c-means clustering algorithm incorporating local information," *IET Image Processing*, vol. 15, no. 3, pp. 805-813. (2021)
21. H. Liu, F. Zhao. "Multiobjective fuzzy clustering with multiple spatial information for Noisy color image segmentation," *Applied Intelligence*, vol. 5, pp. 5280-5298. (2021)
22. Y. Li and Y. Shen. "Robust Image Segmentation Algorithm Using Fuzzy Clustering Based on Kernel-Induced Distance Measure," *2008 International Conference on Computer Science and Software Engineering*, 2008, pp. 1065-1068, doi: 10.1109/CSSE.2008.694.
23. H. Zhang, J. Liu. "Fuzzy c-means clustering algorithm with deformable spatial information for image segmentation," *Multimedia Tools and Applications*, vol. 81, no. 8, pp. 11239-11258. (2022)
24. X. Jia, T. Lei, X. Du, S. Liu, H. Meng and A. K. Nandi. "Robust Self-Sparse Fuzzy Clustering for Image Segmentation," *IEEE Access*, vol. 8, pp. 146182-146195, (2020).
25. W. Shi, J. Zhuo, Y. Lan, et al. "A Novel Fuzzy Clustering Algorithm Based on Similarity of Attribute Space," *Journal of Electronics & Information Technology*, vol. 41, no. 11, pp. 2722-2728. (2019)
26. M. Gong, Y. Liang, J. Shi, et al. "Fuzzy C-Means Clustering With Local Information and Kernel Metric for Image Segmentation," *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 573-584. (2013)
27. I. Cherfa, A. Mokraoui, A. Mekhmoukh and K. Mokrani. "Adaptively Regularized Kernel-Based Fuzzy C-Means Clustering Algorithm Using Particle Swarm Optimization for Medical Image Segmentation," *2020 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, 2020, pp. 24-29, doi: 10.23919/SPA50552.2020.9241242.
28. Memon, Kashif, Hussain, et al. "Generalised kernel weighted fuzzy C-means clustering algorithm with local information," *Fuzzy Sets And Systems*, vol. 340, no. 1, pp. 91-108. (2018)
29. H. B. Seidel, M. M. A. da Rosa, G. Paim, E. A. C. da Costa, S. J. M. Almeida and S. Bampi. "Approximate Pruned and Truncated Haar Discrete Wavelet Transform VLSI Hardware for Energy-Efficient ECG Signal Processing," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 68, no. 5, pp. 1814-1826. (2021)
30. A. Hy, B. Yang, C. Rw, et al. "Associations between trees and grass presence with childhood asthma prevalence using deep learning image segmentation and a novel green view index," *Environmental Pollution*, (2021)

31. A. Dz, A. Bh, G. Xun, et al. "ASS-GAN:Asymmetric Semi-supervised GAN for Breast Ultrasound Image Segmentation," *Neurocomputing*, (2022).
32. A. R. Beeravolu, S. Azam, M. Jonkman, B. Shanmugam, K. Kannoopatti and A. Anwar. "Pre-processing of Breast Cancer Images to Create Datasets for Deep-CNN," *IEEE Access*, vol. 9, pp. 33438-33463. (2021)
33. Gao H, Jun-Wei Z. "Estimation of pile settlement applying hybrid radial basis function network with BBO, ALO, and GWO optimization algorithms," *Journal of Applied Science and Engineering*, vol. 25, no. 6, pp. 1031-1044, 2022.
34. Ibrahim, J., Gajin, S. "Entropy-based Network Traffic Anomaly Classification Method Resilient to Deception," *Computer Science and Information Systems*, Vol. 19, No. 1, pp. 87-116. (2022).
35. A. Skd, B. Kd et al. "Opposition-based Laplacian Equilibrium Optimizer with Application in Image Segmentation using Multilevel Thresholding," *Expert Systems with Applications*, vol. 174, (2021)
36. A. Joshi, M.S. Khan, A. Niaz, et al. "Active Contour Model with Adaptive weighted function for Robust Image Segmentation under Biased Conditions," *Expert Systems with Applications*, vol. 175, (2021).
37. J. Fang, H. Liu, J. Liu, et al. "Fuzzy region-based active contour driven by global and local fitting energy for image segmentation," *Applied Soft Computing*, vol. 100, (2021).
38. A. Srinivasan, S. Sadagopan. "Rough fuzzy region based bounded support fuzzy C-means clustering for brain MR image segmentation," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, pp. 3775-3788, (2021).
39. C. L. Seng, B.A. Macdonald, M. Parsons, et al. "Accelerated superpixel image segmentation with a parallelized DBSCAN algorithm," *Journal of Real-Time Image Processing*, vol. 11, pp. 1-16 (2021).

Rui Yang is with the School of Electronic and Electrical Engineering, Zhengzhou University of Science and Technology, Zhengzhou, 450015 China. Research direction: image processing and cloud computing.

Dahai Li is with the School of Electronic and Electrical Engineering, Zhengzhou University of Science and Technology, Zhengzhou, 450015 China. Research direction: image processing.

Received: March 21, 2022; Accepted: September 15, 2022.

BiSeNet-oriented Context Attention Model for Image Semantic Segmentation

Lin Teng and Yulong Qiao*

College of Information and Communication Engineering,
Harbin Engineering University
Harbin 150001, China
tenglinheu@163.com
qiaoyulong@hrbeu.edu.cn

Abstract. When the traditional semantic segmentation model is adopted, the different feature importance of feature maps is ignored in the feature extraction stage, which results in the detail loss, and affects the segmentation effect. In this paper, we propose a BiSeNet-oriented context attention model for image semantic segmentation. In the BiSeNet, the spatial path is utilized to extract more low-level features to solve the problem of information loss in deep network layers. Context attention mechanism is used to mine high-level implied semantic features of images. Meanwhile, the focus loss is used as the loss function to improve the final segmentation effect by reducing the internal weighting. Finally, we conduct experiments on open data sets, and the results show that pixel accuracy, average pixel accuracy, and average Intersection-over-Union are greatly improved compared with other state-of-the-art semantic segmentation models. It effectively improves the accuracy of feature extraction, reduces the loss of feature details, and improves the final segmentation effect.

Keywords: image semantic segmentation, BiSeNet, context attention, focus loss.

1. Introduction

The different elements in the image are formed by combining various pixels together. Therefore, the method of classifying these pixels by elements is called image semantic segmentation. As a core technology in computer vision research, semantic segmentation has many advantages in pixel-level prediction and image classification by using advanced semantic features of images [1,2]. At present, it has been widely used in medical and health care, storage management, traffic safety and many other fields, it has important research value and significance. Image semantic segmentation based on deep learning is a hot topic in recent years. As a large number of deep learning methods which have been successful in image classification, object detection, natural language processing and other fields have been improved and migrated to the field of semantic segmentation. The semantic segmentation technology has made great breakthrough and gradually changed the development trend of various industries [3,4].

In the process of image semantic segmentation, difficulties and challenges are usually faced in the aspects of target, category and background [5]. For the target, even if it is the

* Corresponding author

same target, if the illumination, Angle of view and distance are different or in the static and moving state, the images will be different. And even there will be mutual occlusion between adjacent targets. In terms of categories, there are still differences between the same category, and there are also similarities between the different categories. As for the background, the background in the real scene is relatively complex, which brings great difficulties to the semantic segmentation [6]. For traditional semantic segmentation such as gray segmentation and conditional random fields, the underlying features of the image are usually used to divide the region of the image, and its segmentation accuracy needs to be further improved. At present, with the development of convolutional neural network algorithm (CNN) and its application in semantic segmentation, a large number of semantic segmentation models based on deep learning have been proposed, which can solve the problem of difficult feature selection in traditional semantic segmentation [7].

The application of convolutional neural network (CNN) has made rapid progress in image semantic segmentation. Various semantic segmentation networks based on convolutional neural networks have been proposed. At present, there are roughly three kinds of researches in the field of image segmentation. 1) Improving the segmentation performance by improving the structure of the convolutional network [8,9] and combining with deeper neural networks. Russo et al. [10] proposed an image segmentation algorithm with low parameter number based on convolutional neural network. By improving the deep-level neural network and applying multi-scale dilated convolution, the algorithm increased the scale standardization layer and optimized the network to improve the segmentation effect while reducing the number of parameters. 2) Based on the encoder-decoder architecture [11,12], a variety of methods are adopted to extract feature information to improve the resolution of feature maps and thus improve the segmentation effect. Tian et al. [13] proposed an improved DUpsampling algorithm based on DeepLabV3+ architecture. In the decoding module, a DUpsampling method was adopted to replace bilinear interpolation, which improved the segmentation accuracy and reduced the computational complexity while restoring the size of the feature graph. 3) The attention mechanism [14,15] is used to model the target feature information, so as to highlight the detail information of the feature map and improve the segmentation effect. Li et al. [16] proposed a pyramid attention network (PAN) for image segmentation, which combined feature pyramid attention network (FPA) and global up-sampling attention network (GAU) to replace the dilated space convolution pooling pyramid (ASPP) structure for feature extraction. In the face of complex scenes, the above methods are prone to large segmentation errors. Therefore, in order to solve the above problems and improve the utilization rate of high-level and low-level feature information, a new method based on BiSeNet and context attention model is proposed in this paper.

The rest of this paper is organized as follows. In the second section, we reviewed more related work. In the third section, the description method of the proposed image semantic segmentation is introduced. Then, in the fourth section, we conduct an experimental analysis. Finally, we summarize the paper.

2. Related works

Traditional image segmentation algorithms are based on the color, texture information and spatial structure of the image, and the semantic information of the same region is consis-

tent, but the attributes of different regions are also different. There are many segmentation methods, mainly including simple broad-value segmentation, region growth, edge feature detection and graph division [17]. Reference [18] proposed to use the structured forest method to generate edge probability, and utilize watershed algorithm to transform edge probability into initial cut blocks. In order to avoid over-segmentation, the hypermetric contour graph algorithm was used to select appropriate broad values to generate segmentation blocks to obtain more accurate contour information, and random forest was used to train segmentation blocks to obtain semantic segmentation results. Reference [19] proposed a hierarchical graph partition method, namely, the Oriented Image Foresting Transform (OIFT), which could be customized for the target object group according to its boundary polarity. This method had a small number of image partitions and could accurately isolate the desired target region with known polarity. The local contrast of the image region was used to make it robust to illumination variation and non-uniformity effect. Because no data training was required, the calculations were relatively simple. However, if the segmentation task was difficult, the performance of segmentation should be further improved. Reference [20] proposed an image segmentation method combining global image features with complete convolutional networks. The method used the parameter learning process of the unified deep learning model embedded in the full convolutional network to encode the whole image content and make the segmentation more reasonable and accurate. This kind of method basically obtained the underlying features through the use of artificial design features, and its segmentation efficiency could not well meet the actual requirements.

The semantic segmentation method based on deep learning automatically learns data features instead of using artificial data features, which is different from traditional image segmentation methods. End-to-end semantic segmentation prediction can be completed by using deep neural networks [21]. The three most important processes in deep learning include feature extraction, semantic segmentation and post-processing. After that, many models such as FCN, VGG16, ResNet or deep network semantic segmentation are developed. Reference [22] proposed a method based on ResNet network to fuse shallow feature image information with deep feature image by defining parallel branches. The features were extracted and fused by parallel dilated convolution with different sampling rates, so as to effectively extract the features and context information of different layers. In order to improve the stability of parameter tuning, batch normalized calculation was introduced into the new module. The defect of the convolutional network was that it was insensitive to image details due to its low spatial resolution, and the edge of segmentation was relatively rough. Reference [23] proposed a weakly supervised learning algorithm with size constraints based on improved deep convolutional neural network for image segmentation. Compared with the existing complete supervision methods, the image segmentation process only used the image-level label and the boundary box label to guide, which was easier to implement. Its disadvantage was that the target information was not enough, and the context information would be lost, so that the boundary could not be accurately located. Reference [24] proposed a novel DenseGram network, which could reduce gaps and segmented degraded images more effectively than traditional strategies. Experimental results showed that the proposed dense-Gram network produced the latest semantic segmentation performance on degraded images using PASCAL VOC 2012, SUNRGBD, CamVid and CityScapes data sets. At present, there is no mechanism or structure that can

make the current network deliberately learn the differences between different categories, which also leads to the high-level semantic features sometimes share the information of the target and its own background. The segmentation of the target is not accurate. In reference [25], a multi-scale semantic segmentation model based on deep residual network was proposed. It was mainly used to enhance the segmentation accuracy of remote sensing image of different scale objects in small sample remote sensing image dataset. Although the end-to-end semantic segmentation model structure was implemented, due to the emphasis on feature understanding and target category prediction, the problem of inaccurate positioning between target and background or the boundary of different targets was caused.

Jiang et al. [26] proposed a deep architecture that could run in real time, using residual connection and decomposition convolution to maintain high efficiency and good accuracy. Yi et al. [27] proposed an efficient spatial pyramid module (ESP) based on extended convolution, enabling it to perform efficiently in terms of computation, memory and accuracy. Grant-Jacob et al. [28] proposed a new context-guided network (CGNet), which could effectively learn the joint features of local features and surrounding context, and further improve the joint features through the surrounding context features, so as to improve the real-time performance and accuracy of the network. Li et al. proposed a single lightweight backbone network to aggregate and identify features respectively through sub-network and sub-cascade, so as to reduce the number of parameters and still obtain enough receptive fields, thus enhancing the learning ability of the model and achieving a balance between speed and segmentation performance. BiSeNet(Bilateral Segmentation Network) [29] divided the Segmentation task into two parallel modules (spatial path module and context path module), which took advanced features and receptive fields into account and significantly improved the detection speed of the network. In conclusion, due to the limitations of the image segmentation method based on manual design features, this paper proposes a BiSeNet semantic segmentation network based on context content. Firstly, the overall structure of the improved segmentation network and differences from the original bilateral segmentation network are described, and the role of the proposed feature fusion module in the context path is emphasized. Secondly, the subnetwork feature fusion module which is used to aggregate features of different depth is described in detail. At the same time, focal loss is used as the loss function to solve the problems of unbalanced sample number of different categories and different object differentiation difficulties, so as to improve the accuracy of target recognition and improve the segmentation efficiency. Finally, experiments are carried out on public image data and comparison with other state-of-the-art networks. The results prove the effectiveness of the proposed method.

3. Proposed Image Sematic Segmentation Model

Semantic segmentation technology is one of the main tasks of computer vision. It is based on the pixel level of the image to some regions of the image corresponding semantic labels. In recent years, in order to meet the requirement of semantic segmentation accuracy, semantic segmentation model technology has made some progress. However, the current mainstream real-time semantic segmentation model acceleration methods are compromise accuracy for speed. For example, in image processing, it cuts the original image or

directly changes the size of the original image to limit the network input size to reduce the computational complexity. Although these methods are simple and effective in improving network speed, the loss of spatial detail still affects the detection effect, especially the boundary part, resulting in a decrease in measurement and visualization accuracy.

BiSeNet consists of two components: Spatial Path (SP) and Context Path (CP). The former solves the problem of spatial information loss in deep network by acquiring more low-level features. The latter mainly solves the problem of receptive field constriction. The BiSeNet is shown in figure 1.

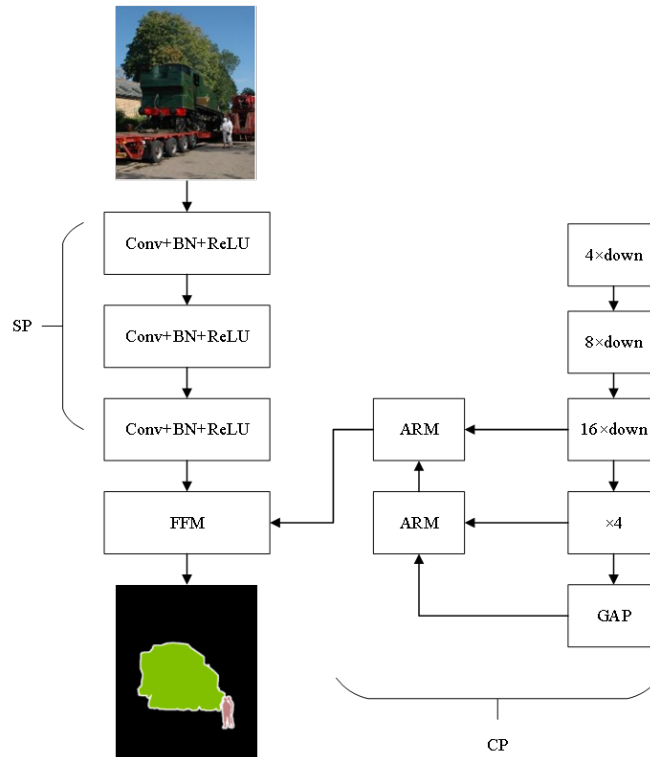


Fig. 1. BiSeNet structure

The overall structure of the improved BiSeNet semantic segmentation network model is also divided into two branches: Spatial Path (SP) and Context Path (CP). SP module is used to obtain high-resolution feature maps and obtain more accurate Spatial information. Its structure consists of three convolutional layers, each of which contains a convolutional layer with step size of 2. After batch standardization processing and ReLU nonlinear activation, the size of the output image through this path is 1/8 of the original image.

CP module enables the network to obtain a larger receptive field. In order to ensure the accuracy and improve the computational speed, Xception model is adopted in the backbone network as a lightweight feature extraction network [30]. Xception can perform fast

down-sampling operation to obtain a large receptive field. The improved BiSeNet semantic segmentation network model is shown in figure 2. The features of different depths are aggregated to obtain the sub-network feature fusion module, and the advanced features are further processed to refine the advanced features. At the same time, feature maps of the same size at each stage of the backbone network are fused to make the context path module possess more low-level features and spatial information, retain spatial details of image structure, and improve its judgment ability of large-scale targets and fine structure edges.

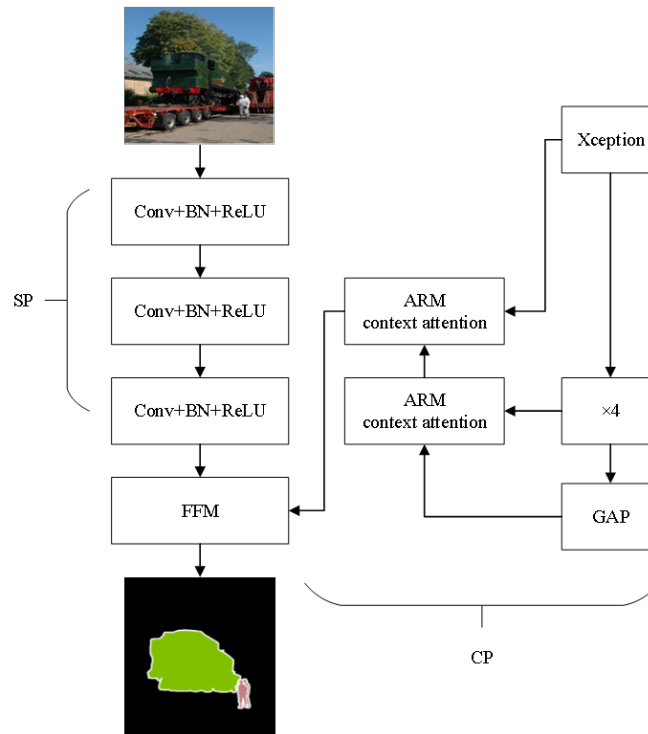


Fig. 2. Overall structure of proposed network model

The sub-network feature fusion module uses the output of the main Xception network as input to refine the features and further improve the network performance. Then, a Global Average Pooling (GAP) layer is added to the tail of the subnetwork feature fusion module to obtain larger receptive fields. Then, using the Content Attention Module (CAM), On the basis of obtaining the global context of the original image through global average pooling, CAM further calculates the attention vector to guide feature learning, and then up-sampling through bilinear interpolation is used to make the size of feature map and SP spline feature the same size. The Feature Fusion Module (FFM) is used to connect the output features of spatial path and context path. Then, the connected feature pool is converted into a feature vector and a weight vector is calculated by batch nor-

malization and balancing the scale of features. This weight vector can be re-weighted to achieve the feature output combining SP and CP.

3.1. Xception model

Xception model is a network based on deeply separable convolution, which is implemented by replacing Inception module with deeply separable convolution on the basis of Inception-v3. Xception model shows good image classification results in ImageNet [31], and the calculation speed is very fast. Chen et al. [32] proposed a encoder-decoder with detachability convolution for semantic image segmentation network. The encoder-decoder introduces Xception model to complete the task of semantic image segmentation, and improves the Xception model by combining TensorFlow deeply detachability convolution. A more dense feature graph is extracted by using depth-separable convolution instead of dilated convolution, and the structure of the Xception model is shown in figure 3. Where A, B, C denote entry flow, middle floe and exit flow respectively.

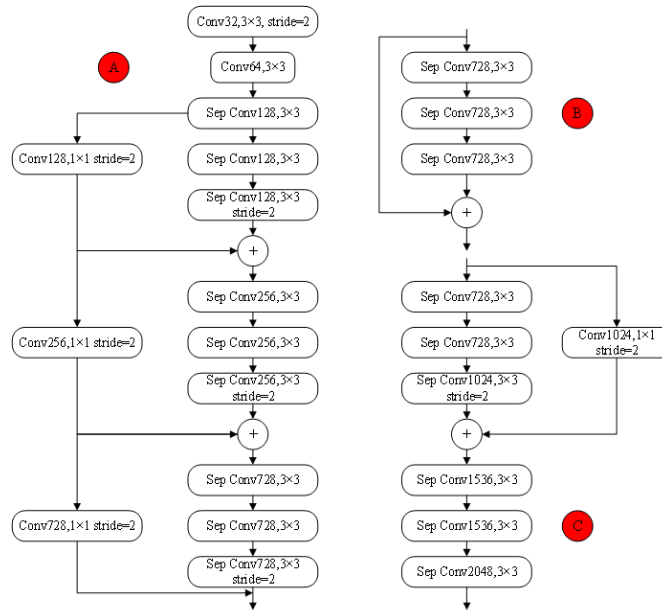


Fig. 3. Structure diagram of Xception module

On the basis of the original model, more layers are added and the network structure of inlet flow is not modified, which achieves fast computation and high storage efficiency. All maximum pooling operations are replaced by step-depth detachable convolution in order to extract feature images by using voidable convolution at arbitrary resolution. Additional batch normalization and ReLU activation are added after each 3×3 deep convolution.

3.2. Context attention model

In the practical application of image semantic segmentation, there is a large amount of data, so the calculation cost can be reduced by establishing a k-nearest neighbor graph $G = (V, E)$ to represent local areas [33]. $V = 1, 2, \dots, N$ is the point set, and $E \subseteq V \times \phi_i$ is the adjacent edge of the adjacent point pair. ϕ_i is the set of points in the neighborhood of point x_i . In order to prevent the point set from being affected by rotation transformation, the coordinate x_{ij} of the point in the local area is transformed into the relative coordinate of the central point x_i . The obtained edge characteristics are expressed as follows:

$$Fy_{ij} = (x_i, x_{ij} - x_i), x_i \in \mathbb{R}^F. \quad (1)$$

$$\forall x_{ij} \in Neighbors(x_i), x_i \in V, x_{ij} \in \phi_i. \quad (2)$$

In order to fully mine fine-grained details and multi-scale context information of images, context attention convolution layer is established based on Point Net. The encoding methods mainly include attention coding and context recurrent neural network coding. Attention encoding mainly learns fine-grained features in local regions. Context recurrent neural network coding learns multi-scale context geometry features between local regions. Where, multi-layer perceptron operation is represented by $MLP(*)$, and the number of convolution kernels is represented by $*$.

The attention encoding mechanism generally selects MLP first [34], and the output channel of the selected MLP is F1. Then, the selected MLP is used to map the original point features and edge features to the feature space with higher dimensions. The following is the specific representation.

$$u'_i = \sigma\Theta(\kappa(f_{F \times 1}(x_i))). \quad (3)$$

$$h'_i = \sigma\Theta(\kappa(f_{F \times 1}(y_i))). \quad (4)$$

Where, the nonlinear activation function is denoted as σ after parameterization. The set of parameters that can be learned in the convolution kernel is expressed as Θ . κ stands for batch normalization. f is convolution operation. The subscript $F \times 1$ is the size of the convolution kernel. In this experiment, the value of F1 is 16, that is, the number of feature channels is 16. MLP is used to process u'_i and h'_i , and the self-attention coefficient and neighborhood attention coefficient of x_i are respectively generated. By combining the two coefficients, the attention coefficient c_{ij} from the center point x_i to k neighboring points in the neighborhood can be obtained as:

$$c_{ij} = Selu(\sigma\Theta(\kappa(f_{1 \times 1}(u'_i)))) + \sigma\Theta(\kappa(f_{1 \times 1}(h'_i))). \quad (5)$$

Where, the nonlinear activation function is $Selu()$. Softmax function is used to normalize the attention coefficient, so that the convergence efficiency of the model is improved.

$$a_{ij} = exp(c_{ij}) / \sum_{j=1}^k exp(c_{ij}). \quad (6)$$

In order to mine fine-grained local features, attention coefficient a_{ij} is multiplied by local image feature h'_{ij} . At this moment of attention as feature selector, in describing the point x_i , on which the concentration coefficient can identify ability of neighborhood characteristics of adaptive capacity, to strengthen the neighborhood features such as noise, meaningless effectively suppressed, thus mining the fine-grained detail information fully and effectively.

By inputting the feature sequence $S_k = s_k^1, \dots, s_k^t, \dots, s_k^T$ of the sampling points into BiSeNet, the correlation between the sampling points in different scale neighborhoods is obtained. In order to fully mine context information, the hidden layer d is used to encode neighborhood feature vectors of different scales in sampling points successively. In addition, BiSeNet is used to encode neighborhood feature vectors of different scales of sampling point x_i , and the state of the hidden layer will be updated successively. Specific updates are as follows:

$$d_t = \zeta(d_{t-1}, s_{k-1}^t), t \in [1, T]. \quad (7)$$

In the equation, ζ is a nonlinear activation function; s_{k-1} is the $t - 1$ neighborhood feature vector, and d_{t-1} is the hidden layer state of s_{k-1} . Then the $t - th$ neighborhood feature vector in the sampling point is s_k . When BiSeNet is used to encode s_k , the corresponding output o_t is:

$$o_t = \omega_a d_t. \quad (8)$$

Where, the weight matrix that can be learned is ω_a . All feature sequences will get the hidden layer state after learning, and the hidden layer state is denoted as d_T . The multi-scale context geometric feature o_T of the sampling point can be obtained by multiplying ω_a and d_T .

The introduction of attention encoding is certainly helpful to improve the network's ability to capture fine-grained details in local areas to a certain extent. But it does not pay attention to the contextual geometry information between local areas. This is extremely important for image semantic segmentation [33]. The advantage of context BiSeNet encoding is that it can fully mine the high-level features of multi-scale context, which makes it possible to compensate each other for the fine-grained local features of a relatively low level and the multi-scale context geometric features of a relatively high level. By selecting Selu function, all fine-grained local features at different levels in the sampling point are fused into context geometric features. After the fusion of the two features, the sample point size of context fine-grained geometric features can be obtained as $N \times F2$. Before feature fusion, the $N \times 128$ image is sampled by interpolation operation on $R \times 128$ image. F_{\sum_i} after fusion is calculated as follows:

$$F_{\sum_i} = Selu(o_T + l_i). \quad (9)$$

where l_i is the local fine granularity feature.

3.3. Focal Loss

The data set is considered unbalanced, if the samples of a certain class of targets are greatly superior in number to those of other classes. This imbalance will lead to two problems. 1) Low training efficiency. Since most samples are simple targets, these samples

provide the model with less useful information during training. 2) The advantages of simple sample size will affect the training of the model and degrade the model performance. Guo et al. [34] proposed focal Loss function to solve the problem of category imbalance by reducing the internal weighting.

There are many kinds of target objects, the size and shape of objects of the same type are also different. It contains very few individual objects that stand out. CE loss function can not balance the learning of a small number of samples well, so focal Loss is introduced as a loss function to solve the sample imbalance problem in the segmentation task. Focal Loss is an improvement on the cross-drop function. By modifying the cross drop function and adding the sample difficulty weight adjustment factor $(1 - p_t)^\gamma$, the imbalance of sample categories and sample classification difficulty is alleviated and the model accuracy is improved. The mathematical expression is:

$$L_{FL}(p_t) = -(1 - p_t)^\gamma \log p_t. \quad (10)$$

We add a category weight α , and equation (7) is rewritten as:

$$L_{FL}(p_t) = -\alpha_t(1 - p_t)^\gamma \log p_t. \quad (11)$$

Where α is the weight parameter between categories. $(1 - p_t)^\gamma$ is the simple/difficult sample regulator. γ is the focusing parameter. When the prediction of a class is accurate, that is, when p_t is close to 1, the value of $(1 - p_t)^\gamma$ is close to 0. When the prediction of a category is inaccurate, that is, when p_t approaches 0, the value of $(1 - p_t)^\gamma$ approaches 1. Set $\gamma = 2$, $\alpha = 0.25$.

4. Experiments and Analysis

Experiments are carried out on PASCAL VOC 2012 benchmark data set [35]. The dataset is published by the International Computer Vision Challenge for image classification, detection or semantic segmentation. It contains 20 foreground object classes and one background class, including people, animals, traffic vehicles and indoor household items. There are 1464 images for training set, 1449 images for validation set and 1456 images for testing set. The experiment is implemented on TensorFlow, a deep learning framework. The operating system used in the experiment is Windows 11, and the graphics card is NVIDIA RTX3060. A dense feature graphs are extracted using pre-trained Xception by ImageNet. Adam optimizer and Poly learning strategy are adopted. In the experiment, the image is cut to 256×256 for training. In the initial training process, a small learning rate is used to achieve smooth start. Set the initial learning rate as 1×10^{-4} , momentum as 0.9, and select iteration training as 50000 times.

4.1. Evaluation index

Mean intersection over Union (MIOU), pixel accuracy (PA) and mean pixel accuracy (MPA) are used to evaluate the segmentation effect of the proposed method on the data set. The higher values of MIOU PA, and MAP denote the better image semantic segmentation effect.

Given that there are $k + 1$ segmentation classes in the image (including k target classes and 1 background class). p_{ij} (False Positives) represents the number of pixels that belong to class i but are predicted to be class j . p_{ji} (False Negatives) represents that the number of pixels that belong to class j but are predicted to be class i . p_{ii} (True Positives) indicates the true number of pixels.

Pixel accuracy (PA) is defined as follows:

$$PA = \frac{\sum_i^k p_{ii}}{\sum_{i=0}^k \sum_{j=0}^k p_{ij}}. \quad (12)$$

Mean pixel accuracy (MPA) is defined as:

$$MPA = \frac{1}{k + 1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij}}. \quad (13)$$

Mean intersection over Union (MIOU) is defined as:

$$MIOU = \frac{1}{k + 1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k (p_{ji} - p_{ii})}. \quad (14)$$

Where formula (12) represents the proportion of correctly classified pixel points and all pixel points in the image. Formula (13) represents the proportion of correctly classified pixel points of each category and all pixel points of that category, and then calculates the average value. Formula (14) represents the intersection of the predicted region and the actual region in the image divided by the union of the predicted region and the actual region.

4.2. Results and Analysis

During the experiment process, a context attention mechanism is introduced to improve the accuracy of feature images, filter background information and reduce the loss of details. In the context attention module, two convolution levels $1 \times k$ and $k \times 1$ are applied to the high-level features to obtain spatial concerns. The segmentation results of different k values are shown in Table 1. As can be seen from Table 1, different convolution operations on feature graphs result in significantly different segmentation results. When $k = 8$, MIOU and MPA are the highest. Therefore, the training model with $k = 8$ is finally selected for verification.

Then, we conduct experiments on the PASCAL VOC 2012 dataset. Compared with different image semantic segmentation algorithms including PSPNet [36], DANet [37], DeepLabV3+ [38], SANet [39], DUpsampling and reference [40-42]). The experimental results are shown in Table 2. As can be seen from Table 2, the MIOU value of the proposed algorithm is 5.88% higher than PSPNet algorithm, 3.48% higher than DANet algorithm, 3.30% higher than DeepLabV3+ algorithm, and 1.24% higher than SANet algorithm. Compared with the , it is improved by 2.41%, 2.18%, 1.50% and 1.04% compared with DUpsampling algorithm, reference [40], reference [41], and reference [42] respectively. Also, in terms of the AP and MAP, the proposed method obtains the better results.

Table 1. MIOU and MPA values with different k

k	MIOU%	MPA%
1	79.18	82.54
2	80.22	82.67
3	81.74	83.54
4	82.06	83.96
5	82.78	84.63
6	83.47	85.21
7	84.73	85.88
8	88.91	89.25
9	86.46	87.25
10	85.41	86.93

Table 2. Comparison results of different image semantic segmentation algorithms

Method	AP%	MAP%	MIOU%
PSPNet	89.21	90.33	79.67
DANet	90.42	91.25	82.07
DeepLabV3+	90.88	91.35	82.25
DUpsamling	91.26	91.87	83.14
Reference [40]	91.78	92.06	83.37
Reference [41]	92.34	92.58	84.05
SANet	93.54	93.88	84.31
Reference [42]	93.89	94.12	84.51
Proposed	94.56	95.71	85.55

Table 3. Comparison between BiSeNet and proposed method

Method	MPA%	MIOU%
BiSeNet	90.23	82.25
Proposed	93.61	85.55

In order to better verify the performance of the proposed algorithm, we make comparison between BiSeNet and proposed method, as shown in Table 3. As can be seen from Table 3, the MIOU value of the proposed algorithm is improved, also the MPA is improves by 3.38% compared with BiSeNet.

Based on BiSeNet algorithm, ablation experiments are performed to verify the better results of the proposed method. The image semantic segmentation results of different combination methods are shown in Table 4.

Table 4. Ablation experiments

Number	BiSeNet	Xception	Context attention	Focal loss	MIOU%	MAP%
a	Yes	No	No	No	82.25	89.31
b	Yes	Yes	No	No	82.67	90.24
c	Yes	No	Yes	No	82.29	89.57
d	Yes	Yes	Yes	No	85.03	92.45
e	Yes	Yes	Yes	Yes	85.55	95.71

From the comparison of results a and b in Table 4, it can be seen that the MIOU value increases by 0.42% by fine-tuning the Xception model and adding a low-level feature extraction path. The comparison between the results of a and c shows that the MIOU value increases by 2.16% when the attention mechanism is introduced on the basis of the original network, indicating that the accuracy of feature extraction is effectively improved by the attention mechanism. The comparison of c and d results shows that multi-path extraction of low-level features can increase MIOU value by another 0.62%. According to the comparison of results of d and e, the use of Focal Loss improves MIOU value by 0.52%.

The training loss curves of BiSeNet algorithm and the proposed algorithm are shown in figure 4. The x-axis is training time and the y-axis is loss value. As can be seen from figure 4, the loss value is relatively high at the beginning of the model training. With the increase of training times, the loss curve gradually stabilizes. Compared with the original algorithm, the loss value of the proposed algorithm decreases greatly, indicating that the proposed algorithm can effectively reduce the loss of feature information and improve the final segmentation effect.

Different algorithms are used to segment the test set in PASCAL VOC 2012 dataset. The experimental comparison results are shown in figure 5. As can be seen from the figure, compared with reference[40] algorithm and reference [41] algorithm, the proposed algorithm has clearer target segmentation boundary. Reference [42] algorithm obviously has the problem of unbalanced bicycle segmentation. Compared with the Reference [42] algorithm, the segmentation results of the proposed algorithm are significantly more balanced. Experimental results show that the proposed algorithm has a significant improvement in the boundary segmentation effect of the target background, refines the target boundary, improves the segmentation effect of the target object, and has a better object resolution ability.

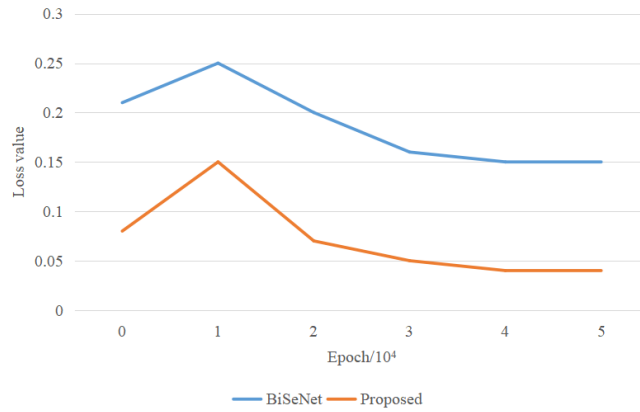


Fig. 4. Loss values with different methods

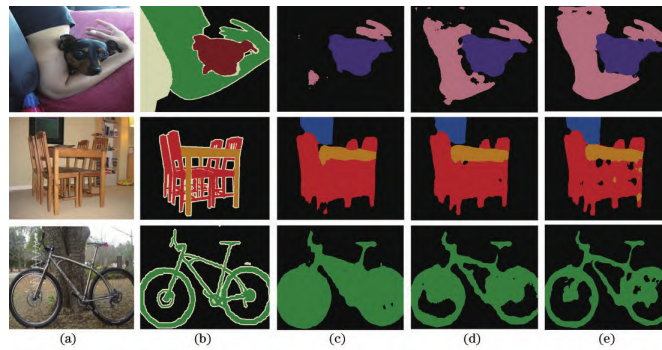


Fig. 5. Comparison of segmentation results. (a) Original images; (b) Ground Truth; (c) reference [41]; (d) reference [42]; (e) Proposed algorithm

5. Conclusion

In this paper, an image semantic segmentation algorithm based on BiSeNet and context attention mechanism is proposed. A low-level feature extraction path is added in BiSeNet to increase feature information and expand receptive fields. Without affecting the network speed, it improves the segmentation accuracy. Context content attention mechanism is introduced to extract high-level features and low-level features, and the two features are fused to obtain rich context information effectively, filter background information, and obtain more detailed feature maps. In order to solve the problem that the sample number of different categories is not balanced and the difficulty of distinguishing different objects is different, the focus loss function is used instead of the cross drop loss function to reduce the loss of feature details. Experiments are carried out on PASCALVOC 2012 data set, and the experimental results show that the proposed algorithm has a great improvement in image segmentation accuracy compared with other algorithms. In the future, we will apply image semantic segmentation to a wider range of fields, such as aerospace, remote sensing, medicine. At the same time, we will also develop more advanced methods to further improve accuracy.

References

1. Zhang G, Zhao K, Hong Y, et al. "SHA-MTL: soft and hard attention multi-task learning for automated breast cancer ultrasound image segmentation and classification," *International Journal of Computer Assisted Radiology and Surgery*, vol. 16, pp. 1719-1725, (2021).
2. H. Zhang et al. "Multiscale Visual-Attribute Co-Attention for Zero-Shot Image Recognition," *IEEE Transactions on Neural Networks and Learning Systems*, (2021). doi: 10.1109/TNNLS.2021.3132366.
3. X. Lei and H. Ouyang. "Kernel-Based Intuitionistic Fuzzy Clustering Image Segmentation Based on Grey Wolf Optimizer With Differential Mutation," *IEEE Access*, vol. 9, pp. 85455-85463, (2021).
4. Fan Wang, Chen Chen, Haitao Zhang and Youhua Ma. "Short-term Load Forecasting Based On Variational Mode Decomposition And Chaotic Grey Wolf Optimization Improved Random Forest Algorithm," *Journal of Applied Science and Engineering*, Vol. 26, No. 1, pp. 69-78, (2020).
5. Fung D, Liu Q, Zammit J, et al. "Self-supervised deep learning model for COVID-19 lung CT image segmentation highlighting putative causal relationship among age, underlying disease and COVID-19," *Journal of Translational Medicine*, vol. 19, no. 1, (2021).
6. Xian S, Cheng Y, Chen K. "A novel weighted spatial T-spherical fuzzy C-means algorithms with bias correction for image segmentation," *International Journal of Intelligent Systems*, vol. 37, no. 2, (2022)
7. Zhang L, X Hu, Zhou Y, et al. "Memristive DeepLab: A hardware friendly deep CNN for semantic segmentation," *Neurocomputing*, vol. 451, pp. 181-191 (2021).
8. H. -Y. Han, Y. -C. Chen, P. -Y. Hsiao and L. -C. Fu. "Using Channel-Wise Attention for Deep CNN Based Real-Time Semantic Segmentation With Class-Aware Edge Information," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 2, pp. 1041-1051, (2021).
9. Jisi A and Shoulin Yin. "A New Feature Fusion Network for Student Behavior Recognition in Education," *Journal of Applied Science and Engineering*, vol. 24, no. 2, pp. 133-140. (2021)
10. Russo G. "On Unsupervised Methods for Medical Image Segmentation: Investigating Classic Approaches in Breast Cancer DCE-MRI," *Applied Sciences*, vol. 12, no. 1. (2022)

11. Gurita A, Mocanu I G. "Image Segmentation Using Encoder-Decoder with Deformable Convolutions," *Sensors*, vol. 21, no. 5, pp. 1570. (2021)
12. C. Lyu, G. Hu and D. Wang. "HRED-Net: High-Resolution Encoder-Decoder Network for Fine-Grained Image Segmentation," *IEEE Access*, vol. 8, pp. 38210-38220, (2020)
13. Z. Tian, T. He, C. Shen and Y. Yan. "Decoders Matter for Semantic Segmentation: Data-Dependent Decoding Enables Flexible Feature Aggregation," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3121-3130, (2019). doi: 10.1109/CVPR.2019.00324.
14. Cai W, Zhai B, Liu Y, et al. "Quadratic Polynomial Guided Fuzzy C-means and Dual Attention Mechanism for Medical Image Segmentation," *Displays*, vol. 70, no. 6, pp. 102106. (2021)
15. A. Bera, Z. Wharton, Y. Liu, N. Bessis and A. Behera. "Attend and Guide (AG-Net): A Keypoints-Driven Attention-Based Deep Network for Image Recognition," *IEEE Transactions on Image Processing*, vol. 30, pp. 3691-3704, (2021).
16. Yang T, Yoshimura Y, Morita A, et al. "Pyramid Predictive Attention Network for Medical Image Segmentation," *IEICE Transactions on Fundamentals of Electronics Communications and Computer Sciences*, vol. E102, no. A(9), pp. 1225-1234. (2019)
17. Al-Huda Z, Zhai D, Yang Y, et al. "Optimal Scale of Hierarchical Image Segmentation with Scribbles Guidance for Weakly Supervised Semantic Segmentation," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 35, no. 10, (2021)
18. Lsel P D, Kamp T, Jayme A, et al. "Introducing Biomedisa as an open-source online platform for biomedical image segmentation," *Nature Communications*, 11(5577). (2020)
19. Guo H, Yang D. "PRDNet: Medical image segmentation based on parallel residual and dilated network," *Measurement*, vol. 173, no. 1, pp. 108661. (2020)
20. Huang M, Huang S, Zhang Y, et al. "Medical Image Segmentation Using Deep learning with Feature Enhancement," *IET Image Processing*, vol. 14, no. 5. (2020)
21. Olimov B, Sanjar K, Din S, et al. "FU-Net: fast biomedical image segmentation model based on bottleneck convolution layers," *Multimedia Systems*, vol. 27, no. 4, pp. 637-650, 2021.
22. Zheng T, Duan Z, Wang J, et al. "Research on Distance Transform and Neural Network Lidar Information Sampling Classification-Based Semantic Segmentation of 2D Indoor Room Maps," *Sensors*, vol. 21, no. 4, pp. 1365. (2021)
23. Shoulin Yin, Hang Li, Desheng Liu and Shahid Karim. "Active Contour Modal Based on Density-oriented BIRCH Clustering Method for Medical Image Segmentation," *Multimedia Tools and Applications*, Vol. 79, pp. 31049-31068, (2020).
24. Wech T, Ankenbrand M J, Bley T A, et al. "A data-driven semantic segmentation model for direct cardiac functional analysis based on undersampled radial MR cine series," *Magnetic Resonance in Medicine*, vol. 87. (2022)
25. Jiang, D., Li, H., Yin, S. "Speech Emotion Recognition Method Based on Improved Long Short-term Memory Networks," *International Journal of Electronics and Information Engineering*, Vol. 12, No. 4, pp. 147-154. (2020)
26. Jiang M, Zhai F, Kong J. "Sparse Attention Module for optimizing semantic segmentation performance combined with a multi-task feature extraction network," *The Visual Computer*, vol. 12. (2021)
27. R. Yi, Y. Huang, Q. Guan, M. Pu and R. Zhang. "Learning From Pixel-Level Label Noise: A New Perspective for Semi-Supervised Semantic Segmentation," *IEEE Transactions on Image Processing*, vol. 31, pp. 623-635, (2022).
28. Grant-Jacob J A, Praeger M, Eason R W, et al. "Semantic segmentation of pollen grain images generated from scattering patterns via deep learning," *Journal of Physics Communications*, vol. 5, no. 5, 055017 (11pp). (2021)
29. Yu C, Wang J, Peng C, et al. "BiSeNet: Bilateral Segmentation Network for Real-time Semantic Segmentation," *ECCV 2018. Lecture Notes in Computer Science*, vol. 11217, pp. 334-349. Springer, Cham. (2018).

30. Polat Z. "Detection of Covid-19 from Chest CT Images using Xception Architecture: A Deep Transfer Learning based Approach," *Sakarya University Journal of Science*, vol. 25, no. 3, pp. 813-823, (2021)
31. Xiaowei Wang, Shoulin Yin, Ke Sun, et al. "GKFC-CNN: Modified Gaussian Kernel Fuzzy C-means and Convolutional Neural Network for Apple Segmentation and Recognition," *Journal of Applied Science and Engineering*, vol. 23, no. 3, pp. 555-561, (2020).
32. George B, Assaiya A, Roy R J, et al. "CASSPER is a semantic segmentation-based particle picking algorithm for single-particle cryo-electron microscopy," *Communications Biology*, vol. 4, no. 1. (2021)
33. Dai, Y., Xu, B., Yan, S., Xu, J.: Study of cardiac arrhythmia classification based on convolutional neural network. *Computer Science and Information Systems*, Vol. 17, No. 2, 445-458. (2020), <https://doi.org/10.2298/CSIS191229011D>
34. Ge, Y., Zhu, F., Huang, W., Zhao, P., Liu, Q.: Multi-Agent Cooperation Q-Learning Algorithm Based on Constrained Markov Game. *Computer Science and Information Systems*, Vol. 17, No. 2, pp. 647-664. (2020), <https://doi.org/10.2298/CSIS191220009G>
35. Wong C C, Yeh L Y, Liu C C, et al. "Manipulation Planning for Object Re-Orientation Based on Semantic Segmentation Keypoint Detection," *Sensors*, vol. 21, no. 7, 2280. (2021)
36. Guo X, Xiao R, Lu Y, et al. "Cerebrovascular Segmentation from TOF-MRA based on Multiple-U-net with Focal Loss Function," *Computer Methods and Programs in Biomedicine*, vol. 202, no. 3, pp. 105998. (2021)
37. Liu R, He D. "Semantic Segmentation Based on Deeplabv3+ and Attention Mechanism," *2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*. IEEE, (2021).
38. H. Zhao, J. Shi, X. Qi, X. Wang and J. Jia. "Pyramid Scene Parsing Network," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 6230-6239, doi: 10.1109/CVPR.2017.660.
39. J. Fu et al. "Dual Attention Network for Scene Segmentation," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 3141-3149, doi: 10.1109/CVPR.2019.00326.
40. Chen LC., Zhu Y., Papandreou G., Schroff F., Adam H. "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," *ECCV 2018. Lecture Notes in Computer Science, vol 11211*. Springer, Cham. (2018)
41. Z. Zhong et al. "Squeeze-and-Attention Networks for Semantic Segmentation," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13062-13071, (2020). doi:10.1109/CVPR42600.2020.01308.
42. Li X, Chen J, Ye Y, et al. "Fast Semantic Segmentation Model PULNet and Lawn Boundary Detection Method," *Journal of Physics: Conference Series*, vol. 1828, no. 1, pp. 012036 (16pp). (2021)
43. Trajanovski S, Shan C, Weijtmans P, et al." Tongue Tumor Detection in Hyperspectral Images Using Deep Learning Semantic Segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 4, pp. 1330-1340. (2021)
44. Wang K, Xiang K, Yang K. "Polarization-driven Semantic Segmentation via Efficient Attention-bridged Fusion," *Optics Express*, vol. 29, no. 4. (2021)

Lin Teng is a doctoral student at the School of Information and Communication Engineering, Harbin Engineering University. Her research interests include image processing, image segmentation.

Yulong Qiao is a professor at the School of Information and Communication Engineering, Harbin Engineering University. His main research areas: statistical image processing, image/video processing and applications.

Received: March 21, 2022; Accepted: September 10, 2022.

DRN-SEAM: A Deep Residual Network Based on Squeeze-and-Excitation Attention Mechanism for Motion Recognition in Education

Xinxiang Hua

College of Marxism, Zhengzhou University of Science and Technology
Zhengzhou, 450015 China
zxcvfdsa5024@foxmail.com

Abstract. In order to solve the shortcomings of the traditional motion recognition methods and obtain better motion recognition effect in education, this paper proposes a residual network based on Squeeze-and-Excitation attention mechanism. Deep residual network is widely used in various fields due to the high recognition accuracy. In this paper, the convolution layer, adjustment batch normalization layer and activation function layer in the deep residual network model are modified. Squeeze-and-Excitation (SE) attention mechanism is introduced to adjust the structure of network convolution kernel. This operation enhances the feature extraction ability of the new network model. Finally, the expansibility experiments are conducted on WISDM(Wireless Sensor Data Mining), and UCI(UC Irvine) data sets. In terms of F1, the value exceeds 90%. The results show that the proposed model is more accurate than other state-of-the-art posture recognition models. The proposed method can obtain the ideal motion recognition results.

Keywords: motion recognition; Deep residual network; Squeeze-and-Excitation; attention mechanism, education.

1. Introduction

A large number of education videos have been collected in the process of education training and teaching. Accurate recognition of education motions in the videos can prevent accidental injuries and protect the health of students. Therefore, it is of great significance to construct an excellent motion recognition method [1-3].

At present, a variety of portable devices have been developed rapidly, such as smart bracelets and smart-phones, etc.. These emerging adaptive mobile applications can use the collected big data by embedded sensors to conduct motion recognition and behavior analysis. For example, the medical system can use motion recognition to effectively monitor the movement behavior. This not only guides medical staffs to perform the correct treatment, but also solves the shortage of hospital staff. In rehabilitation training, motion recognition and behavior analysis can assist patients in rehabilitation activities, analyze patients' movements and behaviors, and monitor the elderly for ensuring safety. In education, behavior recognition technology as an auxiliary means can be used to analyze various data of students. Effective data analysis improves the score of students, thus improving the overall competitive level [4-6].

Related researches have been striving to achieve the following two goals: enhancing the recognition accuracy and reducing the reliance on engineering features. However, they are difficult to achieve, because the difficulty of motion recognition lies in the great variety of specific motion patterns. In other words, the motion trajectory pattern of different individuals completing the same action is not exactly the same, so it is difficult for engineering features to fully and correctly express the motion trajectory pattern in each action process resulting in the unsatisfactory recognition results.

In this paper, the convolution layer, adjustment batch normalization layer and activation function layer are improved to build a new deep residual network model. Meanwhile, Squeeze-and-Excitation (SE) attention mechanism is introduced to adjust the structure of network convolution kernel. The structure of this paper is as follows. Section 2 introduces the related works. In section 3, the proposed motion recognition method is displayed. Experiments and analysis are shown in section 4. There is a conclusion in section 5.

2. Related Works

This section briefly overviews previous studies on posture recognition based on deep learning, then introduces CNN applications in image classification. In order to achieve the above two goals, some researchers propose the feature extraction method based on deep learning.

Deep learning is organized hierarchically with the latter layer processing the output of the former layer [7,8]. It is a neural network that uses multiple nonlinear information processing layers to extract and classify features. Two classical network structures commonly used in deep learning are convolutional neural network (CNN) and recurrent neural network (RNN). CNN is a deep neural network (DNN) with feature extraction capability. It stacks several convolution operations to create a feature map that becomes progressively more abstract. It automatically extracts the valid features from raw data, and no longer depends on prior knowledge. This method can not only enhance the accuracy and generalization ability of the model, but also form an end-to-end model. It reduces the complexity of the model training. RNN can process serial information [9-11]. The LSTM (long short term memory) [12] and the other extended forms of RNN contain a memory module which can simulate the time dependence in a time series, this way can better handle the time sequence dependence information.

Motion recognition can be generally divided into the following two main steps. The first step is the segmentation of time series. The current mainstream method is to use a fixed length sliding window to segment the entire motion time series into equal segments [13-15]. For example, most studies adopt the time series segmentation method on WISDM data set, that is, fixed window length and fixed moving direction. The second step is to extract the valid features from the obtained original segment. Feature extraction is the most critical part in the whole project, which will directly affect the overall recognition accuracy of the model. In the previous recognition algorithms, engineering features are often used [16-19]. Although this method can also show excellent performance, it requires rich domain knowledge. It can also be time-consuming for domain experts to find good engineering features. The common engineering features in relevant studies include spectral entropy, autoregression coefficient and fast Fourier transform coefficient.

The classical solutions are template matching method, hidden Markov model and support vector machine, etc.,

Many researchers combine the deep learning method with motion classification. Qamar [20] trained deformation postures by combining R-FCN with HyperNet network. Ahmad [21] designed a new convolutional neural network, which could effectively complete the classification of motion images. Lin [22] proposed a novel matching R-CNN framework based on mask R-CNN to perfect motion detection, posture estimation, segmentation and retrieval. Yan [23] adopted a sparse algorithm to extract the spatial and temporal features of sports motion, and then used the neural network to establish sports motion recognition model. Wen [24] extracted the energy diagram and motion descriptor of sports movements, and established the sports movement recognition model by using the support vector machine. The above researches all use a deep convolutional neural network to recognize and classify sports images. In order to improve the recognition accuracy, the convolutional layer number in CNN is usually modified to improve the recognition and classification performance of the model.

However, the deep convolutional neural network still has some problems:

1. with the deepening of the deep learning network, the stacked effect of the network is not good;
2. If the network is more complex, it will result in some problems such as gradient dispersion or gradient explosion in the training process.

This paper draws on the advantages of ResNet in solving the gradient dispersion problem in deep network training and proposes a new deep residual network to improve the performance of action recognition and classification. The deep residual network is composed of residual blocks as shown in figure 1.

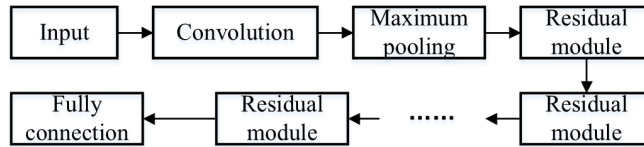


Fig. 1. The deep residual network structure

Each residual block can be expressed as:

$$y_i = h(x_i) + F(x_i, w_i). \tag{1}$$

$$x_{i+1} = f(y_i). \tag{2}$$

Where F is the residual function, f is the ReLU function. W_i is the weight matrix. x_i and y_i are the input and output of the $i - th$ layer, respectively.

The definition of residual function F is:

$$F(x_i, w_i) = w_i \cdot \sigma(B(w'_i) \cdot (B(x_i))). \tag{3}$$

$B(x_i)$ is batch normalization. (\cdot) denotes convolution. $\sigma(x) = \max(x, 0)$.

The basic idea of residual learning is a branch of the gradient propagation path. For CNN, this idea is first introduced into the Inception model with a parallel form. Residual networks share some similarities with Highway Network. It is connected through residual blocks and shortcuts. The gradient loss problem associated with increasing layers in ResNet is mitigated. However, the output of each path in the Highway Network is controlled by the gate function learned in the training stage.

Unlike the convolutional layer in traditional CNN, the residual units in ResNet are not stacked. Instead, there are some shortcut connections from the input to the output in each convolutional layer. Using identity mapping as shortcut connection reduces the complexity of residual networks and makes deep networks be quickly trained. A residual network can be thought of as a collection of many paths, rather than a very deep architecture. However, all network paths in the residual network have different lengths. There is only one path through all the residual units. In addition, all of these signal paths do not propagate gradients, which is why the residual network is optimized and trained faster. With the network layer increasing, the accuracy rate does not decrease. The structure of the residual block is shown in figure 2.

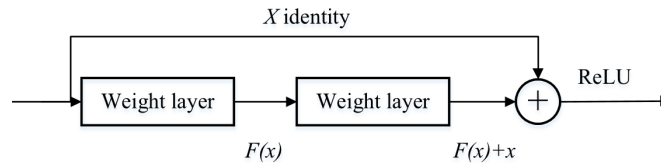


Fig. 2. Residual block structure

Most state-of-the-art methods for action recognition rely on a two-stream architecture that processes appearance and action independently. Aiming at the action features in the context of complex surveillance, and color, texture, edge and other features are extracted manually. The feature process is cumbersome and the classification accuracy is low, so this paper proposes a residual network based on Squeeze-and-Excitation attention mechanism. The recognition performance of the model with shallow network layers is not ideal, while the residual block is introduced into the residual network. When the layer of the model is increased, the residual block can solve the degradation problem well. Therefore, we focus on these problems and propose a deep residual network based on Squeeze-and-Excitation attention mechanism for posture recognition in basketball motion.

The rest of this paper is organized as follows. In Section 2, the proposed posture recognition method is introduced. Section 3 elaborates the detailed experiment process. Finally, a summary alongside with the future research direction is provided in Section 4.

3. Proposed DRN-SEAM Model

At present, most researchers select CNN-based methods to extract the features of motion images [25]. However, if there are many sports postures, CNN network layers are

relatively few, which directly affects the feature learning ability of the network. Then, researchers improve the network by increasing the layer number of the deep convolutional neural network such as GoogleNet and ResNet. Generally, if the layer is deeper in the image recognition and classification model, then the model recognition performance will be better [26]. The flow chart of the proposed method in this paper is shown in figure 3.

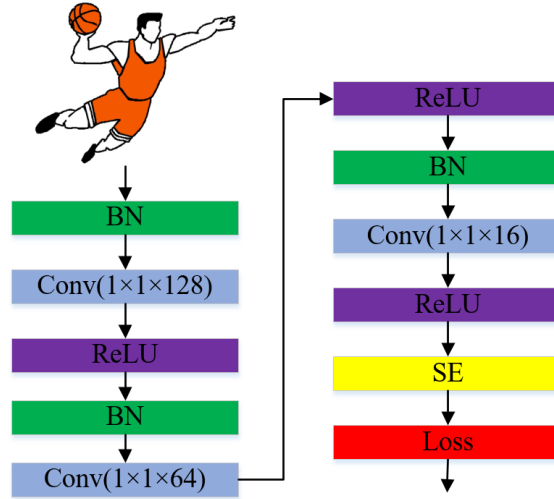


Fig. 3. Flow chart of proposed method

3.1. Modified residual neural network

During the deep convolutional neural network training, the weight of a certain layer changes, and the output feature map of that layer also changes accordingly. The weight of the next layer needs to be relearned, and the network weight of each subsequent layer will be affected. Adding activation function in ResNet can improve the non-linear capability of network model construction. ReLU is used as the activation function in the deep residual network. When $x > 0$, the gradient of ReLU function is always 1, the gradient is not attenuated, which alleviates the problem of gradient dispersion. Assuming that no activation function is used, the network can only be constructed through linear mapping. Even if there are many convolutional layer networks, each layer in the whole network is equivalent, but the convolutional feature map will not change much. So the network should use an activation function.

With the increase of the convolutional neural network layers, the convergence speed of the network will drop sharply and the gradient dispersion will appear in the training process. The Batch Normalization (BN) is an effective solution to this problem, the detailed explanation is shown in reference [27,28]. The specific solution is to normalize the input signals at the same layer. The formula is as follows:

$$\hat{x} = \frac{X - E(x)}{\sqrt{Var(x) + \varepsilon}}. \tag{4}$$

Where, \hat{x} is the activation value of the network normalization. x is the activation value of a certain layer in the network. $E(x)$ is the average value. $Var(x)$ is the variance, and ε is the minimum value. The BN algorithm formula is as follows:

$$y^k = r^k \hat{x}^k + \beta^k. \tag{5}$$

Where, each neuron x^k has the parameters r and β . In this way, when $r^k = \sqrt{Var[x^k]}$, $\beta^k = E[x^k]$, the original learning features in a certain layer can be maintained. The parameters r and β can be reconstructed to restore the initial network learning feature distribution. BN layer is a normalized neural network activation method, it adds batch normalization algorithm to normalize the input signal of each layer, stabilizes its data distribution, and sets a higher learning rate during training to make the network convergence speed and training speed faster. Figure 4 shows the sequence of "convolutional layer+BN layer+ReLU layer" in the traditional residual network.

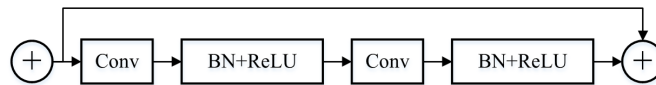


Fig. 4. The sequence of traditional residual blocks

The sequence of traditional residual blocks is defective in deep convolution ResNet, for example, the input of identical blocks is transferred to the deep network from two paths.

The right path indicates that the feature map goes through the convolutional layer and then to BN and ReLU. The input feature map has not been normalized first, so the existence of BN layer is not meaningful. According to the above defects, a new nonlinear branch "BN layer+convolutional layer+ReLU layer" is adopted in this paper to arrange the identical block structure. As shown in figure 5, the network structure is still the same as the traditional residual block network structure (figure 3), while the typical residual block in ResNet is composed of three convolutional layers. In this paper, the new residual block arrangement method not only preserves the identity mapping of the left path, but also maintains the learning ability of the right nonlinear network path.

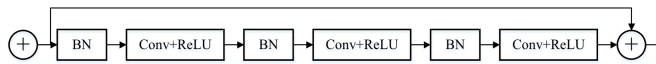


Fig. 5. The sequence of modified residual blocks

3.2. Squeeze-and-Excitation attention mechanism

The attention mechanism is a weighted change for object data that uses top information to guide the bottom-up feed-forward process. The attention model of the human brain is a resource allocation model. At any given moment, attention is always focused on one focal point of the image, while the rest is invisible. In recent years, many attempts have been made to apply attention to deep neural networks. Therefore, the attention mechanism is further located for the discriminative site features. The Squeeze-and-Excitation (SE) algorithm [26] is used for the improved network. The performance of the network model is improved due to the precise modeling of the interaction between the channels for the convolution feature. A mechanism for network models to calibrate features enables networks to selectively enhance valuable feature channels and inhibit useless feature channels from the perspective of global information.

The SE Network modules are shown in figure 6. To ensure the sensitivity of valuable information after adding the network, and make the valuable features be effectively used in the subsequent network layer, accurate modeling of the dependency relationship between channels can be achieved, the features are redefined using the Squeeze, Excitation and Reweight:

1. Squeeze(global information embedding). In order to solve the problem of channel dependence, spatial dimension compression features are used. Each two-dimensional eigenvector is represented by a variable with global spatial information, and the output dimension matches the input channel number. The formula is expressed as follows:

$$Z_c = F_{sq}(U_c) = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H u_c(i, j). \quad (6)$$

2. Excitation (The adaptive recalibration). In order to utilize the information gathered in the compression operation, the dependency of the channel is captured comprehensively. The nonlinear interaction and non-exclusive relationship between learning channels must be satisfied.

$$s = F_{ex}(z, w) = \sigma(g(z, w)) = \sigma(w_2 \delta(W_1 z)). \quad (7)$$

3. Reweight. The weights generated after the Excitation are multiplied by the original features.

$$\tilde{X} = F_{scale}(u_c, S_c) = S_c \cdot U_c. \quad (8)$$

After the structures "BN layer+convolutional layer+ReLU layer", the SE algorithm is introduced to the new residual block. The dynamic features of the network are recalibrated to improve the performance of the network, and the soft attention mechanism is successfully applied to the deep network. Embedding the SE module into the new residual module is as shown in figure 7. Table 1 shows the main structure of the original ResNet50 network and the modified structure.

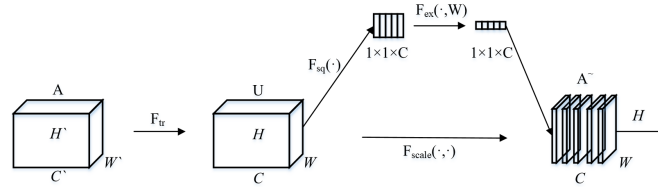


Fig. 6. SE model

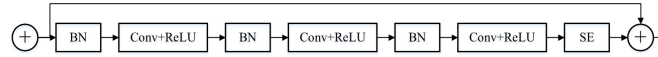


Fig. 7. The residual block with SE module

4. Experiment and Analysis

To verify the performance of the proposed model (DRN-SEAM), we conduct experiments on UCI, WISDM data set and real basketball motion images. Experimental environment is GPU GTX1080, Memory 16 GB, Windows 10 system, MATLAB7a, Tensroflow. It uses stochastic gradient descent (SGD) to optimize the network parameters. The loss function adopts mutual entropy loss. In order to improve the efficiency, the model training is divided into 200 mini-batches. The learning rate is set as 0.0001. All models are trained with 1000 epochs.

4.1. Performance evaluation indexes

In this paper, Precision (P), Recall (R), F1 Measure, average accuracy are used to evaluate the performance.

$$Precision(i) = T(i)/(T(i) + F(i)). \tag{9}$$

$$Recall(i) = T(i)/D(i). \tag{10}$$

$T(i)$ represents the number of i-th correct recognized motions. $F(i)$ represents the number of i-th incorrect recognized motions. $D(i)$ represents the number of i-th motion samples. F value is the weighted average between accuracy and recall, as shown in equation (11).

Table 1. Comparison of two network convolution structures

ResNet50	New network structure
$[1 \times 1, 64; 3 \times 3, 64; 1 \times 1, 256] \times 3$	$[1 \times 1, 128; 3 \times 3, 128; 1 \times 1, 128] \times 3$
$[1 \times 1, 128; 1 \times 1, 128; 1 \times 1, 512] \times 4$	$[1 \times 1, 256; 1 \times 1, 256; 1 \times 1, 256] \times 4$
$[1 \times 1, 256; 1 \times 1, 256; 1 \times 1, 1024] \times 6$	$[1 \times 1, 512; 1 \times 1, 512; 1 \times 1, 512] \times 6$
$[1 \times 1, 512; 1 \times 1, 512; 1 \times 1, 2048] \times 3$	$[1 \times 1, 1024; 1 \times 1, 1024; 1 \times 1, 1024] \times 3$

$$F = \frac{(\vartheta^2 + 1) \cdot P \cdot R}{\vartheta^2(P + R)}. \quad (11)$$

When $\vartheta = 1$, $F = F1$, namely,

$$F1 = \frac{2 \cdot P \cdot R}{P + R}. \quad (12)$$

Average accuracy=correct recognized motions/the total number of samples.

4.2. Datasets description and the compared methods

The experiment datasets in the experiment are WISDM, UCI and real basketball data sets. The WISDM data contains Walking, Jogging, Upstairs, Downstairs, Sitting, Standing [30]. We only select 4526 samples in this datasets, and randomly select 80% of the dataset for training, and retained 20% as the test set. The UCI dataset includes a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities (Walking, Walking_upstairs, Walking_downstairs, Sitting, Standing, Laying). The obtained dataset has been randomly partitioned into two sets, where 80% of the volunteers was selected for generating the training data and 20% the test data [31]. The real basketball data is composed of ten football students, each student conducts 4 actions (Jumping, Shooting, Falling back, Defending) with a total of 40 samples. Then, the data sets are enhanced through rotation, translation, scaling. Finally, we obtain a total of 1000 samples. They are collected by professional PE student belonging to in-house data with mobile phone. Similarly, we randomly select 80% of the dataset for training, and retained 20% as the test set. Some sample images from UCI and real basketball data are shown in figure 8 and figure 9.



Fig. 8. The sample images in dataset

We compare the DRN-SEAM model with one classical algorithm ResNet50 and other three state-of-the-art motion recognition methods including LTPCNN [32], TDFD [33], DEAPP [34].

LTPCNN: The main idea of the method is the action mapping image classification via convolutional neural network (CNN) based approach. Firstly, we project the raw frames onto three orthogonal Cartesian planes and stack the results into three still images (corresponding to the front, side, and top views) to form the Depth Motion Maps (DMMs). Secondly, Local Ternary Pattern (LTP) is introduced as an image filter for DMMs, thus to



Fig. 9. The sample images in basketball data

improve the distinguishability of similar actions. Finally, we apply CNN to action recognition by classifying corresponding LTP-encoded images.

TFDF: It proposed a shoulder motion recognition optimization method based on the maximizing mutual information from multiclass CSP selected spatial feature channels and wavelet packet features extraction.

DEAPP: It proposed a novel edge-aware end-to-end deep network method, which used the edge-aware pooling module to improve contour accuracy and captured video sequences using multi-scale pyramid pooling layer spatial-time context feature.

We use the related code to transform the non-uniform dataset as matrix format to train and test this model.

4.3. Average accuracy performance test

Average accuracy testing will be conducted on three data sets in this experiment. We test all the categories and get the test results. Assume $a_i (1 \leq i \leq k)$ is the i -th accuracy rate, the average accuracy is $acc_{average} = \frac{a_1 + \dots + a_k}{k}$, k is the action class. The experimental results are shown in tables 2-4. It can be seen that the average accuracy of the DRN-SEAM recognition model in this paper is higher than other methods on the three different data sets.

In the WISDM data set, DEAPP has the best average accuracy in relevant studies due to the use of CNN. And it also adds the additional mathematical statistics features by manual extraction in the dense layer. However, the accuracy of the DRN-SEAM model is higher than that of the DEAPP in the absence of any artificial features. The recognition rate of the proposed DRN-SEAM is higher than that of other methods. The experimental results also show that the feature extraction ability of the Squeeze-and-Excitation structure is higher than the normal CNN and deep separable convolution.

Table 2. Results with different methods on WISDM

Method	Average accuracy (%)
ResNet50	86.07
LTPCNN	86.78
TFDF	86.57
DEAPP	96.52
DRN-SEAM	98.49

Table 3. Results with different methods on UCI

Method	Average accuracy (%)
ResNet50	77.25
LTPCNN	78.82
TFDF	84.53
DEAPP	96.35
DRN-SEAM	98.67

Table 4. Results with different methods on basketball data

Method	Average accuracy (%)
ResNet50	74.86
LTPCNN	76.47
TFDF	82.17
DEAPP	86.54
DRN-SEAM	91.37

Table 5, table 6 and table 7 display the detailed classification results of each action on WISDM, UCI and basketball data, and compare them with other methods. It can be seen from the two tables, jogging, walking, and standing are the easiest to recognize on WISDM. The accuracy of all the methods is more than 90%. Since the changes in the three actions have the biggest difference than other actions. Downstairs and downstairs are difficult to recognize. Because the two actions are the easiest to confuse. However, the recognition accuracy of the DRN-SEAM is bigger than 90%. Figures 10,11,12 are the statistical analysis for these data sets, which also shows that the proposed method has better result.

Table 5. Classification accuracy on WISDM with different methods (%)

Motion type	ResNet50	LTPCNN	TFDF	DEAPP	DRN-SEAM
downstairs	63.48	51.64	87.35	80.12	92.16
jogging	92.54	94.83	97.98	98.25	99.26
sitting	82.96	83.58	83.74	98.61	98.76
standing	93.87	95.87	93.45	92.72	98.87
upstairs	71.55	66.78	72.34	84.38	91.85
walking	83.79	84.65	98.61	97.81	99.32

In order to observe the performance of DRN-SEAM model in a more detailed way, the accuracy rate, recall rate and F1 value of the six motions are counted respectively as shown in table 8, table 9 and table 10.

By recording the accuracy of each network on the test set after each epoch training, the accuracy is obtained as shown in figure 13 and figure 14.

It can be seen that the accuracy of DRN-SEAM and DEAPP during the training process has obvious advantages over the other two networks after 200 epochs. This indicates that the feature map extracted by the DRN-SEAM feature extraction module can be accu-

Table 6. Classification accuracy on UCI with different methods (%)

Motion type	ResNet50	LTPCNN	TFDF	DEAPP	DRN-SEAM
downstairs	97.45	97.65	98.24	98.87	99.86
jogging	93.77	97.88	98.67	100.00	100.00
sitting	91.85	92.36	94.87	95.87	99.87
standing	98.36	98.67	98.96	99.24	99.98
upstairs	98.66	98.67	98.56	99.41	100.00
walking	99.45	99.67	99.85	100.00	100.00

Table 7. Classification accuracy on basketball with different methods (%)

Motion type	ResNet50	LTPCNN	TFDF	DEAPP	DRN-SEAM
Jumping	81.32	87.32	88.34	90.23	92.54
shooting	82.46	88.47	89.75	89.67	93.57
Falling back	83.64	85.63	84.16	91.22	91.83

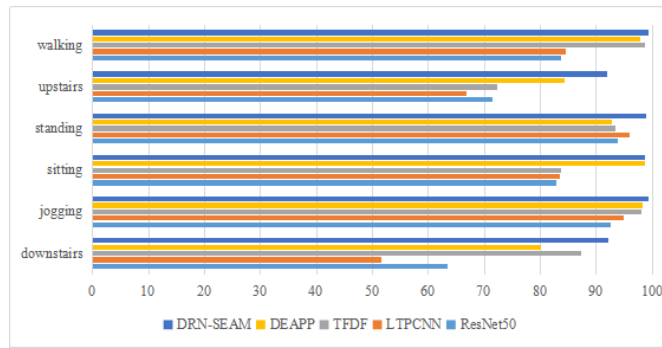


Fig. 10. Bar chart analysis for WISDM data set

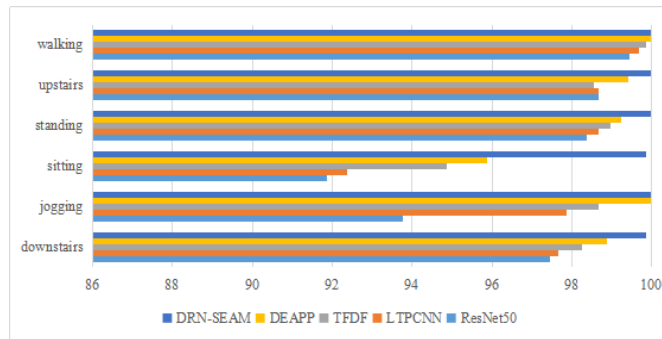


Fig. 11. Bar chart analysis for UCI data set

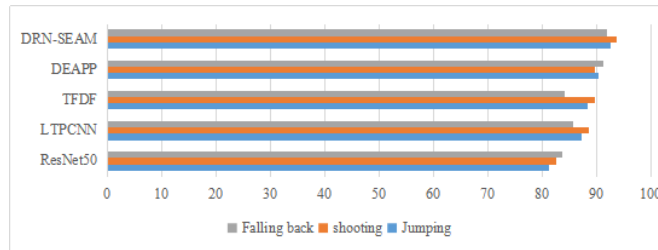


Fig. 12. Bar chart analysis for basketball data set

Table 8. Precision, recall and F1-score of action recognition on WISDM (%)

Method	Precision	Recall	F1
ResNet50	89.64	87.25	88.41
LTPCNN	96.35	94.65	96.85
TFDF	96.87	95.87	96.99
DEAPP	97.65	96.84	97.82
DRN-SEAM	98.26	97.98	98.96

Table 9. Precision, recall and F1-score of action recognition on UCI (%)

Method	Precision	Recall	F1
ResNet50	91.65	86.57	89.77
LTPCNN	96.87	97.21	98.24
TFDF	97.25	98.25	98.65
DEAPP	98.13	98.67	99.12
DRN-SEAM	99.69	99.54	99.57

Table 10. Precision, recall and F1-score of action recognition on basketball (%)

Method	Precision	Recall	F1
ResNet50	82.54	76.21	77.59
LTPCNN	89.32	82.14	83.65
TFDF	90.54	83.46	85.28
DEAPP	91.23	84.55	86.95
DRN-SEAM	92.87	91.32	90.77

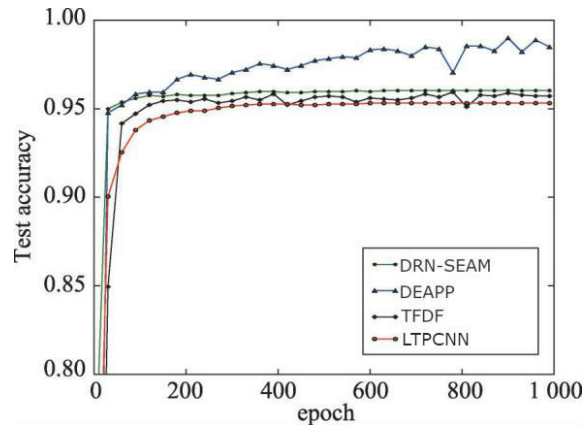


Fig. 13. Accuracy curve on UCI

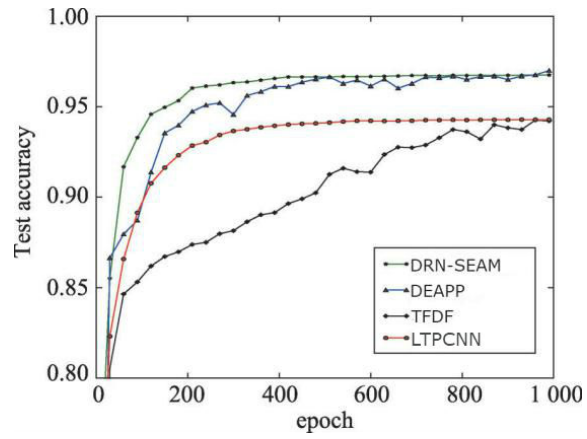


Fig. 14. Accuracy curve on WISDM

rately classified after 200 epochs. In this paper, the precision-recall (PR) curve is used to identify the accuracy index to evaluate the performance of action recognition algorithms. First, the PR value is calculated by predicting the contour processing of the action, and then the PR curve is drawn. Figure 14 is the basketball action PR curve. As shown in figure 15, the recognition PR curve on the basketball action dataset with proposed method can achieve better performance.

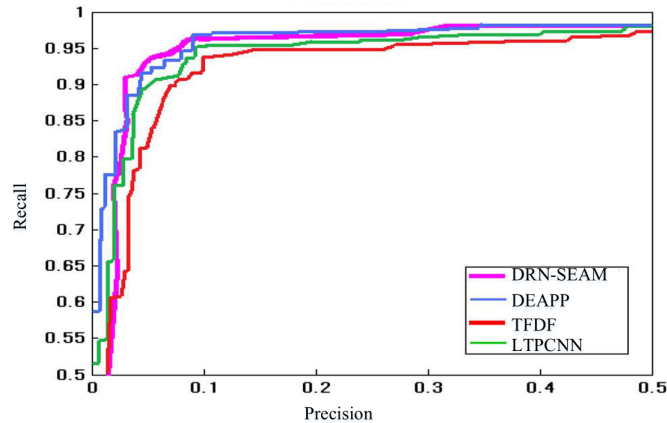


Fig. 15. PR curve on the basketball action dataset

4.4. Time consuming test

This experiment verifies the time-consuming advantages of the proposed DRN-SEAM structure in this paper. The experiment counts the time when the four networks are trained for 500 iterations simultaneously on the UCI, WISDM and basketball data set. The advantages of the DRN-SEAM are obtained by the time comparison. The experimental results are shown in tables 11-13. It can be seen from table 11 that the real-time performance of DRN-SEAM network has great advantages than other three networks. It shows that the proposed method has been greatly improved in efficiency as well as accuracy.

Table 11. Time consuming on UCI/s

Method	Epoch=100	Epoch=200	Epoch=300	Epoch=400	Epoch=500
DRN-SEAM	85	90	95	97	100
DEAPP	120	140	160	180	220
TFDF	150	180	210	240	270
LTPCNN	500	900	1300	1700	2100

Table 12. Time consuming on WISDM/s

Method	Epoch=100	Epoch=200	Epoch=300	Epoch=400	Epoch=500
DRN-SEAM	90	100	110	120	130
DEAPP	140	160	180	200	220
TFDF	160	200	240	280	320
LTPCNN	450	850	1250	1650	2050

Table 13. Time consuming on basketball/s

Method	Epoch=100	Epoch=200	Epoch=300	Epoch=400	Epoch=500
DRN-SEAM	74	84	89	97	105
DEAPP	120	145	170	195	215
TFDF	155	195	235	275	315
LTPCNN	480	891	1400	1871	2390

5. Conclusions

Action recognition is an important research direction in computer vision, which has worldwide applications, such as video surveillance, human-robot interaction and so on. Due to the influence of complex background and multi-angle changes, accurate recognition and analysis of human action in real-life scenarios is still a challenging problem. In order to improve the accuracy of action detection and recognition, this paper modifies the residual network by improving the order of "BN+ReLU+convolutional layer" in the residual block. And we introduce the attention mechanism and adjust the structure of the network convolution kernel to improve the recognition and classification effect of the model. The experimental results show that the proposed network model is better than the traditional deep residual network in terms of classification accuracy and convergence speed. In this study, we confine our technique to static images of human action.

The difficulty of action recognition lies in the huge changes in the specific actions. So the models often have poor accuracy. It performs better on one data set, but it is inferior to other models on the other data sets. Meanwhile, error data exists in the real-time situation of mobile phone sensor, this requires that the model needs to have high accuracy, adaptability, robustness and better data fault tolerance ability. Next work, RNN-based models will be researched with its better ability of handling the time sequence dependence information by combining the SE attention mechanism. Also, future studies will dynamically test our system on humans action with a robot system in real-world settings.

Availability of data and materials. The data used to support the findings of this study are available from the corresponding author upon request.

Competing interests. The authors declare that they have no conflicts of interest.

References

1. Peng L, Chen Z, Yang L T, et al. "Deep Convolutional Computation Model for Feature Learning on Big Data in Internet of Things," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 2, pp. 790-798, Feb. 2018.

2. Rajini A R, Abishek E, Ramesh S, et al. "Compact Printed Planar Eye Shaped Dipole Antenna for Ultra-Wideband Wireless Applications," *Journal of Applied Science and Engineering*, vol. 25, no. 5, pp. 761-766, 2021.
3. Yeh, J., Tsai, C. "A Graph-based Feature Selection Method for Learning to Rank Using Spectral Clustering for Redundancy Minimization and Biased PageRank for Relevance Analysis," *Computer Science and Information Systems*, Vol. 19, No. 1, pp. 141-164. (2022).
4. Zhong X, Huang W, Luo R, et al. "Video Human Behavior Recognition Based on ISA Deep Network Model," *International Journal of Pattern Recognition and Artificial Intelligence*, 2020. doi: 10.1142/S0218001420560121
5. S. Yin and H. Li. "Hot Region Selection Based on Selective Search and Modified Fuzzy C-Means in Remote Sensing Images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 5862-5871, 2020, doi: 10.1109/JSTARS.2020.3025582
6. Li M, Chen T, Du H. "Human Behavior Recognition Using Range-Velocity-Time Points," *IEEE Access*, vol. 8, pp. 37914-37925, 2020. doi: 10.1109/ACCESS.2020.2975676
7. Mandić, M. "Semantic Web Based Platform for the Harmonization of Teacher Education Curricula," *Computer Science and Information Systems*, Vol. 19, No. 1, pp. 229-250. (2022).
8. L. Jiao and J. Zhao. "A Survey on the New Generation of Deep Learning in Image Processing," *IEEE Access*, vol. 7, pp. 172231-172263, 2019, doi: 10.1109/ACCESS.2019.2956508.
9. Chen W. "A Novel Long Short-Term Memory Network Model For Multimodal Music Emotion Analysis In Affective Computing," *Journal of Applied Science and Engineering*, vol. 26, no. 3, pp. 367-376, 2022.
10. Ding S, Sun Y, An Y, et al. "Multiple birth support vector machine based on recurrent neural networks," *Applied Intelligence*, vol. 50, no. 7, pp. 2280-2292, 2020.
11. R. Jiao, T. Zhang, Y. Jiang and H. He, "Short-Term Non-Residential Load Forecasting Based on Multiple Sequences LSTM Recurrent Neural Network," *IEEE Access*, vol. 6, pp. 59438-59448, 2018.
12. Jiang F, Yuen K K R, Lee E W M. "A long short-term memory-based framework for crash detection on freeways with traffic data of different temporal resolutions," *Accident Analysis & Prevention*, vol. 141:105520, 2020.
13. Ronao C.A., Cho SB. "Deep Convolutional Neural Networks for Human Activity Recognition with Smartphone Sensors," *Neural Information Processing. ICONIP 2015. Lecture Notes in Computer Science*, vol. 9492. Springer, Cham.
14. M. Zeng et al., "Convolutional Neural Networks for human activity recognition using mobile sensors," *6th International Conference on Mobile Computing, Applications and Services*, Austin, TX, pp. 197-205, 2014.
15. Moya Rueda F, Grzeszick, René, Fink G, et al. "Convolutional Neural Networks for Human Activity Recognition Using Body-Worn Sensors," *Informatics*, vol. 5, no. 2, 2018.
16. Jain S, Rustagi A, Saurav S, et al. "Three-dimensional CNN-inspired deep learning architecture for Yoga pose recognition in the real-world environment," *Neural Computing and Applications*, pp. 1-15, 2020.
17. Yadav S K, Singh A, Gupta A, et al. "Real-time Yoga recognition using deep learning," *Neural Computing and Applications*, vol. 31, no. 12, pp. 9349-9361, 2019.
18. Alghyaline S. "Real-time Jordanian license plate recognition using deep learning," *Journal of King Saud University-Computer and Information Sciences*, 2020.
19. Lei, Zhang, Yang, et al. "RFR-DLVT: a hybrid method for real-time face recognition using deep learning and visual tracking," *Enterprise Information Systems*, 2020.
20. Qamar S, Jin H, Zheng R, et al. "3D Hyper-Dense Connected Convolutional Neural Network for Brain Tumor Segmentation," *IEEE, 14th International Conference on Semantics, Knowledge and Grids (SKG)* 2018. IEEE, 2019.
21. A. P. Tafti, F. S. Bashiri, E. LaRose and P. Peissig, "Diagnostic Classification of Lung CT Images Using Deep 3D Multi-Scale Convolutional Neural Network," *2018 IEEE Interna-*

- tional Conference on Healthcare Informatics (ICHI)*, New York, NY, 2018, pp. 412-414, doi: 10.1109/ICHI.2018.00078
22. Lin K, Li C, Zhao H, et al. "Face Detection and Segmentation Based on Improved Mask R-CNN," *Discrete Dynamics in Nature and Society*, 2020. doi: 10.1155/2020/9242917
 23. Guoli Yan, Huiyan Wang, et al. "Semantic annotation for complex video street views based on 2D-3D multi-feature fusion and aggregated boosting decision forests," *Pattern Recognition the Journal of the Pattern Recognition Society*, vol. 62, pp. 189-201, 2017.
 24. Weng Z, Guan Y. "Action recognition using length-variable edge trajectory and spatio-temporal motion skeleton descriptor," *EURASIP Journal on Image and Video Processing*, vol. 2018, no. 1, 2018.
 25. M. Zhou, "Feature Extraction of Human Motion Video Based on Virtual Reality Technology," *IEEE Access*, vol. 8, pp. 155563-155575, 2020, doi: 10.1109/ACCESS.2020.3019233.
 26. Jahandad, Suriani Mohd Sam, Kamilia Kamardin, Nilam Nur Amir Sjarif, Norliza Mohamed. "Offline Signature Verification using Deep Learning Convolutional Neural Network (CNN) Architectures GoogLeNet Inception-v1 and Inception-v3," *Procedia Computer Science*, vol. 161, pp. 475-483, 2019.
 27. Shoulin Yin, Ye Zhang, Shahid Karim. "Large Scale Remote Sensing Image Segmentation Based on Fuzzy Region Competition and Gaussian Mixture Model," *IEEE Access*, vol. 6, pp. 26069-26080. 2018.
 28. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv: 1502.03167, 2015.
 29. J. Hu, L. Shen, S. Albanie, G. Sun and E. Wu. "Squeeze-and-Excitation Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 8, pp. 2011-2023.
 30. Jeffrey W. Lockhart, Gary M. Weiss. "Limitations with Activity Recognition Methodology & Data Sets," *Proceedings of the 2014 ACM Conference on Ubiquitous Computing (UBICOMP) Adjunct Publication (2nd International Workshop on Human Activity Sensing Corpus and its Application)*, Seattle, WA, 2014.
 31. Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, Jorge L. Reyes-Ortiz. "Energy Efficient Smartphone-Based Activity Recognition using Fixed-Point Arithmetic," *Journal of Universal Computer Science*, vol. 19, no. 9, May 2013.
 32. Z. Li, Z. Zheng, F. Lin, H. Leung, and Q. Li. "Action recognition from depth sequence using depth motion maps-based local ternary patterns and CNN," *Multimedia Tools Appl.*, vol. 78, no. 14, pp. 19587-19601, 2019.
 33. Bai D, Chen S, Yang J. "Upper Arm Motion High-Density sEMG Recognition Optimization Based on Spatial and Time-Frequency Domain Features," *Journal of Healthcare Engineering*, 2019, 2019:1-16.
 34. Xu L, Yan S, Chen X, et al. "Motion Recognition Algorithm Based on Deep Edge-Aware Pyramid Pooling Network in Human-Computer Interaction," *IEEE Access*, vol. 7, pp. 163806-163813, 2019.

Xinxiang Hua is with College of Marxism, Zhengzhou University of Science and Technology, Zhengzhou, 450015 China. His research interests include image processing and education.

Received: March 22, 2022; Accepted: August 20, 2022.

Human Action Recognition Using a Depth Sequence Key-frames Based on Discriminative Collaborative Representation Classifier for Healthcare Analytics

Yuhang Wang¹, Tao Feng^{2,3}, and Yi Zheng^{1,*}

¹ Institute of physical culture, Harbin University
Harbin 150000, China
83008943@qq.com
86959936@qq.com

² Department of Physical Education, Harbin Finance University
Harbin, 150000, China
ancrum@qq.com

³ The Graduate School of Saint Paul University Philippines
Ottawa, Philippines

Abstract. Using deep map sequence to recognize human action is an important research field in computer vision. The traditional deep map-based methods have a lot of redundant information. Therefore, this paper proposes a new deep map sequence feature expression method based on discriminative collaborative representation classifier, which highlights the time sequence of human action features. In this paper, the energy field is established according to the shape and action characteristics of human body to obtain the energy information of human body. Then the energy information is projected onto three orthogonal axes to obtain deep spatial-temporal energy map. Meanwhile, in order to solve the problem of high misclassification probability of similar samples by collaborative representation classifier (CRC), a discriminative CRC (DCRC) is proposed. The classifier takes into account the influence of all training samples and each kind of samples on the collaborative representation coefficient, it obtains the highly discriminative collaborative representation coefficient, and improves the discriminability of similar samples. Experimental results on MSR_Action3D data set show that the redundancy of key-frame algorithm is reduced, and the operation efficiency of each algorithm is improved by 20%-30%. The proposed algorithm in this paper reduces the redundant information in deep map sequence and improves the extraction rate of feature map. It not only preserves the spatial information of human action through the energy field, but also records the temporal information of human action in a complete way. What's more, it still maintains a high recognition accuracy in the action data with temporal information.

Keywords: action recognition, deep map sequence, deep spatial-temporal energy map, discriminative CRC, energy information.

1. Introduction

Human action recognition is a research hotspot in machine vision and artificial intelligence [1,2]. Many research achievements have been applied in the fields of human-

* Corresponding author

computer interaction, biometrics, health monitoring, video surveillance system, motion sensing games, robotics, etc.. Most of the early researches on action recognition were carried out on color video sequences collected by RGB cameras. For example, motion energy image (MEI) and motion history image (MHI) proposed by Bobick and Davis [3] were originally color videos collected by RGB cameras. MEI represents the outline of human action and does not involve the time sequences information of human action. MHI expresses temporal information and spatial contour of human action through brightness attenuation. However, due to the action occlusion, some action information is missing, and the final expressed time sequence information is incomplete. Due to the existence of redundant frames, the gray value of the final generated MHI is relatively concentrated near the redundant frames, which affects the final recognition accuracy [4].

With the development of imaging technology, especially the introduction of depth camera, the research object of human action recognition is transformed from the original RGB image to the depth image. Compared with the previous RGB images, the depth map sequences collected by the structured light depth sensor are not sensitive to light changes, and provide depth human action data. So far, researchers have done a lot of researches on depth map sequences. Zhu et al. [5] proposed 3D points, using a small amount of 3D points to represent human action. Luo et al. [6] proposed the depth cube and established a novel depth cube similarity feature to describe the local 3D depth cube around the depth map sequence. Xuan et al. [7] proposed surface normals and used 4-dimensional surface normal-direction histogram descriptors to capture the structural information of human action changes. Nie et al. [8] proposed bone joints, which were used to represent human action. Chaudhary et al. [9] used a depth motion map (DMM) to represent human action. Where, DMM was to project the depth map onto three orthogonal Cartesian planes, generated 2D projection maps from three perspectives according to the front view, side view and top view, and accumulated the image difference between two continuous projection maps to generate DMM from three perspectives. Mattiev et al. [10] proposed new associative classifiers, called DC, DDC and CDC, that used distance-based agglomerative hierarchical clustering as a post-processing step to reduce the number of its rules, and in the rule-selection step, it used different strategies (based on database coverage and cluster center) for each algorithm. Human action is composed of spatial information and temporal information. Spatial information reflects the spatial distribution of human body information, and temporal information reflects the sequence of human body information. DMM completely describes the spatial information of human action, but it cannot describe the temporal information of depth map sequence. When there are some actions with the same space trajectory and opposite time sequence in the database, the generated feature map is the same, but the two actions cannot be distinguished.

Although human action recognition has made great progress in recent years, it still has many shortcomings. In this paper, a key frame algorithm is proposed to solve the problem of too much redundant information in depth map sequence. Firstly, the redundancy coefficient is utilized to describe the redundancy. Then, according to the sequence of redundant coefficients, the redundant frames in the depth map sequence are located and deleted to obtain the key frame sequence to express human action sufficiently. Our main contributions are as follows:

1. In this paper, a new depth spatial-temporal energy feature expression (abbreviated as DSTEFE) method is proposed to solve the problem of poor temporal information of feature maps extracted from depth map sequences.
2. The energy field is established according to the shape and action characteristics of human body to obtain the energy information of human body. Then the energy information is projected onto three orthogonal axes to obtain deep spatial-temporal energy map.
3. Meanwhile, in order to solve the problem of high misclassification probability of similar samples by collaborative representation classifier (CRC), a discriminative CRC (DCRC) is proposed. The classifier takes into account the influence of all training samples and each kind of samples on the collaborative representation coefficient, it obtains the highly discriminative collaborative representation coefficient, and improves the discriminability of similar samples. This proposed method highlights the action information of human body and further improves the accuracy of action recognition.

This paper is organized as follows. Section 2 introduces the related works. Section 3 detailed illustrate the proposed DSTEFE based on DCRC for human action recognition. Section 4 gives the experiments for the proposed method. Finally, a conclusion is conducted in Section 5.

2. Related Works

In the early stage of human action recognition, people usually use RGB camera to collect the color video sequence of human action, and then extract the feature map from the color video sequence. MEI is initially extracted from the color video sequence. Firstly, the foreground area of human action is extracted, and binarization is carried out to obtain the binarization image sequence $B(x, y, t)$. Then, the union set of the binary image sequence is evaluated to obtain the feature graph of MEI [11]. The calculation of MEI is as follows:

$$M_{\delta}(x, y, t) = \cup_{i=0}^{\delta-1} B(x, y, t - i). \quad (1)$$

Where $M_{\delta}(x, y, t)$ represents MEI generated by δ images at frame t in the video sequence. x and y represent the height and width values of one point in the image respectively. t denotes the serial number of a frame in the image sequence.

MEI expresses the spatial contour of human action through the union of the binary foreground region. However, the video sequence of human action expressed in this way has the following problems: 1) MEI represents the maximum contour boundary of human action. Due to the occlusion of action information from front to back, some action information will be lost during movement; 2) MEI cannot express the time sequence information of human action. When there are actions in the database with the same spatial trajectory and opposite time sequence, the generated feature maps are the same and cannot be distinguished.

In order to show the human action, MHI is a feature map that can express some temporal information of human action. Different from MEI, MHI is a gray image. The gray value at each point is a time-history function. MHI can be represented by a simple substitution and attenuation operator, calculated as:

$$H_\tau(x, y, t) = \begin{cases} \tau & B(x, y, t) = 1 \\ \max(0, H_\tau(x, y, t - 1)) & \text{others} \end{cases} \quad (2)$$

Where $H_\tau(x, y, t)$ is the MHI generated by τ images at frame t in the video sequence. τ is the initial brightness. $B(x, y, t)$ is the binary image sequence.

Compared with MEI, MHI is significantly improved. It not only retains the spatial outline of human action, but also shows the temporal information of human action by brightness attenuation. But there are also some shortcomings: 1) There are many redundant frames in the collected video sequence, so that the gray value distribution of the final generated MHI is concentrated near the redundant frames, which seriously affects the accuracy of recognition; 2) The front and back occlusion of action information makes some action information missing, which makes it impossible to accurately express human action.

With the introduction of depth camera, people also begin to use depth map sequence for human action recognition research. Compared with the previous color video sequences, depth map sequences are not sensitive to light changes, so it is more convenient to extract the foreground area of human action and provides the depth information of human action. Yang et al. [12] proposed DMM, which projected each frame of depth map sequence onto three orthogonal Cartesian planes, and generated 2D projection images from three perspectives according to the front view, side view and top view, respectively. They were represented by map_f , map_s and map_t . DMM is calculated as:

$$S_v = \sum_{i=2}^F (|map_v^i - map_v^{i-1}| > \varepsilon). \quad (3)$$

Where $v \in f, s, t$ represents the projection angle of view. f, s and t denote the front view, side view and top view. S_v is the DMM of projection angle of view v . map_v^i is the i -th frame graph of projection angle of view v . ε is the difference threshold. F is the frame number of the depth map sequence. $|map_v^i - map_v^{i-1}|$ represents the difference image of two consecutive projected images.

Compared with MEI, DMM fully uses the depth information of depth map sequence, but the sequence DMM information is also unable to express the temporal information of human action, and does not have the ability to distinguish positive and negative sequence actions.

3. Proposed DSTEFE Based on DCRC for Human Action Recognition

3.1. Proposed action recognition framework

The human action recognition framework based on DSTEFE and DCRC algorithm is shown in figure 1. Firstly, the redundant frames in depth map sequence are eliminated by the redundancy coefficient of the difference image sequence, and the key frame sequence sufficiently expressing human action is obtained. Then the energy field is established according to the shape and action characteristics of human body to obtain the energy information of human body. Then the energy information of human body is projected onto

three orthogonal axes to obtain DSTEFE. Finally, HOG (histogram of Oriented gradient) features are extracted from each DSTEM and sent to a new DCRC classifier for human action recognition.

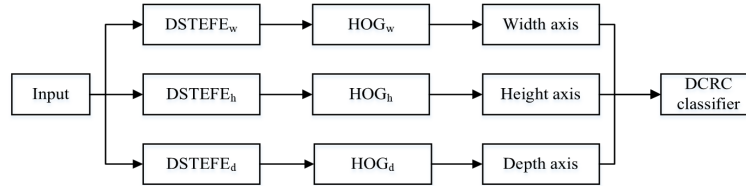


Fig. 1. Proposed human action recognition framework

3.2. Key frame algorithm

Due to the uneven human action rate during sampling, there are a large number of similar frames in the collected depth map sequences. In this paper, the similar frames appearing at the similar time in the depth map sequences are called redundant frames. After the redundant frames are eliminated, the remaining depth map sequences are called key frame sequences.

In the action recognition process, human action can be expressed only by the depth frame of the key position. However, there are a lot of redundant frames in the collected database, which has a great influence on the future research. Aiming at the above problems, the redundancy coefficient is proposed to describe the redundancy degree, and the key frame algorithm is further proposed based on the redundancy coefficient. By eliminating redundant frames of depth map sequence, redundant information is reduced, which makes the same action have approximate time interval, thus improving the operation rate of feature map and recognition accuracy.

The overall flow of the key frame algorithm is shown in figure 2.

1. The image difference between two adjacent frames of depth map sequences is obtained and the image difference sequence is generated.
2. The redundant frames in depth map sequence are located and deleted by the maximum redundancy coefficient.
3. Repeat the above steps until obtain the key frame sequence sufficiently expressing the human action.

This algorithm firstly executes difference processing between adjacent frames of depth map sequence, and then obtains the difference image of adjacent frames, which is calculated as:

$$D(x, y, t) = |I(x, y, t + 1) - I(x, y, t)|. \quad (4)$$

Where $I(x, y, t)$ is the i -th frame image of the original depth map sequence. $D(x, y, t)$ is the difference image between $(t + 1)$ -th frame and t -th frame of the original depth map sequence, namely the t -th frame of the difference image sequence.

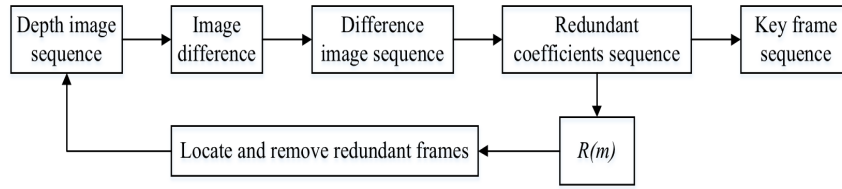


Fig. 2. Key frame flow chart

Then the redundancy coefficient of each frame in the difference image sequence is calculated to show the similarity between adjacent frames in the original depth map sequence. The calculation process of the redundancy coefficient of each frame in the differential image sequence is as follows.

First, the L2-norm of each frame in the difference sequence is calculated.

$$\alpha(t) = \|D(t)\|_2 = \sqrt{\lambda_{max}(D^T(t)D(t))}. \quad (5)$$

Where $\alpha(t)$ represents the L2-norm of t -th frame in the difference image sequence. λ_{max} is the maximum eigenvalue of a difference image.

Second, the L2-norm values of each frame in the difference image sequence are projected to the interval [0,1] to obtain the corresponding redundancy coefficient.

$$R(t) = e^{-\alpha(t)}. \quad (6)$$

Where $R(t)$ represents the redundancy coefficient of t -th frame in the difference image sequence.

Third, the redundancy coefficients of each frame in the difference image sequence are sorted from big to small. It will find out the maximum redundancy coefficient $R(m)$ and its corresponding difference image frame $D(m)$. According to the difference image frame $D(m)$, it finds the corresponding redundant frames in the original depth map sequence and removes them. Repeat the above operations, remove redundant frames in the sequence, and obtain N frame sequences sufficiently expressing human action. N is determined by the experiment results.

3.3. Deep Spatial-temporal Energy Feature Expression

To solve the problem of missing time sequence information of generated feature map from depth map sequence, a feature expression method that can completely express action spatial-temporal sequence information is proposed, namely DSTEFE. DSTEFE reflects the change of human action energy information distribution on three orthogonal axes. Firstly, the energy field is established according to the shape and action characteristics of human body to obtain the energy information of human action. The energy information of human body is projected to 3 orthogonal Cartesian planes to generate 2d-projection images from 3 perspectives. Then two 2d-projection images are selected to continue to project onto the three orthogonal axes to generate a one dimension energy distribution list. DSTEFE with three orthogonal axes is formed after time order splicing. The three

orthogonal axes are width axis (w), height axis (h) and depth axis (d), which correspond to the width direction, height direction and depth direction of the depth frame, respectively. L_w , L_h and L_d represent the corresponding one-dimension energy distribution list. The flow chart of DSTEFE is shown in figure 3.

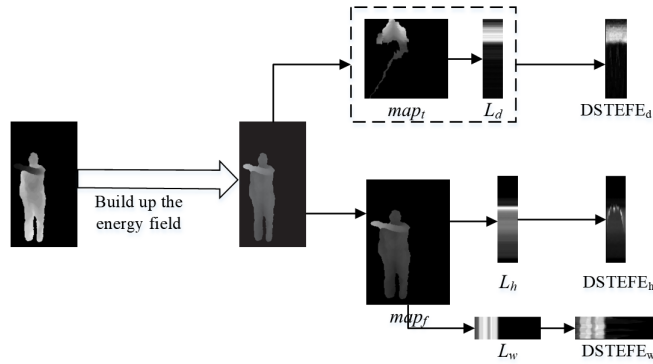


Fig. 3. The flow chart of DSTEFE

Step 1. Building up the energy field.

As shown in figure 4, the energy field of human body is first established to obtain the energy information of human action, so as to highlight the information of human action. The energy field coordinate system is shown in figure 4. It takes the height of the depth map as the x-axis direction and the width of the depth map as the y-axis direction.

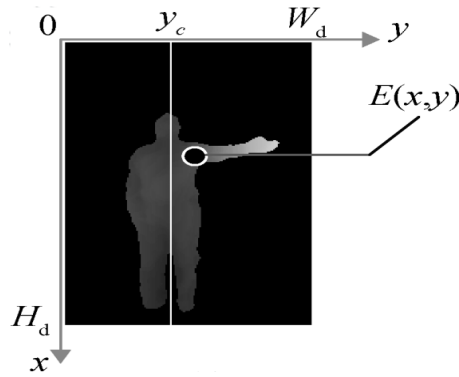


Fig. 4. Energy field coordinate system

According to the characteristics of forward stretching of human action, the depth distance between human foreground and the background is denoted as the forward energy of human body. It is calculated as:

$$E_f(x, y) = 255 - f(x, y). \quad (7)$$

Where $E_f(x, y)$ represents the forward energy of the human body. $f(x, y)$ is the depth value of the human body. According to the characteristics of lateral extension of human action, the distance between the foreground and the central axis of human body is denoted as the lateral energy of human body. It is calculated as:

$$E_s(x, y) = |y - y_c|. \quad (8)$$

Where $E_s(x, y)$ represents the lateral energy of the human body. When y_c is the initial frame of the action, the y-axis coordinate of the human body under the stand-at-attention posture is calculated as:

$$\sum_{i=0}^{H_d} \sum_{j=0}^{y_c} f(i, j) = 0.5 \sum_{i=0}^{H_d} \sum_{j=0}^{W_d} f(i, j). \quad (9)$$

Where W_d is the width of the depth map. H_d is the height of the depth map. Because there are many overlap areas between stretch up and foreground in the human action, this paper does not record the height direction energy of human body. The total energy $E(x, y)$ of human body is calculated as:

$$E(x, y) = \sqrt{E_f^2(x, y) + E_s^2(x, y)}. \quad (10)$$

Because the forward energy and the lateral energy are linear operators, but the total energy is not linear operators. The absolute value is used to calculate the total energy.

$$E(x, y) = |E_f(x, y)| + |E_s(x, y)|. \quad (11)$$

The depth frame comparison of the energy field is shown in figure 5. Figure 5(a) shows the depth frame without the establishment of the energy field. Figure 5(b) shows the depth frame with the establishment of the energy field. Compared to figure 5(a), the establishment of energy field in figure 5(b) can significantly highlight the information of human action, which is conducive to enhancing the effect of human action recognition.

Step 2. Calculating DSTEFE.

The energy information of human body is projected to three Cartesian planes, and the 2D-energy projection diagram of three perspectives is generated according to the front view, side view and top view, which are represented by map_f , map_s and map_t respectively. In order to obtain the energy distribution of the width axis, height axis and depth axis in the action space, the front view and the top view are selected to continue to project onto the corresponding orthogonal axis. That is, the row sum or column sum of the two dimension energy projection graph is computed. Three one-dimension energy distribution lists are generated according to the width axis, height axis and depth axis, denoted as L_w , L_h and L_d , respectively. The formula is:

$$L_u(k) = \sum_{x=1}^{W_m} map_v(x, k) || \sum_{y=1}^{H_m} map_v(k, y). \quad (12)$$

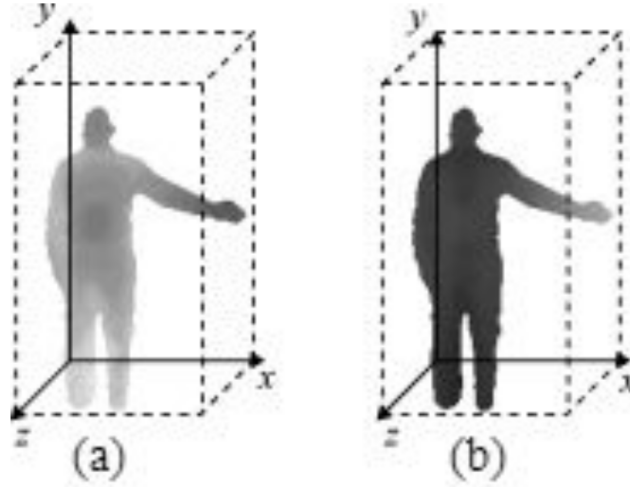


Fig. 5. The depth frame comparison

Where $v \in f, s, t, u = w, h, d$. w is width axis. h is height axis. d is depth axis. W_m is the width of the 2-dimension energy projection graph. H_m is the height of the 2-dimension energy projection graph. $L_u(k)$ is the k -th element of the projection list on the u axis.

$L_u(k)$ is normalized and spliced into DSTEFE of each axis in time order. For depth map sequence with N frames, DSTEFE is calculated as:

$$T_u(t) = L_u^t. \quad (13)$$

Where L_u^t represents the one-dimension energy distribution list of t -th frame on the u -axis. T_u stands for DSTEFE on the u -axis. $T_u(t)$ denotes the t -th row of T_u . Region of Interest (ROI) processing is carried out for each DSTEFE according to the maximum and minimum values of the width, height and depth in human action. That is, the image is cropped and the size is normalized.

3.4. Discriminative Collaborative Representation Classifier (DCRC)

Collaborative Representation Classifier (CRC) [13] is a very effective classifier, which is believed that the test samples can be approximately linearly represented by all training samples. Given a training sample set $D = D_1, \dots, D_k, \dots, D_K$ with K classes, where $D_k (k = 1, 2, \dots, K)$ is the sample vector set corresponding to category K . y is used to represent a test sample, D collaboration can be represented as $y = D\alpha$, where α is the collaboration representation coefficient vector of the test sample. In this subsection, we introduce a new CRC classifier.

Assuming S is used to represent the linear space spanned by the collaboration of all the training samples. S_i represents the linear subspace spanned by a sample $D_i (i = 1, 2, \dots, K)$ of the same class. $L = 1, 2, \dots, K$ denotes the set of all categories. The test samples that do not belong to the space S can be represented as $y \approx D\alpha$, which can only

indicate that the one category in test samples belongs to the one category in L . Then the residuals between the test samples and each category are reconstructed to approximate the category of the samples [14]. However, when the two classes in the training sample are very similar (such as D_i and D_j), the samples reconstructed by the corresponding coefficients α_i and α_j in the representational coefficient vectors obtained by CRC have a high similarity degree, so the probability of misclassification based on the residual classification rule will be increased.

In order to improve the discriminability of CRC for similar actions and improve the performance of the classifier, a highly discriminative cooperative representation coefficient is obtained by quadratic constraint for the coefficients, a DCRC classifier is proposed. First, a shared sample point $\hat{y} = D\alpha^* = D(\alpha_1^*, \dots, \alpha_K^*)$ in space S is determined, where α^* is the corresponding representation coefficient vector of the sample point. The shared sample point should satisfy two conditions: 1) the similarity between the sample point and the test sample is very high; 2) The distance sum from the reconstructed sample point $\hat{y}_i = D_i\alpha_i^*$ to the sample point in each subspace S_i is the minimum. Then, after continuous optimization for the objective function, the optimal collaboration representation coefficient α^* and the optimal shared sample point can be obtained. Finally, it will be stopped until the residual difference between the shared sample point \hat{y} and the reconstructed sample point \hat{y}_i in a subspace is the smallest. So the category of test sample is obtained as shown in figure 6.

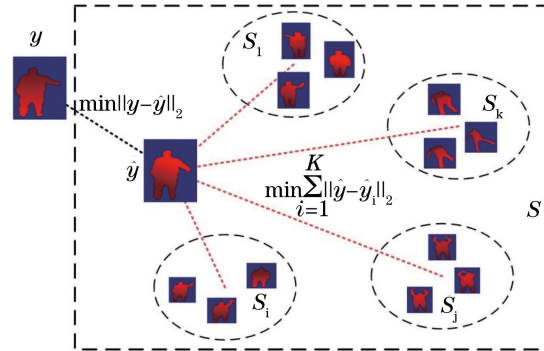


Fig. 6. DCRC process

It can be seen that the closer two samples denotes the greater probability that the two samples belong to a same category. Assuming that all samples are independently distributed in the space. $l(y)$ is used to represent the label of sample y , the probability of test sample y belonging to category i is:

$$\begin{aligned}
 P[l(y) = i] &= P[l(y) = l(\hat{y})|l(\hat{y}) = i]P[l(\hat{y}) = i] \\
 &= P[l(y) = l(\hat{y})|l(\hat{y}) = i]P[l(\hat{y}) = i|l(\hat{y}) \in L]P[l(\hat{y}) \in L]
 \end{aligned}
 \tag{14}$$

Because the samples are independent of each other. So $P[l(y) = l(\hat{y})|l(\hat{y}) = i]P[l(\hat{y}) = l(y)|l(\hat{y}) \in L]$, formula (14) is equivalent to:

$$\begin{aligned} P[l(y) = i] &= P[l(y) = l(\hat{y})|l(\hat{y}) = i]P[l(\hat{y}) = i] \\ &= P[l(y) = l(\hat{y})]P[l(\hat{y}) = i|l(\hat{y}) \in L] \end{aligned} \tag{15}$$

Here, $P[l(y) = l(\hat{y})]$ can measure the distance between test sample y and \hat{y} , that is, $\|y - \hat{y}\|^2$. Because \hat{y}_i falls inside the subspace S_i , so $P[l(\hat{y}) = i|l(\hat{y}) \in L]$ can be considered as measuring the distance between \hat{y} and \hat{y}_i , namely, $\sum_{i=1}^K \|\hat{y} - D_i\alpha_i\|_2^2$. To obtain the label of the test sample, let

$$\max P[l(y) = i] = \min(\|y - D\alpha\|_2^2 + \mu \sum_{i=1}^K \|\hat{y} - D_i\alpha_i\|_2^2). \tag{16}$$

In order to reduce the risk of over-fitting and computational complexity, Tikhonov matrix regularization term is used to constrain this function, and the final objective function is obtained:

$$\hat{\alpha} = \operatorname{argmin}_{\alpha} \|y - D\alpha\|_2^2 + \lambda \|\Gamma\alpha\|_2^2 + \mu \sum_{i=1}^K \|\hat{y} - D_i\alpha_i\|_2^2. \tag{17}$$

Where λ and μ are the regularization constraint parameters. $\sum_{i=1}^K \|\hat{y} - D_i\alpha_i\|$ makes double constraint for α_i based on D_i , which can enhance the discriminability of the final coefficient vector α . when $\mu = 0$, the model of formula (17) becomes CRC. So μ must be larger than zero. Equations (14)-(17) prove the feasibility of DCRC from the perspective of probability. The category of test sample y can be determined by taking the maximum probability belonging to a single category.

The partial derivative of constraint $\sum_{i=1}^K \|\hat{y} - D_i\alpha_i\|$ for coefficient vector α is solved as follows:

$$\begin{aligned} \frac{\partial}{\partial \alpha} \left(\sum_{i=1}^K \|\hat{y} - D_i\alpha_i\|_2^2 \right) &= \frac{\partial}{\partial \alpha} \sum_{i=1}^K \operatorname{tr}[(\hat{y} - D_i\alpha_i)^T (\hat{y} - D_i\alpha_i)] \\ &= \sum_{i=1}^K \frac{\partial}{\partial \alpha} \operatorname{tr}(\alpha^T D^T D \alpha - \alpha^T D^T D_i \alpha_i - \alpha_i^T D_i^T D \alpha + \alpha_i^T D_i^T D_i \alpha_i) \end{aligned} \tag{18}$$

Let $\bar{D}_i = [0, \dots, D_i, \dots, 0]$, so formula (18) can be simplified as:

$$\sum_{i=1}^K 2D^T D \alpha - 2(D^T \bar{D}_i \alpha + \bar{D}_i^T D \alpha) + 2\bar{D}_i^T \bar{D}_i \alpha. \tag{19}$$

Combined with the optimal solution of CRC model, the optimal solution of the discriminant cooperative representation classifier can be obtained:

$$\hat{\alpha} = [D^T D + \lambda \Gamma^T \Gamma + \mu \sum_{i=1}^K (D - \bar{D}_i)^T (D - \bar{D}_i)]^{-1} D^T y. \tag{20}$$

Finally, a new rule is used to determine the category of test sample:

$$e_i = \|D\hat{\alpha} - D_i\hat{\alpha}_i\|_2^2. \quad (21)$$

$$\text{label}(y) = \text{argmin}_i[e(i)]. \quad (22)$$

4. Experiments and Analysis

This experiment is conducted on MATALB2017a, Python3.5, CPU3.4GHz, GTX1060, windows 10. The public MSR_Action3D is selected as the experiment dataset. The database has 557 depth image samples and 20 different actions including high wave (A01), horizontal wave (A02), throw (A03), right hand grasp (A04), punch (A05), high throw (A06), cross (A07), hook (A08), circle (A08)(A09), clap (A10), hand swing (A11), side jab (A12), bend (A13), front kick (A14), side kick (A15), jog (A16), tennis swing (A17), tennis serve (A18), golf swing (A19), pick up throw (A20). Ten persons participate in the experiment, each person conducts action 2 to 3 times. In this paper, the original depth map sequence is called a positive order action, marked as Data1. The inverse action is called negative order action. In this paper, the inverse action is obtained by the reverse order of positive order action. The combination of the positive order action and inverse order action forms Data2. The positive order action in dataset 2 is the same as that in Data1. Inverse action contains inverse high wave (B01), inverse horizontal wave (B02), inverse throwing (B03), inverse grasp (B04), inverse strike (B05), inverse high throw (B06), inverse fork (B07), inverse hook (B08), inverse circle (B09), inverse clap (B10), inverse hands up swing (B11), inverse side jab (B12), inverse bend (B13), inverse forward kick (B14), inverse side kick (B15), inverse jog (B16), inverse tennis swing (B17), inverse tennis serve (B18), inverse golf swing (B19), inverse pick up throw (B20).

4.1. Experiment Set

Setting 1. Divide the actions in the data set into 3 groups, the actions with high similarity in the same group. The actions in Data1 are divided into AS1, AS2 and AS3. The actions in Data2 are classified as AS4, AS5 and AS6. The grouping of Data1 and Data2 is shown in table 1. Each group is tested three times. In test 1, 1/3 samples are used as training data, and the remaining samples are as test data. In test 2, 1/2 samples are used as training data, and the remaining samples are as test data. In test 3, 2/3 samples are used as training data, and the remaining samples are as test data.

Setting 2. Cross-verify is conducted for the whole action in the dataset. The samples are divided into 5 parts, where 4 parts are used for training and 1 part is used for testing. The final recognition result is the average of the five results. In this paper, the image block is 10×10 pixel. HOG feature is extracted by sliding image block with step size of 10 pixels. The local binary pattern features of the image are extracted by setting the parameters with a sampling radius of 2 and a sampling number of 8.

Setting 3. Ablation experiments are conducted to show the effectiveness of proposed method in terms of the addition of Depth Sequence Key-frames and Discriminative Collaborative Representation Classifier on the Data1 and Data2.

Table 1. Subsets of Data1 and Data2

Data1	Data1	Data1	Data2	Data2	Data2
AS1	AS2	AS3	AS4	AS5	AS6
A02	A01	A06	A02+B02	A01+B01	A06+B06
A03	A04	A14	A03+B03	A04+B04	A14+B14
A05	A07	A15	A05+B05	A07+B07	A15+B15
A06	A08	A16	A06+B06	A08+B08	A16+B16
A10	A09	A17	A10+B10	A09+B09	A17+B17
A13	A11	A18	A13+B13	A11+B11	A18+B18
A18	A12	A19	A18+B18	A12+B12	A19+B19
A20	A14	A20	A20+B20	A14+B14	A20+B20

Setting 4. The confusion matrix experiment results show that the unique complete timing of DSTEFE plays an important role in human behavior recognition on the database with both positive and reverse sequence behaviors.

4.2. Recognition results

According to the setting 1, the DSTEFE-HOG feature of each action in the three sub-datasets of Data1 is input into different classifiers for classification. In this paper, we select the four classical classifiers including Gaussian Bayes, Random Forest, K-nearest neighbor, SVM to make comparison. Next work direction is research more classifiers. Table 2 displays the recognition results of DSTEFE-HOG features in different classifiers. Table 3 shows the ablation experiment result.

Table 2. Recognition results of DSTEFE-HOG in different classifiers

Classifier	AS1	AS2	AS3
Gaussian Bayes	79.87	76.53	85.39
Random Forest	85.12	84.79	91.96
K-nearest neighbor	81.34	85.72	82.56
SVM	93.77	92.56	96.14
DCRC	98.91	95.71	99.25

Table 3. Recognition results of ablation experiment

Sub-model	AS1	AS2	AS3
Depth Sequence Key-frame	66.78	75.36	81.28
Depth Sequence Key-frame+DCRC	98.91	95.71	99.25

From table 2, it can be seen that DSTEFE-HOG has a high recognition accuracy on all classifiers, but the proposed DCRC has the best classification effect. In order to achieve

the most ideal recognition effect for DSTEFE-HOG feature, DCRC is used as the classifier in the following experiments. Table 3 shows that through our proposed scheme, the results of human behavior recognition have been greatly improved.

When key frame algorithm is carried out, the number of key frames N must be determined first. N directly affects the extraction speed of feature map and the removal of redundant information. Figure 7 shows the DSTEFE of the width axis of the tennis swing with different N . Figure 7(a) is the DSTEFE without key frame algorithm. It can be clearly seen from the contents in the white box that the feature map contains more redundant information. In figure 7(b), the number of key frames is 40, and many depth frames still belong to redundant frames, so the effect improvement is not ideal. In figure 7(d), the number of key frames is 25, which clearly shows that the depth frames of many key positions are lost, resulting in inaccurate action description. In figure 7(c), the number of key frames is 30, which not only eliminates redundant information in depth map sequence, but also retains the key information completely.

In this paper, in order to obtain the most ideal key frame sequence, the step length is set as 5 frames, and the recognition accuracy of the final extracted DSTEFE-HOG is taken as the standard value to find the most appropriate key frame number N from 2540 frames. According to setting 1, each action in the three sub-datasets of Data1 is extracted by key frame. The DSTEFE-HOG feature is calculated and results are shown in figure 8. Through the analysis of figure 8, it can be seen more intuitively that when $N=30$, the recognition accuracy on any sub-datasets is the highest, which indicates that the key frame sequence can best describe the depth map sequence when $N=30$.

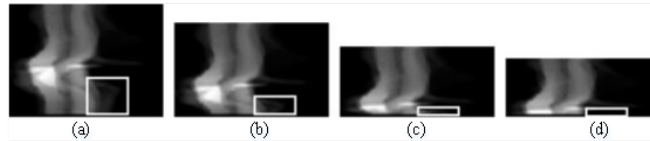


Fig. 7. DSTEFE effect with different key frames. (a) without key frame; (b) $N=40$; (c) $N=30$; (d) $N=25$.

In order to further verify the effectiveness of the key frame algorithm, this paper carries out a comparison experiment on the action recognition effect with/without the key frame algorithm on three sub-datasets according to experiment setting 1. Data1 contains 20 positive order actions. The results of different methods without and with key frame algorithm are shown in table 4 and table 5, respectively.

The recognition comparison results before and after the key frame algorithm in table 4 and table 5 show that the key frame algorithm eliminates the redundant frames in the depth map sequence, reduces the redundant information in the sequence, and improves the final recognition accuracy. Where, the recognition results of DSTEFE-HOG feature processed by key frame are significantly improved compared with those without key frame. The reason is that DSTEFE is formed by projecting the energy information of each frame in depth map sequence onto three orthogonal axes and spliced in time order, which is sensitive to redundant information. The key frame algorithm eliminates redundant frames,

Table 4. Recognition rate of different methods on Data1 without key frame algorithm (%)

Data	AS1	AS1	AS1	AS2	AS2	AS2	AS3	AS3	AS3
Method	Test1	Test2	Test3	Test1	Test2	Test3	Test1	Test2	Test3
MEI+HOG	73.41	86.35	86.41	73.14	81.69	86.95	72.41	71.28	90.65
MEI+LBP	56.27	65.24	71.34	53.40	62.39	73.79	54.84	60.47	75.79
MHI+HOG	69.97	83.60	86.42	64.58	81.69	88.27	72.99	72.18	90.65
MHI+LBP	53.53	66.72	68.61	56.01	65.02	71.16	54.84	54.16	71.73
DMM+HOG	76.14	84.51	87.78	71.82	85.21	86.95	77.81	75.79	82.67
DMM+LBP	57.64	75.34	86.41	63.93	71.16	78.99	64.97	66.87	83.89
DSTEF+HOG	91.24	89.74	94.25	79.78	86.17	90.24	81.38	84.56	94.77

Table 5. Recognition rate of different methods on Data2 with key frame algorithm (%)

Data	AS1	AS1	AS1	AS2	AS2	AS2	AS3	AS3	AS3
Method	Test1	Test2	Test3	Test1	Test2	Test3	Test1	Test2	Test3
MEI+HOG	74.08	84.51	86.41	74.45	81.69	86.95	75.79	72.18	91.99
MEI+LBP	56.27	65.24	75.45	52.09	58.89	68.53	54.16	59.57	75.79
MHI+HOG	70.67	83.61	86.42	69.19	81.69	88.27	76.46	73.11	90.65
MHI+LBP	53.54	67.08	71.34	56.11	65.12	73.79	55.84	61.37	73.08
DMM+HOG	72.03	88.17	91.89	77.08	86.95	86.27	77.81	77.59	94.71
DMM+LBP	63.81	78.09	86.41	66.56	78.18	85.64	68.35	70.38	87.85
DSTEF+HOG	91.84	92.19	98.74	85.93	89.98	92.77	89.68	90.34	98.79

reduces redundant information, and makes DSTEF+HOG with the same action have stronger similarity, so the final recognition accuracy is greatly improved. The confusion matrixes of DSTEF+HOG feature on Data1 and Data2 are shown in figure 8 and 9 respectively.

We also select three other state-of-the-art methods to make comparison containing DHS [15], R-STDP [16] and MMNN [17]. The average results are shown in table 6.

Table 6. Recognition with different methods (%)

Method	Data1	Data2
DHS	76.52	58.17
R-STDP	81.55	79.25
MMNN	93.86	85.44
DSTEF+DCRC	96.48	91.37

As can be seen from table 6, the recognition accuracy of the proposed method in this paper has been improved compared with other methods.

This paper compares the complexity of DSTEF+DCRC with that of other methods, and the comparison results are shown in Table 7.

In Table 7, f is the frame number of depth map sequence, and the upper limit is 30. W , H and D are width value, height value and depth value of depth map sequence respectively. In this paper, $W = 320$, $H = 240$ and $D = 255$. It can be concluded that the time complexity of DSTEF+DCRC is lower than that of DHS, R-STDP and MMNN.

Table 7. Comparison of computational complexity with different methods (%)

Method	Time complexity
DHS	$O(wh) + O(fdhw)$
R-STDP	$O(wh) + O(fwh)$
MMNN	$O(fwh) + O[(f - 1)(wh + wd + hd)]$
DSTEF+DCRC	$O(fwh) + O(fh + fd + wh)$

5. Conclusions

In this paper, we propose a new deep map sequence feature expression method based on discriminative collaborative representation classifier. It solves the problem of missing time sequence information in feature map generated from deep map sequence. The experiment results show that the key frame algorithm improves the extraction rate of feature map and the recognition accuracy of human action. DSTEF+DCRC not only preserves high recognition accuracy on the positive order action data, but also maintains high recognition accuracy on the inverse order action data. In the future, we will continue to research the action recognition based on deep learning methods and apply them into the practical engineering applications.

Acknowledgments. Harbin Federation of Social Science research project, project title "In the "post-Winter Olympics" era, the ice and snow industry was taken as the endogenous driving force to optimize the industrial aggregation mode in Longjiang".

Availability of data and materials. The data used to support the findings of this study are available from the corresponding author upon request.

Competing interests. The authors declare that they have no conflicts of interest.

References

- Berlin S J, John M. "R-STDP Based Spiking Neural Network for Human Action Recognition," *Applied Artificial Intelligence*, vol. 3, pp. 1-18, 2020.
- Jisi A and Shoulin Yin. "A New Feature Fusion Network for Student Behavior Recognition in Education," *Journal of Applied Science and Engineering*, vol. 24, no. 2, 2021.
- Bobick A F, Davis J W. "The recognition of human movement using temporal templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 3, pp. 257-267, 2001.
- S. Yin and H. Li. Hot "Region Selection Based on Selective Search and Modified Fuzzy C-Means in Remote Sensing Images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 5862-5871, 2020. doi: 10.1109/JSTARS.2020.3025582.
- Y. Zhu, W. Chen and G. Guo. "Fusing Spatiotemporal Features and Joints for 3D Action Recognition," *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Portland, OR*, pp. 486-491, 2013, doi: 10.1109/CVPRW.2013.78.
- Luo. "Feature Extraction and Recognition for Human Action Recognition," *Machine Vision & Applications*, vol. 25, no. 7, pp. 1793-1812, 2014.

7. Xuan Son Nguyen, Thanh Phuong Nguyen and F. Charpillet. "Improving surface normals based action recognition in depth images," *2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Colorado Springs, CO, pp. 109-114, 2016. doi: 10.1109/AVSS.2016.7738053.
8. Q. Nie, J. Wang, X. Wang and Y. Liu. "View-Invariant Human Action Recognition Based on a 3D Bio-Constrained Skeleton Model," *IEEE Transactions on Image Processing*, vol. 28, no. 8, pp. 3959-3972, Aug. 2019. doi: 10.1109/TIP.2019.2907048.
9. S. Chaudhary and S. Murala. "Depth-based end-to-end deep network for human action recognition," *IET Computer Vision*, vol. 13, no. 1, pp. 15-22, 2019. doi: 10.1049/iet-cvi.2018.5020.
10. Mattiev, J., Kavek, B. "Distance based Clustering of Class Association Rules to Build a Compact, Accurate and Descriptive Classifier," *Computer Science and Information Systems*, Vol. 18, No. 3, pp. 791-811. (2021), <https://doi.org/10.2298/CSIS200430037M>.
11. Fan, Z., Guan, Y. "Face Recognition Based on Full Convolutional Neural Network Based on Transfer Learning Model," *Computer Science and Information Systems*, Vol. 18, No. 4, pp. 1395-1409. (2021), <https://doi.org/10.2298/CSIS200922028F>.
12. Chao X, Hou ZJ, Li X, Liang JZ, Huan J and Liu H Y. "Action recognition under depth spatial-temporal energy feature representation," *Journal of Image and Graphics*, vol. 25, no. 04, pp. 0836-0850, 2020.
13. Yang X, Zhang C, Tian Y L. "Recognizing actions using depth motion maps-based histograms of oriented gradients," *ACM International Conference on Multimedia*. ACM, 2012:1057.
14. S. Jia, L. Shen and Q. Li. "Gabor Feature-Based Collaborative Representation for Hyperspectral Imagery Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 2, pp. 1118-1129, Feb. 2015. doi: 10.1109/TGRS.2014.2334608.
15. Baofeng Z, Jun K, Min J. "Human Action Recognition Based on Discriminative Collaborative Representation Classifier," *Laser & Optoelectronics Progress*, vol. 55, no. 1, pp. 257-263, 2018.
16. Md Azher, Uddin, Young-Koo, et al. "Feature Fusion of Deep Spatial Features and Handcrafted Spatiotemporal Features for Human Action Recognition," *Sensors*, vol. 19, no. 7, pp. 1599, 2019. doi: 10.3390/s19071599.
17. Berlin S J, John M. "R-STDP Based Spiking Neural Network for Human Action Recognition," *Applied Artificial Intelligence*, vol. 3, pp. 1-18, 2020. Zhao H, Xue W, Li X, et al. "Multi-Mode Neural Network for Human Action Recognition," *IET Computer Vision*, vol. 14, no. 8, pp. 587-596, 2020.

Wang Yuhang is an associate professor in School of Physical Education, Harbin University. His current research interests include physical education, and behavior analysis.

Tao Feng is a lecturer in Physical Education Department of Harbin Institute of Finance, PhD candidate at The Graduate School Saint Paul University. His research interests include video quality of service over wireless networks, adaptation, perceptual modeling, physical education, and behavior analysis.

Yi Zheng is the director of educational affairs office of Physical Education Institute of Harbin University. His current research interests include Machining and Modeling of physical education, and behavior analysis.

Received: March 21, 2022; Accepted: August 28, 2022.

A Novel Deep LeNet-5 Convolutional Neural Network Model for Image Recognition

Jingsi Zhang, Xiaosheng Yu, Xiaoliang Lei, and Chengdong Wu*

Faculty of Robot Science and Engineering, Northeastern University
Shenyang 110819, China
newmansuper@163.com
yuxiaosheng@mail.neu.edu.cn
xiaolianglei@stumail.neu.edu.cn
wuchengdongnu@163.com

Abstract. At present, the traditional machine learning methods and convolutional neural network (CNN) methods are mostly used in image recognition. The feature extraction process in traditional machine learning for image recognition is mostly executed by manual, and its generalization ability is not strong enough. The earliest convolutional neural network also has many defects, such as high hardware requirements, large training sample size, long training time, slow convergence speed and low accuracy. To solve the above problems, this paper proposes a novel deep LeNet-5 convolutional neural network model for image recognition. On the basis of LeNet-5 model with the guaranteed recognition rate, the network structure is simplified and the training speed is improved. Meanwhile, we modify the Logarithmic Rectified Linear Unit (L-ReLU) of the activation function. Finally, the experiments are carried out on the MINIST character library to verify the improved network structure. The recognition ability of the network structure in different parameters is analyzed compared with the state-of-the-art recognition algorithms. In terms of the recognition rate, the proposed method has exceeded 98%. The results show that the accuracy of the proposed structure is significantly higher than that of the other recognition algorithms, which provides a new reference for the current image recognition.

Keywords: CNN, image recognition, feature extraction, deep LeNet-5, L-ReLU.

1. Introduction

With the development of science and technology, computer vision has been widely used in various fields. The core technologies of these applications are image processing [1], image recognition [2] and classification tasks [3]. The recognition technology is to calculate the characteristics of the samples and apply them to the classifier to generate classification for different calculated values.

Since 1980s, research on optical character recognition methods has always been a hot topic in pattern recognition [4]. It is not easy for a computer to correctly recognize a large number of handwritten fonts because different people have different habits of writing numbers. Therefore, it is of great significance to study an accurate and efficient number recognition method.

* Corresponding author

For image recognition methods, the traditional recognition methods such as support vector machine, traditional neural network, the K-nearest neighborhood method (KNN), have some shortcomings. The minimum distance classification algorithm is a traditional recognition algorithm, but it is not suitable for handwritten fonts. The recognition method of KNN is derived from statistics [5]. The principle is to calculate the features of the image and measure the distance between the calculated results of different features for classification. Its advantage is that it is insensitive to abnormal data collection. SVM has been successfully applied to image recognition [6]. In machine learning, SVM can avoid the complexity of high-dimensional space, and it is very prominent in small sample, high-dimensional space calculation and nonlinear problems. However, in the classification problem, the storage space occupied by solving the function is large. These traditional recognition algorithms mentioned above have very poor expression ability for more complex mathematical functions, poor generalization performance, and usually fail to reach the expected effect of data prediction and accuracy.

The emergence of convolutional neural network (CNN) provides the possibility to solve the generalization ability of image recognition [7,8]. Convolutional neural network, as a successful model in deep learning, has been widely applied in the field of image recognition.

CNN can extract the hidden features of human face by using hardware acceleration technology and massive face image data training. This feature is highly invariant to scale changes such as translation, zoom and tilt, and has certain robustness to complex fonts. Therefore, the research on image recognition based on CNN is very active. For example, the reference [9] achieved the fusion of high and low level features by improving the structure of AlexNet. Reference [10] integrated binary tree with ResNet convolutional neural network, and put forward a binary tree CNN information fusion model for image recognition.

In this paper, we propose a novel deep LeNet-5 convolutional neural network model for image recognition. The network structure of LeNet-5 is improved, and a new activation function L.ReLU is used to solve the over-fitting phenomenon in the training process. The experiment of network structure is carried out through MINIST database to improve the operation speed of network structure and the recognition accuracy of the proposed algorithm.

This paper is organized as follows. In section 2, we give the related works including CNN and LeNet-5. Section 3 presents the proposed image recognition method in detail. Experiments and analysis are conducted in section 4. We make a conclusion for this paper in section 5.

2. Related works

2.1. Structure of CNN

The traditional CNN is generally composed of five parts: the input layer, the convolution layer, the down-sampling layer, the fully connection layer and the output layer [11]. The network structure is shown in figure 1.

For general multi-layer neural networks, the first layer is the eigenvector [12]. In general, the image is processed manually to obtain the feature vector, which is used as the

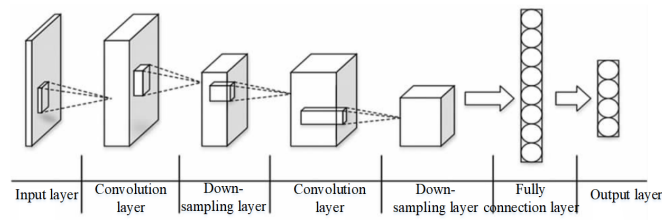


Fig. 1. The Structure of CNN

input of the neural network. The convolution neural network is different from general multi-layer neural network, and the whole image is as the input of the network. For example, the experimental object of this paper is the handwritten digital image in the MINIST database, and the size of the image obtained after processing is 28×28 . In order to facilitate the call and reading of data, the image can be expanded according to the pixel number.

The convolution layer is also the feature extraction layer. Convolution operation is the soul of convolutional neural network, and convolution kernel is the tool of convolution operation[13]. The principle of convolution operation is shown in figure 2.

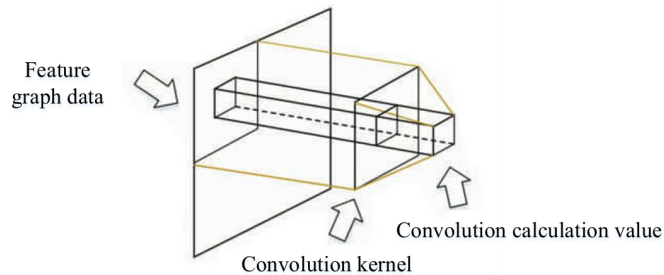


Fig. 2. Principle diagram of convolution operation

The down-sampling layer is also known as the pooling layer. During the pooling of feature graphs, the depth of the image does not change, but the size of the image can be reduced. Pooling can be viewed as converting a high resolution image to a low resolution image with a smaller size. Through multiple pooling layers, the number of parameters in the final fully connection layer can be gradually reduced to reduce the parameters of the whole neural network and improve the training speed. The pooling layer has no parameters that can be trained. The fully connected layer is the same as the normal fully connected layer. Its input layer is the previous feature graph, which will transform all neurons in the feature graph into data in the fully connection layer.

2.2. Structure of LeNet-5

The LeNet-5 convolutional neural network model for image recognition is shown in figure 3, which consists of input layer, hidden layer and output layer. The input layer is a 32×32 single channel target image. The hidden layer is responsible for the extraction and classification of object features. The output layer outputs an integer representing the category.

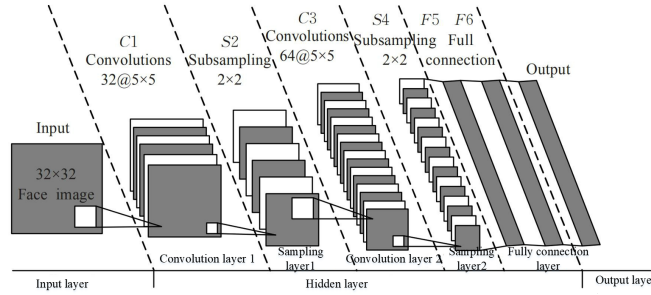


Fig. 3. The LeNet-5 convolutional neural network model

The hidden layer of the convolutional neural network is generally composed of Convolutions, Sub-sampling and Fully connection. The model in figure 3 contains two convolution layers (C1 and C3). C1 and C3 have $32 \ 5 \times 5$ convolution kernels and $64 \ 5 \times 5$ convolution kernels respectively. In the process of image recognition, the convolution kernel and the input image convolve each 5×5 region to extract the feature graphs that are highly robust to scale changes. When the convolution step size is 1, the convolution operation is shown in equation (1):

$$H_{i,j}^s = \sum_{m=0}^k \sum_{n=0}^k W_{m,n}^s X_{m+i,n+j} + b^s. \quad (1)$$

In equation (1), X represents an image with a depth of 1 and a size of $u \times v$. W represents the convolution kernel of $k \times k$. S represents the feature graphs of different convolution kernels. $H_{i,j}^s$ represents the element in row i , column j of the feature graph H^s matrix. b^s is the Bias value corresponding to the convolution kernel.

S2 and S4 are pooling layers. Through multi-layer sampling, CNN can reduce the dimension of feature graph and eliminate repeated features [14], so that features have certain translational invariance. F5 and F6 are fully connection layers similar to traditional multi-layer perceptron neural networks, with 1024 and 67 neurons respectively. The image features obtained by the convolution layer and the sampling layer are dot product with the weight of the multi-layer perceptron neural network, and then the feature classification is completed by the Sigmoid function.

Recently, many researchers developed LeNet-5-based methods to process the images. For example, In reference [15], the recognition of haze images was performed by adjusting the parameters and structure of the classic LeNet-5 model. The image recognition

technology was applied to a haze image field, which showed good performance. Zhang et al. [16] proposed an improved LeNet-5 algorithm for traffic sign recognition. The picture noise elimination and image enhancement on selected traffic sign images were performed. Then, Gabor filter kernel was adopted in the convolution layer for convolution operation. In the convolution process, the normalization layer Batch Normality (BN) was added after each convolution layer and reduced the data dimension. Zhang et al. [17] studied the detection of hyperthyroidism by the modified LeNet-5 network.

The accuracy statistics result is shown in figure 4. After nearly 10000 times training, the recognition accuracy of the network is still far less than 1, indicating that the convergence speed of this network model is slow and its learning ability is poor. The accuracy of the training set is always much higher than that of the test set, which indicates that the network is over-fitting and the generalization ability is poor. Through the above analysis, it can be seen that the LeNet-5 model-based convolutional neural network has poor image recognition effect, and the network needs to be improved.

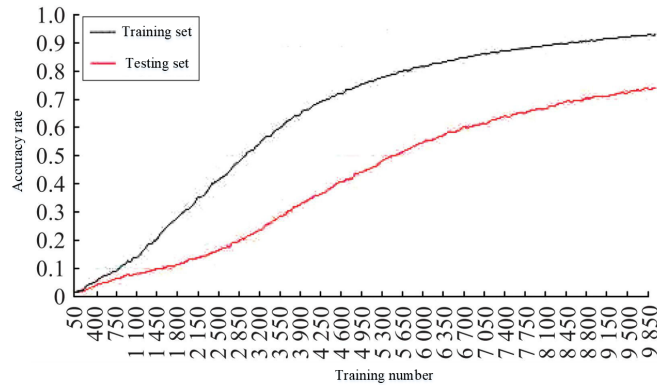


Fig. 4. Training results with LeNet-5 for image recognition

3. Proposed LeNet-FC convolutional neural network

3.1. Activation function optimization

The reason why CNN based on LeNet-5 model has a slow convergence speed in image recognition training is that the Sigmoid activation function appears the gradient disappearance phenomenon when the network is trained by gradient descent method. Sigmoid function is shown in equation (2):

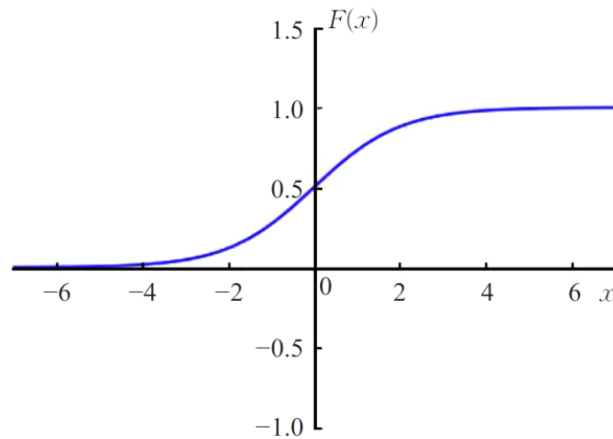
$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}}. \quad (2)$$

In the gradient descent training method, the parameter updating is mainly based on the gradient value to achieve training optimization. This gradient is calculated by backward

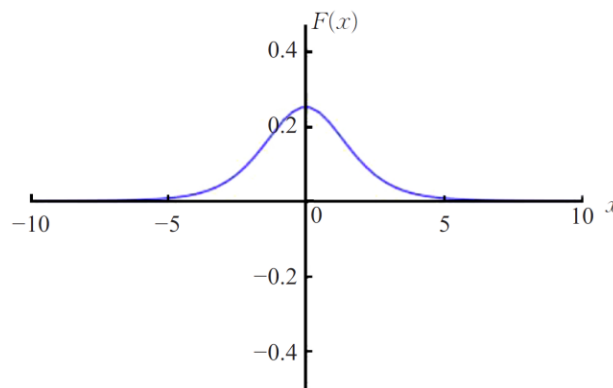
recursion from the output layer based on the error between the recognition result and the real result as shown in equation (3):

$$Grad = Error \times Sigmoid'(x) \cdot x. \quad (3)$$

In equation (3), Grad represents the gradient of the current layer. Error denotes the error. x is the input value of the current layer. $Sigmoid'(x)$ represents the first derivative of the Sigmoid function. According to figure 5(a), the Sigmoid function is double-ended saturated. As shown in figure 5(b), when x value is too large or too small, the first derivative of Sigmoid function will approach 0. This will cause the gradient value to be greatly attenuated in the backpropagation calculation, or even disappear to 0. Therefore, the network parameters are updated with a minimal gradient value in the training, and the convergence speed of the network is slow or even unable to converge.



(a) Sigmoid function graph



(b) First derivative of Sigmoid function

Fig. 5. Sigmoid function and its first derivative graph

In this paper, we present the statistatized Rectified Linear Unit (L_ReLU) in the activation function optimization. Its expression is shown in equation (4):

$$LReLU(x) = \ln\left(\frac{1+e^x}{2}\right) + 0.1x. \quad (4)$$

According to the expression of L_ReLU and the function in figure 6(a), it can be seen that L_ReLU has five basic properties which is necessary to become an activation function:

1. Non-linearity. L_ReLU function is nonlinear, and it can play a good role of nonlinear mapping in CNN.
2. Differentiability. The derivative of L_ReLU function is shown in equation (5):

$$LReLU'(x) = \frac{e^x}{1+e^x} + 0.1. \quad (5)$$

Therefore, the training method based on gradient can be adopted.

3. Monotonicity. $LReLU'(x) > 0$ shows that L_ReLU function is monotonically increasing. This can guarantee that each layer of network in CNN is convex function.
4. $f(x) \approx x$. L_ReLU function satisfies this condition when $x > 0$. The network can be initialized with a small random value to obtain a good training effect.
5. The output value is infinite. The output value of L_ReLU function is infinite. When the model is trained with a small learning speed, it can obtain a higher training efficiency.

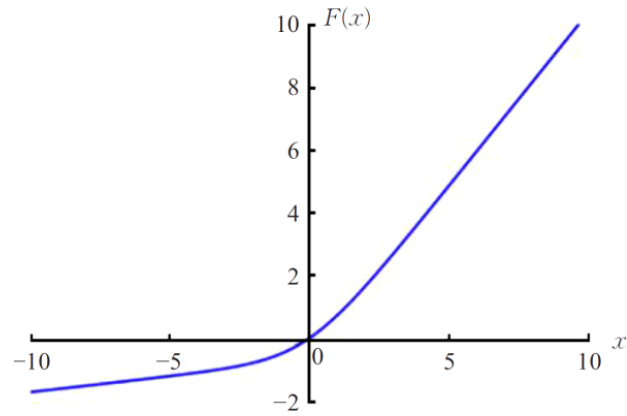
Compared with Sigmoid function, L_ReLU activation function does not appear gradient disappearance in gradient descent method. By computing the limit of the derivative of L_ReLU in equation (5), it can be seen that when x tends to positive infinity and negative infinity, the limits are 1.1 and 0.1 respectively. As shown in figure 6(b), when x is too large, the derivative value of L_ReLU is close to 1.1. When x is too small, its value will be close to 0.1, and will not be 0. Therefore, L_ReLU activation function can be used in CNN to carry out effective gradient descent training.

The Rectified Linear Unit (ReLU), Softplus [18] and the proposed L-ReLU activation function are compared and analyzed. The curves of the three functions are shown in figure 5. The expressions of ReLU and Softplus are in equations (6) and (7) respectively, while the expression of L_ReLU is shown in equation (4).

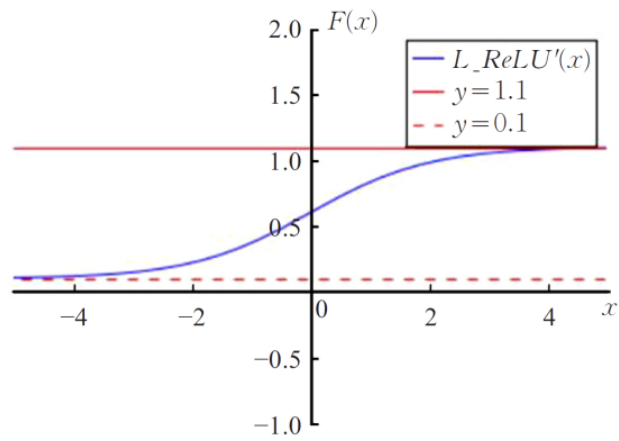
$$ReLU(x) = \max(0, x). \quad (6)$$

$$softplus(x) = \ln(1 + e^x). \quad (7)$$

According to figure 7, it can be seen that the ReLU function has the characteristics of one-sided suppression of negative input values (single-ended saturation, and the output is 0 when negative values are input) and wide excitation boundary (linear mapping for positive input), so the nonlinear mapping is sparsity. According to equation (6), the calculation amount of ReLU is far less than Sigmoid function, and its first derivative is 1, so it will not cause the gradient disappear. The Softplus activation function is also single-ended saturation, so it converges faster than the Sigmoid function. However, it is only a smooth approximation of ReLU, so it does not have sparse activation, and its activation



(a) L_ReLU function graph



(b) First derivative of L_ReLU function

Fig. 6. L_ReLU function and its first derivative graph

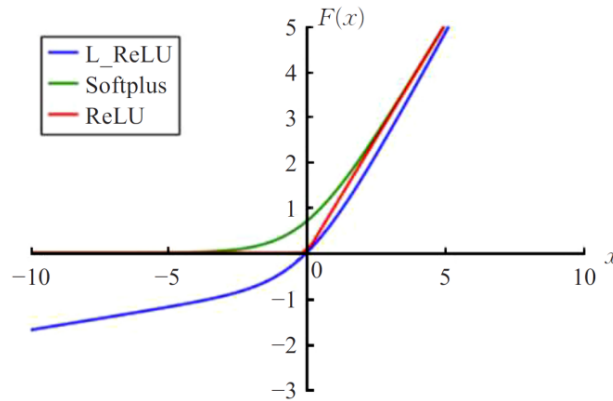


Fig. 7. Three activation function curves

performance is worse than that of ReLU activation function. L_ReLU activation function is double-ended unsaturated, and the gradient will not disappear in the CNN training, so the convergence speed of the network will be faster than the Sigmoid activation function. Moreover, by comparing the L_ReLU in figure 7, it can be seen that L_ReLU is also similar to a smoothing of ReLU, but different from Softplus, and it also has a certain sparse mapping feature, that is, it has an appropriate inhibitory effect on negative input, and an approximate linear mapping for positive input.

In this paper, MNIST data training experiments are carried out on CNNs with the above three activation functions respectively. The convergence of the network is shown in table 1. In table 1, the network convergence speed of L_ReLU and ReLU is higher than that of Softplus, but the convergence speed of L_ReLU is slightly lower than that of ReLU. By comparing equations (4) and (6), it can be seen that the convergence speed of L_ReLU is slightly slower than that of ReLU, because it requires a large amount of calculation.

Table 1. The error of three activation functions in MNIST data training

activation	200	400	600	800	1000	1200	1400
softplus	13.5	13.4	10.1	7.2	6.8	6.1	6.1
L_ReLU	7.7	4.6	3.8	3.7	3.5	3.4	3.4
ReLU	2.8	2.7	2.6	2.5	2.4	2.4	2.4

Although the ReLU has excellent performance, it inputs all negative values into the sparsity map, and two problems are likely to occur in the actual training of CNN: (1) "dead neurons" phenomena appears in the network, that is, the gradient value of the network parameter in this neuron is 0, and it cannot be trained and updated again. (2) it results in the loss of some characteristic information quantized by negative values. By comparing the curves of L_ReLU and ReLU in table 1, it can be seen that L_ReLU realizes the non-zero sparse mapping through the compression of the negative input similar to the

exponential function, which can avoid the above two problems. Therefore, the optimized activation function used in this paper is L_ReLU.

3.2. Structure optimization

It is found that artificial neural networks with appropriate multiple hidden layers can learn the essential features of data and classify data effectively. However, increasing the network depth too much will reduce the performance of the neural network. Therefore, one direction of improving CNN network structure is to properly increase the convolutional layer, while the other direction is to adjust the size of the convolutional kernel. This paper conducts performance testing on CNN with nine different structures, and the results are shown in table 2.

Table 2. CNN performance comparison with nine different structures

No.	Convolutional layer number	Kernel	Training rate	Testing rate
1	2	2×2	1.00	0.89
2	2	3×3	1.00	0.91
3	2	4×4	1.00	0.93
4	3	2×2	1.00	0.84
5	3	3×3	1.00	0.92
6	3	4×4	1.00	0.92
7	4	2×2	0.95	0.61
8	4	3×3	0.98	0.83
9	4	4×4	1.00	0.86

According to table 2, under the condition of the same size of convolution kernel, CNN with three convolution layers has the best accuracy in the testing set. For CNN with the same number of convolutional layers, 3×3 convolutional kernel CNN has better performance. Considering that the performance difference is not large and the scale of CNN is small, the network structure is improved with three convolution layers in this paper. The convolution kernel of each layer is 3×3 .

By comparing the accuracy of the network in the training set and the testing set in table 1, it can be seen that although the improvement of network structure has improved the performance of CNN, the problem of over-fitting still exists in the network. The reason for the over-fitting is that the parameters in the fully connection layer are updated completely according to the feature recognition results of the training data. The classification of training data is "overlearned", which leads to the failure to accurately classify the test data. An effective solution is the Dropout technology. Therefore, Dropout technology is adopted in this paper, where the parameter is set as 0.7, that is, in the training process of deep learning network, the parameter is temporarily dropped from the network with a probability of 0.7 without training update, so as to improve the network generalization ability.

In addition to the above improvements, the new CNN model also carries out the following changes.

1. Single channel image input with a high resolution of 68×68 is adopted, so that CNN can extract deeper, high-scale invariant and strong robustness implicit features of the image.
2. During the convolution, zero padding with the value of 1 is carried out first, and a layer of boundary with the value of 0 is added to the input image to enhance the extraction of image edge and contour features.
3. The sampling layer is improved by using 2×2 maximum pooling to enhance sparse expression of features.
4. The output layer is improved by computing the Softmax function as shown in equation (8):

$$p_j = \frac{e^{f_{y_i}}}{\sum_j e^{f_j}}. \quad (8)$$

In equation (8), y_i is defined as the label of the i -th input feature. f_j represents the j -th element of the output vector f in the output layer. P_j represents the probability that the input feature belongs to the j -th class.

3.3. Performance analysis of LeNet-FC model

The proposed LeNet-FC model is shown in figure 8. The image recognition training of LeNet-FC model adopts sparse data labels, that is, the labels (categories) of the original training data set and test data set conducts one-hot coding according to table 3 before training [19]. Then, the Adam optimization algorithm is used to minimize the value of cross entropy loss function, as shown in equation (9):

$$Loss = -\frac{1}{N} \sum_i^N lb(p_i + \varepsilon), \varepsilon = 1 \times 10^{-10}. \quad (9)$$

In equation (9), N is the number of input samples during each training. p_i is the Softmax function output value corresponding to each sample. The function of ε is to prevent the occurrence of $L = \pm\infty$ when $p_i = 0$ leading to the termination of training.

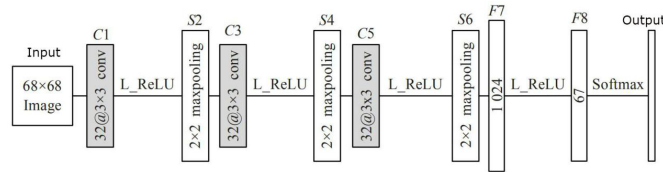


Fig. 8. Proposed LeNet-FC model

This paper conducts experiments on the LeNet-FC model and the LeNet-5 model. The results are shown in figure 9. According to figure 9, the convergence speed of LeNet-FC convolutional neural network is significantly faster than that of CNN based on LeNet-5 model. At the same time, the accuracy change curves of the new LeNet-FC model in the

Table 3. Font sizes

Category label	0	1	2
Coding	[1, 0, 0, ..., 0, 0, 0]	[0, 1, 0, ..., 0, 0, 0]	[0, 0, 1, ..., 0, 0, 0]
...	64	65	66
[0, 0, 0, ..., 1, 0, 0]	[0, 0, 0, ..., 0, 1, 0]	[0, 0, 0, ..., 0, 0, 1]	[0, 0, 0, ..., 1, 0, 0]

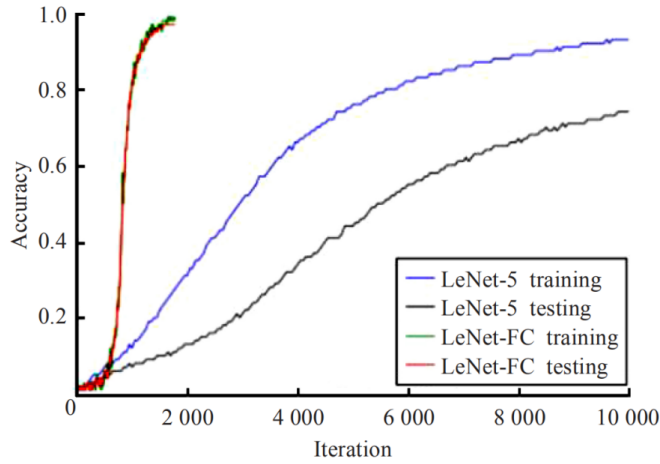


Fig. 9. Training comparison with the two CNN models

training data and test data almost coincide, which indicates that the CNN has a strong ability of image recognition generalization, and there is no over-fitting phenomenon.

As can be seen from table 4, compared with other improved CNN models, the proposed LeNet-FC in this paper has a good performance in image recognition. At the same time, because the other two models are based on AlexNet [20] and ResNet [21] respectively, their models have large size and many parameters, so they need to use hardware with stronger computing performance, such as GPU to perform well. The new model not only has good performance, but also has small scale and low requirement on hardware, and can even be applied to some embedded devices. Therefore, the application of LeNet-FC model is more extensive.

Table 4. Font sizes

Model	Accuracy/%
LeNet-FC	98.9
AlexNet[20]	98.1
ResNet[21]	95.6
DCGAN[22]	92.5
SCNN[23]	94.7
PAMSGAN[24]	95.1

3.4. Experiments and Analysis

The MNIST dataset is from the National Institute of Standards and Technology [25]. The training set is composed of numbers handwritten by 250 different people, some samples of which are shown in figure 10. It contains a total of 70000 images including 60000 training images and 10000 testing images. The images of training set and test set are not repeated. 50% are high school students, and 50% are workers at the Census Bureau. The test set is also handwritten digital data with the same proportion, which is 28×28 pixel data set with labels.



Fig. 10. Some samples of MNIST

The CPU parameters of this experiment are Intel(R) Core(TM) i78700 3.20GHz, 8GB memory, Windows10 system, 64-bit operating system. Anaconda is used to simulate the development environment of TensorFlow. In the experiment, the learning rate is set as 1×10^{-4} , and the cross-entropy cost function is used to update the network parameters. During training, dropout is added to avoid over-fitting problems. Batch size=50. An iteration is defined that it traverses all the training sets one time. The initial variable is assigned with a normal distribution with a standard deviation of 0.1. The final prediction results are output after 50 training times in total, and then the model training data is analyzed and verified.

The structure of the image recognition based on LeNet-FC model consists of three parts: image preprocessing, feature extraction and Euclidean distance comparison. The image preprocessing includes the grayscale and normalization, and its purpose is to change the input image X into a single channel image with a resolution of 68×68 , so that it is consistent with the input of LeNet-FC model.

Feature extraction is realized by the convolution-sampling layer in LeNet-FC model, and the obtained dimension of feature vector is 17776. The eigenvector of the measured number is denoted as Y . The average feature vector of N standard number images in the same category is denoted as the standard feature Y^S . The calculation formula is shown in equation (10):

$$Y^S = \frac{1}{N} \sum_{i=0}^N Y_i^S. \quad (10)$$

Where s stands for different numbers. By adding or deleting features in the library, the number of recognized categories can be changed.

The Euclidean distance comparison mainly calculates the Euclidean distance between the feature vectors to be recognized extracted from the measured number and all the feature vectors in the standard feature library. The expression is $d = \sqrt{(Y - Y^S)^2}$. Then, the

function $f(x) = \text{argmid}(d)$ outputs the index T corresponding to the minimum value in vector d , namely, the category. Threshold value η is an important parameter in Euclidean distance. If $d \leq \eta$, then the algorithm will output the result. Otherwise, it needs to re-input the number image.

In the test, the numbers of correct recognition, false recognition and rejection recognition are counted. The corresponding accuracy α , error recognition rate β , and rejection recognition rate δ are calculated [25], and the expressions are expressed as equations (11)-(13) respectively. The test results are shown in table 5.

$$\alpha = \frac{\text{correctly detected and } d \leq \eta}{\text{the tested total number}}. \quad (11)$$

$$\beta = \frac{\text{error detected and } d \leq \eta}{\text{tested total number}}. \quad (12)$$

$$\delta = \frac{\text{number of } d > \eta}{\text{tested total number}}. \quad (13)$$

Table 5. Font sizes

η	Correctly detected number	Error detected number	Rejected detected number	α	β	δ
0.05	130	0	70	0.66	0	0.34
0.06	150	0	50	0.76	0	0.24
0.07	165	0	35	0.84	0	0.16
0.08	177	1	22	0.89	0.01	0.11
0.09	182	1	17	0.92	0.01	0.09
0.10	186	2	12	0.94	0.01	0.06
0.11	190	2	8	0.95	0.01	0.04
0.12	190	6	4	0.95	0.03	0.02
0.13	191	7	2	0.96	0.04	0.01
0.14	191	8	1	0.96	0.04	0.01
0.15	191	8	1	0.96	0.04	0.01

Through the analysis of the test results in table 5, it can be seen that the image recognition based on LeNet-FC model has high recognition accuracy and relatively low error recognition and rejection rate. In order to test and compare the performance of the improved deep neural network, the traditional LeNet-5 structure is used to adjust the output structure, and the same data training set is also used for simulation comparison. In the traditional network model, a convolution layer and a fully connection layer are added, and the dropout method is not added to prevent over-fitting. The simulation results are shown in table 6 and table 7. We also make comparison with other recognition methods. It can be seen from tables 6, 7 that the recognition rates of LeNet-FC and traditional LeNet neural network are 98.6% and 97.8%, respectively. Compared with the traditional LeNet recognition rate, the recognition performance of LeNet-FC is better.

By comparing tables 6, 7, it can be seen that with the increase of iteration number, the accuracy rate is also continuously improved. Finally, the network gradually reaches

Table 6. Structure recognition rate of LeNet-FC and traditional LeNet neural network (%)

Iteration number	LeNet-FC	LeNet	DCGAN[22]	SCNN[23]	PAMSGAN[24]
10	96.8	96.2	94.8	95.7	95.9
20	97.1	96.9	93.1	95.7	94.6
30	98.6	97.3	94.1	95.7	95.5
40	98.6	97.5	91.2	95.6	96.7
50	98.6	97.8	96.6	94.7	95.8

Table 7. Average cross entropy error (%)

LeNet	LeNet-FC
1.42	0.89

the state of convergence. In terms of convergence effect, the convergence effect of the improved neural network structure tends to be stable, and the final recognition rate is 98.6% after 50 iterations. However, in the training of traditional LeNet, the recognition rate increases with the iteration number, presenting an unstable state with relatively large fluctuation of recognition. In the case of 50 iterations, the final recognition rate is 98.6%.

Meanwhile, this experiment also studies the impact of batch size input on the recognition rate. The single Batch is set as 50, 100 and 200, and other conditions in the experiment remain unchanged. The results are shown in table 8. The number of training iterations is 50, and every 5 iterations show the current training recognition accuracy.

Table 8. Test recognition rate of different batch conditions

No.	5	10	15	20	25
50batches	98.11	98.69	98.89	98.97	99.03
100batches	97.70	98.38	98.68	98.81	98.92
200batches	96.84	97.99	98.38	98.55	98.67
No. 30	35	40	45	50	
50batches	99.12	99.11	99.28	99.22	99.25
100batches	98.88	99.09	99.13	99.09	99.15
200batches	98.79	98.79	98.95	98.97	98.88

As can be seen from table 8, after 35 iterations, the 50 batches, 100 batches can achieve recognition accuracy of more than 90%. With the increasing number of iterations, the change of recognition rate becomes stable, so it can be considered that the network model reaches the convergence state at this time. Where, the fastest convergence rate is at 50batch, and the slowest is at 200batch. After about 45 iterations, the recognition accuracy will fluctuate around 99%, and does not improve in 50 iterations. Overall, the recognition rate of 200batch model is lower that of 50batch and 100batch model in the preliminary data training. After 50 iterations of the 100 batches model, the recognition accuracy fluctuates around 99.1%. In the training process, the training speed of 50 batches is not the fastest.

4. Conclusion

On the basis of LeNet-5 neural network, the structure of the network is improved, which greatly reduces the number of neuron parameters, improves the training time, increases the number of feature extraction layers, and improves the recognition accuracy. The LeNet-FC model shows strong generalization ability in image recognition training. The comparative experiment shows that the proposed method greatly improves the recognition effect. In the future recognition training experiments, appropriate parameters should be selected according to the size of training batch, so as to improve the recognition rate and the training speed as much as possible.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China under Grant nos. U20A20197, 61973063, U1713216, 61901098, 61971118, Liaoning Key Research and Development Project 2020JH2/10100040, the China Postdoctoral Science Foundation 2020M670778, the Northeastern University Postdoctoral Research Fund 20200308, the Scientific Research Foundation of Liaoning Provincial Education Department LT2020002, the Foundation of National Key Laboratory (OEIP-O-202005) and the Fundamental Research Fund for the Central Universities of China N2026005, N181602014, N2026004, N2026006, N2026001, N2011001.

Availability of data and materials. The data used to support the findings of this study are available from the corresponding author upon request.

Competing interests. The authors declare that they have no conflicts of interest.

References

1. Maruo S, Fujishiro Y, Furukawa T. "Simple autofocusing method by image processing using transmission images for large-scale two-photon lithography," *Optics Express*, vol. 28, no. 8, 2020.
2. Chen J, Zheng H, Xiong H, et al. "FineFool: A Novel DNN Object Contour Attack on Image Recognition based on the Attention Perturbation Adversarial Technique," *Computers & Security*, vol. 9:102220, 2021.
3. Shoulin Yin, Hang Li, Desheng Liu and Shahid Karim. "Active Contour Modal Based on Density-oriented BIRCH Clustering Method for Medical Image Segmentation," *Multimedia Tools and Applications*, vol. 79, pp. 31049-31068, 2020.
4. Khan M A, Rizvi S, Abbas S, et al. "Deep Extreme Learning Machine-Based Optical Character Recognition System for Nastalique Urdu-Like Script Languages," *The Computer Journal*, vol. 65, no. 2, pp. 331-344, 2022.
5. Murata M, Kanamaru T, Shirado T, et al. "Automatic F-term Classification of Japanese Patent Documents Using the k-Nearest Neighborhood Method and the SMART Weighting," *Information & Media Technologies*, vol. 14, no. 1, pp. 163-189, 2007.
6. Xia, B., Han, D., Yin, X., Gao, N. "RICNN: A ResNet & Inception Convolutional Neural Network for Intrusion Detection of Abnormal Traffic," *Computer Science and Information Systems*, vol. 19, no. 1, pp. 309-326, 2022.
7. Gorban A N, Mirkes E M, Tukin I Y. "How deep should be the depth of convolutional neural networks: a backyard dog case study," *Cognitive Computation*, vol. 12, no. 1, pp. 388-397, 2020.

8. Kim M J, Yi L, Song H O, et al. "Automatic Cephalometric Landmark Identification System Based on the Multi-Stage Convolutional Neural Networks with CBCT Combination Images," *Sensors*, vol. 21, no. 2, pp. 505, 2021.
9. X. Yu, W. Long, Y. Li, X. Shi and L. Gao. "Improving the Performance of Convolutional Neural Networks by Fusing Low-Level Features With Different Scales in the Preceding Stage," *IEEE Access*, vol. 9, pp. 70273-70285, 2021.
10. Wen L, Li X, Gao L. "A transfer convolutional neural network for fault diagnosis based on ResNet-50," *Neural Computing and Applications*, vol. 32, pp. 6111-6124, 2020.
11. Kg A, Nc A. "Analysis of Histopathological Images for Prediction of Breast Cancer Using Traditional Classifiers with Pre-Trained CNN - ScienceDirect," *Procedia Computer Science*, vol. 167, pp. 878-889, 2020.
12. nan Güler a, B E B. "Expert systems for time-varying biomedical signals using eigenvector methods," *Expert Systems with Applications*, vol. 32, no. 4, pp. 1045-1058, 2007.
13. Glorot X, Bordes A, Bengio Y. "Deep Sparse Rectifier Neural Networks," *Journal of Machine Learning Research*, vol. 15, pp. 315-323, 2011.
14. Gao S. "A Two-channel Attention Mechanism-based MobileNetV2 And Bidirectional Long Short Memory Network For Multi-modal Dimension Dance Emotion Recognition," *Journal of Applied Science and Engineering*, vol. 26, no. 4, pp. 455-464, 2022.
15. Fan Y, Rui X, Poslad S, et al. "A better way to monitor haze through image based upon the adjusted LeNet-5 CNN model," *Signal Image and Video Processing*, vol. 14, no. 2, 2020.
16. Zhang C, Yue X, Wang R, et al. "Study on Traffic Sign Recognition by Optimized Lenet-5 Algorithm," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 34, no. 1, pp. 2055003.1-2055003.21, 2020.
17. Zhang Q, Hu X, Zhou S. "The Detection of Hyperthyroidism by the Modified LeNet-5 Network," *Indian Journal of Pharmaceutical Sciences*, vol. 82, 2020.
18. A. Senior and X. Lei. "Fine context, low-rank, softplus deep neural networks for mobile speech recognition," *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7644-7648, 2014.
19. F. Jafarzadehpour, A. Sabbagh Molahosseini, A. A. Emrani Zarandi and L. Sousa. "Efficient Modular Adder Designs Based on Thermometer and One-Hot Coding," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 27, no. 9, pp. 2142-2155, 2019.
20. Sarp, S., Kuzlu, M., Zhao, Y., Cetin, M., Guler, O. "A Comparison of Deep Learning Algorithms on Image Data for Detecting Floodwater on Roadways," *Computer Science and Information Systems*, vol. 19, no. 1, pp. 397-414, 2022.
21. Wu Z, Shen C, Hengel A. "Wider or Deeper: Revisiting the ResNet Model for Visual Recognition," *Pattern Recognition*, vol. 90, pp. 119-133, 2019.
22. L. Sun, K. Liang, Y. Song and Y. Wang. "An Improved CNN-Based Apple Appearance Quality Classification Method With Small Samples," *IEEE Access*, vol. 9, pp. 68054-68065, 2021.
23. M. Zhang, M. Gong, H. He and S. Zhu. "Symmetric All Convolutional Neural-Network-Based Unsupervised Feature Extraction for Hyperspectral Images Classification," *IEEE Transactions on Cybernetics*, vol. 52, no. 5, pp. 2981- 2993, 2022.
24. Z. Zhang. "PAMSGAN: Pyramid Attention Mechanism-Oriented Symmetry Generative Adversarial Network for Motion Image Deblurring," *IEEE Access*, vol. 9, pp. 105131-105143, 2021.
25. S. B. Ahmed, I. A. Hameed, S. Naz, M. I. "Razzak and R. Yusof. Evaluation of Handwritten Urdu Text by Integration of MNIST Dataset Learning Experience," *IEEE Access*, vol. 7, pp. 153566-153578, 2019.
26. Chuang Bai, Xiang Chen. "Research on New LeNet-FC Convolutional Neural Network Model Algorithm," *Computer Engineering and Applications*, vol. 55, no. 5, pp. 105-111, 2019.

Jingsi Zhang is with Faculty of Robot Science and Engineering, Northeastern University, Shenyang 110819, China. His research interests include image processing and robot.

Xiaosheng Yu is with Faculty of Robot Science and Engineering, Northeastern University, Shenyang 110819, China. His research interests include image processing and robot.

Xiaoliang Lei is with Faculty of Robot Science and Engineering, Northeastern University, Shenyang 110819, China. His research interests include image processing and robot.

Chengdong Wu is with Faculty of Robot Science and Engineering, Northeastern University, Shenyang 110819, China. His research interests include image processing and robot.

Received: January 20, 2022; Accepted: August 29, 2022.

Application of Wearable Motion Sensor in Business English Teaching

Dan Lu¹ and Fen Guo^{2,*}

¹ School of Foreign Languages, Shunde Polytechnic,
Shunde 528300, Guangdong, China
rachellu2009@126.com

² School of Software Engineering, South China University of Technology,
Guangzhou, 510006, China
csguofen@scut.edu.cn

Abstract. With the advancement of science and technology, portable motion sensors are becoming increasingly popular in life and have become a research point for improving life and learning, and are widely used in medical care and smart terminals. Based on the advantages of portable motion sensors, this paper focuses on its application in learning English business. Collects speech information through special motion sensors, analyzes the accuracy of students' reading through speech recognition, to help students better learn business English. Firstly, the wearable sensor is used to collect and preprocess the speech information of students' business English reading as the input of speech recognition. Secondly, the linear predictive cepstrum coefficient (LPCC) and Meier frequency cepstrum coefficient (MFCC) of students' business English reading speech are extracted, and the mixed parameters of LPCC and MFCC are taken as speech features. Finally, the correctness of reading speech is recognized by combining HMM and WNN. Through the simulation analysis of students' reading speech recognition, it is shown that the speech recognition based on wearable motion sensor is feasible and the recognition method has good performance. In addition, the feasibility of wearable motion sensors in business English teaching is verified by the establishment of an experimental classes, which can promote students' English learning better.

Keywords: Wearable motion sensor, Business English, Business English Teaching.

1. Introduction

Judging by the current situation, based on the professional curriculum of the English company in the country, especially the higher vocational education that focuses on the training of qualified talents, the current curriculum does not meet the respective national standards. Vocational training and basic knowledge go beyond their basic needs. Continuing this strategy, a large number of key skills and foreign trade staff will have the opportunity to work and study abroad. For them, having the same language expertise

* Corresponding author

is a fundamental requirement of foreign business and entrepreneurship, and needs to be improved as well. In this economic environment, the ability of "foreign languages + skills" is a necessary demand. With the global economic integration and the rapid development of China's opening process, China has steadily lost touch with other countries around the world. Learning foreign languages, and especially learning English, has become an essential means of living and working for people. With the rapid growth of the demand for learning English, many language schools, teaching tools and teaching methods are appearing in an endless stream. However, teaching and learning spoken English has always been a difficult problem for Chinese to learn English. The main reasons lie in the following two aspects: on the one hand, the characteristics of Chinese pronunciation are quite different from those of English pronunciation, which makes Chinese learners of foreign languages make many pronunciation errors that are difficult to detect or even impossible to detect under the influence of their mother tongue. On the other hand, there is a great lack of qualified oral English teachers in China. Even primary and secondary schools in large and medium-sized cities lack their own pronunciation standards and English teachers who can accurately guide oral English learning. However, general media teaching can only be imparted unilaterally, but not in accordance with the specific situation of students. Teachers and students interact in oral teaching, so it can not play a very effective role.

The current computer-aided language learning system [1], mostly focuses on the learning of words and grammar. Only a few oral English learning programs, its function is relatively uniform, can only give students a whole pronunciation score. However, due to their own level limitations, it is difficult for self-learners to find errors and correct incorrect pronunciation by themselves. The application of speech recognition technology makes the software have the function of correcting pronunciation errors. It can help learners correct pronunciation errors in time and avoid repetition of errors in habits. Greatly improving the efficiency of learners' oral English learning will achieve great social benefits and market value.

Looking at the current domestic research on curriculum design, it is found that the theoretical results of its research are mainly concentrated in three aspects: curriculum design principles, curriculum design analysis, and curriculum design suggestions. Once pointed out that the speciality of the business English professional curriculum design lies in the "complex professional pain". He believes that in addition to completing the corresponding development concepts and training objectives, in addition to completing the corresponding development concepts and training goals, the complex pain is the key, and the important thing is to change the teaching model of "learning business through English and learning English through business". Speech recognition applied in business English teaching can achieve business English teaching reform and improve the effect of English learning. However, the current system of teaching speech recognition in English is limited to computers. People have to learn business English through software on computers or terminals, which makes it incapable of meeting the needs of people learning business English anytime, anywhere. Therefore, people urgently need to develop a portable device to facilitate people to learn business English with interest at any time and anywhere. Wearable motion sensors provide technical conditions for this purpose.

Wearable Motion Sensor [2-4] was developed from wearable computing. From the 1960s to 1970s, wearable computing was in its infancy. In 1966, a wearable computer

for roulette was developed by Thorp and Shannon MIT students. It was the first wearable device in history. After the 21st century, wearable computing has made remarkable progress and entered the vision and life of ordinary people. The rapid development of computer information technology provides a good technological basis for the development of wearable devices. In addition, the demand for more free, healthier, and personalized information processing opens up the application market of wearable devices.

Examining the bibliography for English business course design at home and abroad, it can be seen that there is little in-depth research for a course in the research field. Wearable Computing is a new computer technology, which is very different from traditional computer technology. It breaks the traditional interactive mode, makes people and computers become one, and improves the overall human interaction and computing ability. It provides a ubiquitous way of computing and interaction from time to time, and allows users to use the computer while freeing up their hands or one-handed operation of other tasks. The reduction and mobility of computing devices, the anytime and anywhere nature of information and tasks make the interactive computing model develop towards mobility, accessibility, naturalness, and simplicity. Compared with the traditional interactive mode, the interactive mode of the wearable computer requires high flexibility and real-time. Simple interactive operation is also necessary (for example, a wearable computer GUI strives for simplicity and naturalness). Wearable computing technology involves many disciplines, such as ergonomics, advanced interactive technology, electronic engineering, advanced material technology, etc. It is a very complex computer technology. It is in the early stages of development, and various key theories and technologies are in the process of being perfected and developed.

Wearable computing technology enlarges the scope of human-computer interaction to a certain extent. Specific interactive tasks and devices promote the research of new interactive technologies. For example, in certain ways and environments, when interactive devices such as handwriting and voice cannot be used, it can support other interactive media, such as eye movement functions. Wearable computing makes the human-machine relationship very close. At the same time, because all kinds of equipment are equipped with the human body, their installation, location, shape, and operation convenience should be closely combined with the natural attributes of human, forming a comprehensive and harmonious human-machine interface. This poses a new challenge to the research of the new generation of human-computer interaction.

Through wearable computing technology, the data needed can be accessed at any time. When some emergencies occur, wearing a computer can save a lot of time, such as not having to rush to the phone or work next to the computer. In the military, this demand is more urgent. It is important to receive military information in a timely manner and to receive proper judgment in military operations. In the field of medicine, doctors sometimes fail to record the examination results immediately in front of patients because of various reasons when they treat patients, which easily leads to omissions afterwards and affects the diagnosis of the disease. The wearable computing system can solve the above problems. In addition, if the diagnostic equipment is embedded in the wearing system, it can further improve the diagnostic accuracy and reduce the occurrence of human errors. In English learning, wearable motion sensors can collect students' speech information anytime and anywhere, and then analyze the correctness of students' pronunciation to provide good guidance.

We can say that laptops and their technology represent an important direction in the future development of computers, so that computers can provide better facilities to users in the future and become real people and tools. General Chat Chat Lounge It also reduces human-computer coordination, reduces human-computer interaction, and completes human-computer integration. Industry experts have said that "research that is helpful in reducing human-machine interaction is essential." Technology should be integrated with the application. Particularly because portable computing technology and speech recognition technology are likely to have high research value and application, it is very important to research on speech recognition technology based on portable computing platform. The professional English learning system in this article focuses on Motion Sensor Speech Recognition.

This paper mainly collects and preprocesses the voice information of students' business English reading through wearable motion sensors as the input of the speech recognition system, and puts forward voice features through the speech recognition system, then judges the accuracy of students' reading voice, and then provides pronunciation guidance. The specific contributions of this paper are as follows:

- (1) Wearable sensors are used to collect and preprocess the speech information of students' business English reading as an input of speech recognition.
- (2) Linear Predictive Cepstrum Coefficient (LPCC) and Meier Frequency Cepstrum Coefficient (MFCC) are extracted from students' business English reading pronunciation, and their mixed parameters are used as phonetic features.
- (3) The combination of hidden Markov model (HMM) and wavelet neural network (WNN) can identify the correctness of students' reading pronunciation and guide students' pronunciation.

2. Some Methods are Proposed

2.1. Acquisition and Preprocessing of Speech Information

In this paper, the motion sensor is made into a wearable form, which is convenient for students to carry with them. Then, the spoken speech information of students is collected by a wearable motion sensor. It mainly converts the vibration energy of the sensor into current energy, which passes through the signal amplifiers, gain devices, sampling device, and A/D conversion devices, and finally converts into the digital signal stored in the computer. The WAV format is then used as the signal storage format. The WAV format file is also called a waveform file. It does not change the sampling amplitude of the original sound signal, but adds some control information to it. This format file is convenient for later data analysis and processing.

After obtaining the spoken speech information, it is necessary to preprocess the speech information before entering the speech recognition system. The preprocessing operations adopted in this paper are pre-emphasis [5, 6], frame windowing [7], and endpoint detection [8, 9].

Signal preemphasis.

The collection of audio compositions is the same as the filter function. In the audio signal, the power spectrum of many signals is proportional to the frequency and the energy is concentrated in the low frequency range, thus reducing the signal to noise ratio in the high frequency band, and data loss in the signal transmission process. Pre-emphasis uses the difference between signal characteristics and noise characteristics to process the signal. It amplifies the high-frequency section with low energy at the input end of the signal, so that the entire spectrum from low frequency to high frequency becomes flat and easy to analyze, and then performs the reverse processing at the output end. In the process of processing, the high frequency component of noise is correspondingly weakened, so the problem of signal-to-noise ratio decrease in the high frequency band of signal is effectively solved. Pre-emphasis is the pre-emphasis function processing of all input signals. The commonly used functions are:

$$H(z) = 1 - \mu z^{-1}, z \in (0.9, 1) \quad (1)$$

This function is expressed as the relationship between input and output:

$$y(n) = x(n) - \mu x(n-1) \quad (2)$$

Signal framing and windowing.

All signals require frame processing and the amount of audio signal data will increase as time goes on. When the computer processes the sound signal, it is unrealistic to perform an arithmetic process on an infinitely long sound signal. Therefore, it is necessary to take a limited time segment for analysis, that is, framing [10, 11]. The voice of students' spoken language is continuous, and its voice has a certain periodicity, which can be regarded as short-term stable. Therefore it can be subframe processed to speed up the computer operation and reduce the occupancy of memory space. In the process of framing, a second is usually divided into 33 to 100 frames, that is, each frame is 10 to 30 ms long. To ensure a smooth transition between two consecutive frames, two adjacent frames should overlap each other. This part is defined as a frame shift, and its length is generally half of the frame length. Although the overlap increases the burden of data processing, it also improves the accuracy of signal processing.

To reduce the leakage of spectrum energy, different interception functions are used to truncate the signal. This function is called window function [12-14]. The process of processing with a window functions is called windowing, that is, multiplying the window function by the sound signal.

Fast Fourier Transform (FFT) [15-17] is a necessary step for processing and analyzing acoustic signals. When processing frame signals with FFT, if the frame signal does not have periodicity, then after processing with FFT, an additional energy distribution will appear at both ends of the frame signal, resulting in deviation of the results, and the window function can solve the periodicity problem.

Correct selection of the window function can play an auxiliary role in signal analysis and processing [18]. The most window function for sound signal windowing is a rectangular window [19-20], but the essence of this window function is not windowing, and it does not play an auxiliary role in the FFT process of sound signal. In this paper, Hamming window function is selected to process the sound signal. Hamming window

meets the requirements of the sound signal. Its advantages are that the main lobe is widened and reduced, the side lobe is reduced relatively significantly, the change of both ends of the window is smooth and smooth, and the truncation effect of the signal frame is effectively avoided. The Hamming window function formula is as follows:

$$w(n) = \begin{cases} 0.54 - 0.46(2\pi n / (N-1)), & 0 \leq n \leq N-1 \\ 0, & 0 \end{cases} \quad (3)$$

Where N is the frame length.

Signal endpoint detection.

In the process of voice signal recognition, endpoint detection is an essential key link. Accurate identification of the starting and ending points of spoken reading voice in voice signals can reduce misjudgments in the process of recognition and increase the recognition rate of the recognition system. Therefore, the quality of the endpoint detection method directly affects the performance of the recognition system. Under the time-domain characteristics of sound signals, the noise of the surrounding environment is persistent. When the spoken reading sound appears, the collected sound contains the reading sound and the environmental noise, that is, the sum of the two, that is, the short-term energy and the short-term zero-crossing rate of the reading sound segment are higher than those of the silent segment. However, it is not effective to detect the endpoint only by a single short-term energy or short-term zero-crossing rate. For example, the wind noise generated by a small wind will cause a short-term energy increase. To avoid the misjudgment of single feature recognition, we can use the method of combining two features to judge, that is, setting threshold values for both features at the same time and setting two threshold values.

This method of setting two thresholds simultaneously for short-time energy and short-time zero-crossing rate is called double-threshold endpoint detection method. Its principle is that when the short-time energy or short-time zero-crossing rate of the signal has a value beyond the low threshold, the endpoint detection enters the preparatory stage. However, it is not certain that this point is the signal endpoint. Only when one or two of the two features exceed the high threshold at the same time, can this point be determined as the starting point. Conversely, the process of determining the end point is to first compare whether the feature is below the high threshold, and then determine whether it is below the low threshold.

Business English.

Mental health can be set in the public elective courses, and English reading, business, English translation, English-speaking country overview, international trade, geography, exhibition business, business planning, archives management, and other courses can be added to the professional elective courses. And according to the needs of the students, the communication theory courses can be expanded appropriately to cultivate their own qualities. In the basic training courses, such as comprehensive training of business, English correspondence writing, comprehensive training of international business, document preparation, comprehensive training of business, English business, etc., these

should be arranged in a dedicated training week for completion. For the special training week for orientation courses, it should also be added in an appropriate amount at the right time according to the needs of the school curriculum. In addition, the school also needs to work hard to introduce corporate culture and gradually build a high-level curriculum structure with a steady development of school-enterprise cooperation.

Business English is a covering term, which originated from ESP. Like other categories of ESP Business, English is a definition of a special corpus that emphasizes special language communication in a special environment. It includes English for General Business Purposes (ESGP) and English for Special Business Purposes (ESBP). Business English textbooks need to carefully select and design teaching materials, study specific materials for adult learners, and the textbooks need to meet the specific needs of learners.

2.2. Feature Extraction of Spoken Speech Information

Linear Predictive Cepstrum Coefficient (LPCC).

Speech signal is the common result of both channel frequency characteristics and excitation signal source. The speaker's personality characteristic largely depends on the speaker's pronunciation channel, i.e., the channel spectrum characteristics. Therefore, it is necessary to separate the two effectively. Therefore, in speech recognition systems, LPC coefficients are seldom used directly, but another parameter is derived from LPC coefficients: LPC (Captrum Linear Prediction More. The LPCC can completely eliminate the multiplication of the speech production process, especially reflecting the frequency characteristics of the sound system. The caprum is actually a homomorphis signal. Is the processing method. The standard Cepstrum coefficients calculation process is complex. The LPCC solution process is shown in Fig. 1.

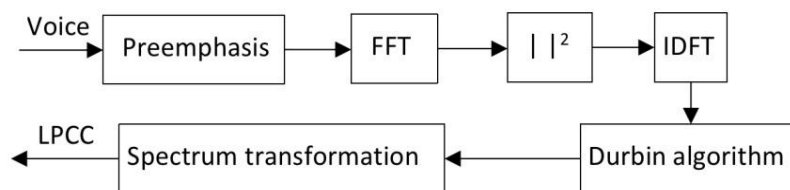


Fig. 1. LPCC solution flow

However, in practice, LPCC coefficients can be obtained by LPC. The direct recurrence relationship is as follows:

$$c_0 = \log G^2, c_m = a_m + \sum_{k=1}^{m-1} \left(\frac{k}{m}\right) c_k a_{m-k} \quad (4)$$

Where G^2 is the gain of the model, c_0 is actually a DC component, which is usually not used in the recognition and is not calculated, p is the order of LPC. LPCC can improve the stability of parameters. Its main advantages are small computation, easy implementation, good vowel description ability, and good effect in speech recognition. However, LPCC has its drawbacks, such as poor ability to describe consonants and poor ability to resist noise. At the same time, LPCC inherits the disadvantage of LPC. One of the main points is that LPC is a linear approximation of speech at all frequencies, which is inconsistent with human auditory characteristics.

Meier Frequency Cepstrum Coefficient (MFCC)

Scrams play an important role in the general identification of the language to the human ears in a variety of noise environments and in a variety of ways. Coachella is basically the equivalent of a filter bank. Angular filtration is performed on a Logarithmic frequency, with a line under 1000Hz and a Logarithmic scale above 1000Hz, which makes the human ear more sensitive to low frequency signals than to low frequency signals. For this acoustic model of the human ear, mel-frequency cepstrum coefficients (MFCCs) have been proposed. In recent years, MFCC has been widely used and many studies show that it improves the efficiency of system identification.

The calculation of MFCC parameters is based on "bark" as its frequency benchmark. Its conversion relationship with linear frequency is as follows:

$$f_{mel} = 2595 \log_{10}(1 + f / 700) \quad (5)$$

MFCC parameters are also calculated by frame. Firstly, the power spectrum $x(n)$ of the frame signal is obtained by FFT and converted to the power spectrum at Mel frequency. This requires that several band-pass filters $H_m(n), m = 0, 1, L, M - 1, n = 0, 1, L, N / 2 - 1$ be set in the frequency range of speech before calculation. M is the number of filters, usually 24. N is the number of points in a frame of the speech signal. The filter is a simple triangle in the frequency domain, and its center frequency is f_m , they are uniformly distributed on the Mel frequency axis. In linear frequency, when m is small, the adjacent f_m interval is small. With the increase of m , the adjacent f_m intervals gradually open. The parameters of band-pass filters are calculated in advance and used directly in calculating MFCC parameters.

The business English major was born under the background of China's continuous deepening of reform and development and the globalization of world economy, culture, and politics. As a derivative major of the English major, Business English has changed its name, from "Business English" when it was founded in the 1950s to the later "Finance English", "Business English", "Business English" to today's "Business English". The course of this major has also experienced a development process from the course of "Business English Correspondence" to a major that has begun to take shape (Chen Zhunmin, Wang Lifei 2009). After tortuous development, Business English was finally officially approved as an independent subject by the Ministry of Education in 2007. The specific steps of calculating MFCC parameters are as follows:

(1) Firstly, the number of points in each frame of the speech sampling sequence is determined, and $N = 256$ points are selected in this paper. After preemphasis of each frame sequence, the discrete power spectrum $S(n)$ is obtained by discrete FFT transform and the square of the mode.

(2) Calculate $S(n)$. The M parameters $P_m (m=0,1,L,M-1)$ are obtained by calculating the sum of the product of $S(n)$ and $H_m(n)$ at each discrete frequency point through the power value obtained after M $H_m(n)$, that is, calculating the sum of the product of $S(n)$ and $H_m(n)$ at each discrete frequency point.

(3) Calculate the natural logarithm of P_m and get $L_m, m=0,1,L,M-1$.

(4) For L_m , the discrete cosine transform is calculated and D_m is obtained. $m=0,1,L,M-1$.

(5) Abandoning the D_0 representing the DC component, take D_1, D_2, L, D_K as MFCC parameters. In this paper, $K=12$.

Voice information is mostly concentrated in the low-frequency part, while the high-frequency part is vulnerable to environmental noise interference. MFCC parameters emphasize the low-frequency information of the voice, thus highlighting the information that is beneficial to the recognition and shielding the interference of noise. However, the calculation and extraction of MFCC parameters are complex and time-consuming.

2.3. Recognition of Speech Information

Hidden Markov model

Hidden Markov Model (HMM), as a statistical model of speech signals, is widely used in various fields of speech processing today. Its theoretical basis was established by Baum et al. around 1970, and then applied to speech recognition by Baker of CMU and Jelinek of IBM et al. Because of the deep and simple introduction of HMM by Robiner et al. of Bell Laboratory in the mid-1980s, HMM has gradually become a recognized research hot spot for researchers engaged in speech processing all over the world.

As HMM is a stochastic probability model, it not only describes the dynamic change of speech signal characteristics, but also describes the statistical distribution of speech signal characteristics very well. It is a powerful tool for quasi-static speech signal analysis and speech recognition.

Hidden Markov process is a double stochastic process: one is used to describe the statistical characteristics of the short-term stationary period of the nonstationary signal (the transient characteristics of the signal can be observed directly); the other is used to describe how each short-term stationary period can be transformed into the next short-term stationary period, that is, the dynamic characteristics of the short-term statistical characteristics (implicit in the observation sequence). Based on these two stochastic processes, HMM can effectively solve the problems of how to identify short-term

stationary signal segments with different parameters and how to track the transformation between them.

The human speech process is also such a double random process. Because the speech signal itself is an observable sequence, and it is a parameter stream of phonemes (words and sentences) produced by the brain (not observable), according to speech needs and grammatical knowledge (state selection). At the same time, a large number of experiments show that HMM can really describe the process of speech signal production very accurately.

Wavelet neural network

The main idea of Wavelet Neural Network (WNN) is to use the wavelet function as the neuron activation function, thus combining the wavelet and BP network. Based on a similar idea, Pati introduced discrete wavelet transform into neural network in 1993 and proposed a discrete affine wavelet neural network with single hidden layer forward structure. Because the wavelet neural network inherits some advantages of wavelet analysis, such as multiresolution, compact support, and even orthogonality of the basis function, the research of wavelet network has attracted extensive attention from the beginning, and a variety of network forms and design methods have emerged.

From the structural point of view, the wavelet network is a BP network based on the analysis of wavelets, so it can usually be regarded as a generalization of Radial Basis Function (RBF) network, but it has different characteristics from the general forward network and RBF network. For example, the determination of wavelet primitives and the whole network is based on the theory of wavelets, which can avoid the blindness in structure design, such as BP network, have strong function learning and generalization ability, have good ability of feature extraction and shielding random noise, especially suitable for the classification of non-stationary and non-linear signals. In signal classification and recognition, the wavelet space can be used as the feature space of signal classification, and the rule of signal feature extraction can be realized by the function of neural network classification. Because of its unique mathematical background, wavelet network has been widely used in image compression, data classification, signal representation, nonlinear approximation, feature extraction, pattern recognition, and adaptive control.

Speech recognition based on HMM/WNN Hybrid Model

Large English companies are in dire need of teaching models and teaching materials that incorporate English and business knowledge. The teaching described here rejects the teaching methodology and textbook integration. In terms of teaching methodology, teachers who are proficient in language proficiency and subject matter are required to teach, and not just teachers who excel in a particular area. By integrating itself into the atmosphere of both business and English knowledge at the same time, students can combine these two areas of knowledge and skills and achieve learning outcomes that complement business knowledge and basic English skills. Usually, the HMM parameters of each digit are trained by EM algorithm. The training criterion is maximum likelihood

estimation, that is, the output probability of each digit is maximized by adjusting the HMM parameters. Training is carried out in the same kind of sample data, and classification decision-making needs to be carried out in different types. Therefore, when the differences between words are small, such as confusing numbers, it is difficult for HMM methods to ensure that the probability calculated by HMM parameters in a certain class of patterns is greater than that calculated by HMM parameters outside the class. The hybrid system adds a WNN classifier with the ability to distinguish between classes on the basis of the optimal state sequence, which not only guarantees the original classification characteristics of HMM, but also makes use of the nonlinear classification ability of WNN to classify small pattern differences in different intervals. The method proposed in this paper mainly focuses on distinguishing similar sequential patterns and considering all training samples in the training process, which is why the performance of HMM/WNN hybrid model is better than that of HMM.

The feature of HMM is that it can extract temporal features effectively, but it can only use the state with the greatest accumulation probability in each model, and it does not make full use of the accumulation probability of other states. At the same time, it ignores the similar features of each pattern, which affects the performance of HMM speech recognition. The wavelet neural network is applied to HMM speech recognition, and its ability of subdivision is utilized to recognize speech. The specific method is to use the cumulative probability $[\delta_r^1(1,L), \delta_r^1(N), \delta_r^2(1,L), \delta_r^2(N), \dots, \delta_r^K(1,L), \delta_r^K(N)]$ of all States in HMM as the input characteristic of the wavelet neural network classifier. Where K is the number of speech primitives to be recognized. Take the number recognition in Chinese as an example, $K=10$. The neural network model consists of the input layer, the hidden layer, and the output layer. The input layer consists of L neurons ($L=N*K=40$ in this experiment), which corresponds to the state accumulation probability of each speech element in HMM. The hidden layer is a dynamic tissue layer containing P neurons. P Dynamically changes in network training, here take $P=10$. The output layer consists of K neurons, each corresponding to a speech unit to be recognized (Take $K=10$ in this experiment).

The block diagram of HMM/WNN hybrid model speech recognition system is shown in Fig. 2. The whole system consists of two parts: HMM recognition subsystem and wavelet neural network recognition subsystem.

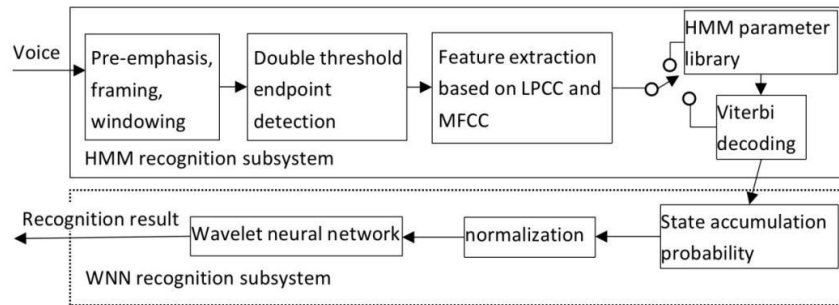


Fig. 2. HMM/WNN hybrid model speech recognition block diagram

As can be seen from the process in Figure 2, first, the voice enters the frame and window we designed, and then the dual-threshold endpoint is detected. The detection results are based on LPCC and MFCC for feature extraction, and then the HMM parameter library is calculated to perform Viterbi decoding to get the cumulative probability of transition, normalize the probability, enter the wavelet neural network, and finally get the recognition result. Among them, the received symbols are first judged by the demodulator, outputting 0 and 1 codes, and then sent to the decoder in a form called Viterbi decoding; and HMM is the simplest dynamic Bayesian network, and it is a particularly famous directed graph structure.

3. Concrete Experiments

In this article, the wearable motion sensor is used to collect students' voice information, and through the extraction of voice features, recognized vocal information is obtained in order to correct and guide students' pronunciation, and to reform the way English is taught. The flow chart of this article is shown in Figure 3.

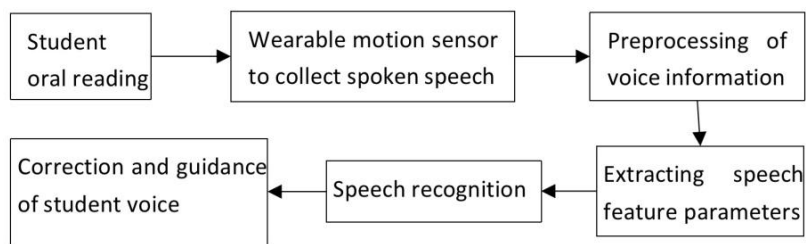


Fig. 3. Article idea flow chart

It can be seen from the process in Figure 3 that the student transmits the voice through oral reading, then collects the voice through the wearable motion sensor, first preprocesses the voice information, and corrects and guides the student to make the correct voice, and finally through voice recognition extract the voice feature parameters.

To better evaluate the performance of the speech recognition method designed in this paper, the recognition rate is used as the evaluation index. The expression is as follows:

$$\text{Recognition rate} = \frac{\text{Number of samples correctly identified}}{\text{Total number of test samples}} \times 100\% \quad (6)$$

In addition, questionnaires were used to investigate the students' sense of the business English teaching equipment based on wearable motion sensor, speech recognition designed in this paper, and the teaching effect of the business English teaching equipment based on wearable motion sensor, speech recognition designed in this paper was tested by the experimental class method.

4. Discussion of Experimental Results

The knowledge involved in business English majors is very different from that of ordinary English majors. In a broad sense, English in all business-related fields (economics, management, law, etc.) is considered business English. Strictly speaking, it is not a pure linguistics, but belongs to the category of applied linguistics. It belongs to the category of English for special purposes. First, before confirming the doctor's findings, it is imperative that the performance of the proposed speech recognition method be decided. At different signal-to-noise rates, this document sets out the HMM speech recognition method for speech recognition on the neural network, and the identification rate to reflect the performance of the speech recognition system built into this document. Uses as the criterion. Results are shown in Table 1.

Table 1. Identification rate of different speech recognition methods

Signal to noise ratio (dB) Recognition methods	HMM speech recognition	Speech recognition based on neural network	Method of this paper
30	97.5	98.1	99.2
25	97.1	97.4	99.3
20	91.3	92.7	97.1
15	85.4	86.9	91.4
10	73.7	75.8	87.6
5	67.3	68.6	83.2
0	61.7	63.9	75.7

For the visual representation, draw a line chart according to Table 1 as shown in Fig. 4.

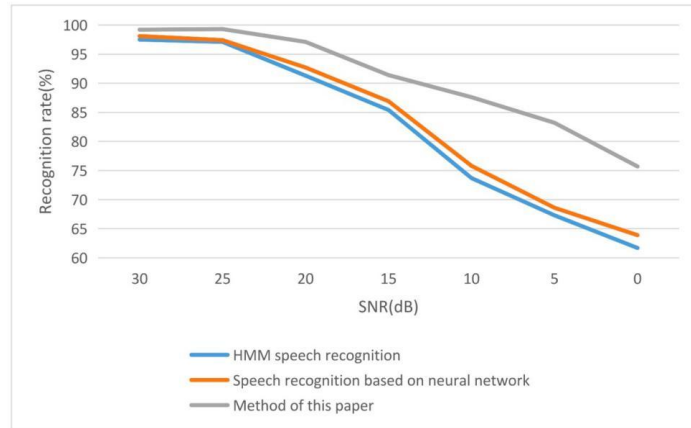


Fig. 4. Comparison of the performance of different speech recognition methods

Combining Table 1 and Figure 4, we can see that, first, by decreasing the SNR, i.e. increasing the noise, the recognition rates of the three speech recognition methods are reduced. Among them, HMM-based speech recognition method and neural network-based speech recognition method have a faster downward trend, which shows that the antinoise performance of this method is better than HMM-based speech recognition method and neural network-based speech recognition method. Secondly, under the same SNR condition, the speech recognition rate of this method is the highest, followed by the speech recognition method based on neural network, and the speech recognition method based on HMM is the worst. This shows that this method has good speech recognition performance and is feasible for students to read speech recognition, and has good recognition performance and antinoise performance.

After verifying the feasibility of this method in students' reading and speech recognition, this paper selected 100 students through a questionnaire survey to investigate their views on the business English teaching system based on wearable motion sensor, speech recognition designed in this paper. The results are shown in Fig. 5.

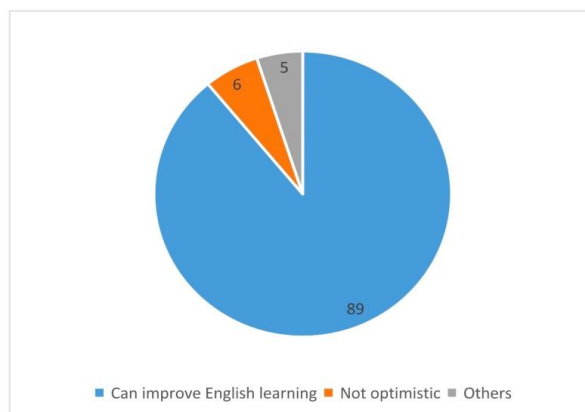


Fig. 5. Students' views on the teaching method of this article

As can be seen from Figure 5, 89 of the 100 students are optimistic about the business English teaching system based on wearable motion sensor speech recognition proposed in this paper. The method proposed in this paper can promote students to learn business English better. Only 6 students are not optimistic about the teaching system designed in this paper. This shows that the business English teaching system based on wearable motion sensor, speech recognition designed in this paper meets the needs of today's students in learning English and can arouse students' interest in learning English.

The above survey of students is only a survey of students' opinions. It does not represent the real effect of the business English teaching system based on wearable motion sensor, speech recognition designed in this paper. To verify the teaching effect of the business English teaching model designed in this paper, non-experimental classes and experimental classes are set up for testing. The students with almost the same level of English are divided into two classes. The non-experimental classes are taught in the traditional way of business English teaching, while the experimental classes are taught in the traditional way of business English teaching. The business English teaching system based on wearable motion sensor, speech recognition designed in this paper is also used to assist learning. After one semester, the two classes were tested by book test and oral expression test. The full score was 100, and the oral expression test was scored by three English teachers. The average score was used as the result. The test results are shown in Table 2 and the bar chart is shown in Fig. 6.

Table 2. Test results of experimental class

Class	Book test class average score	Oral expression grade class average score
Non-experiment class	72.5	63.3
Experimental class	83.1	84.7

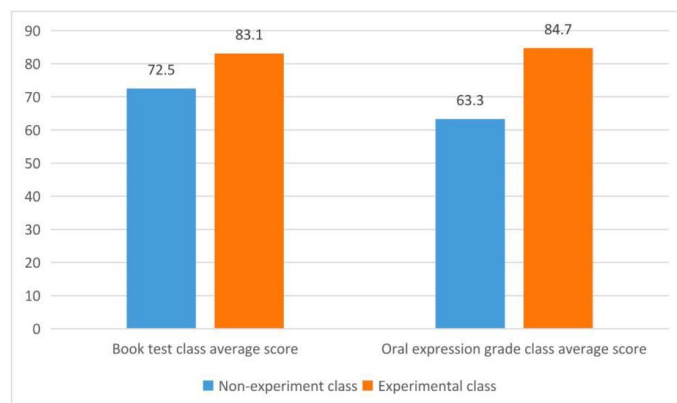


Fig. 6. Experimental class test analysis

The most important part of Business English course design is the perfection of course implementation. The quality of implementation links directly affects students' ability for theoretical knowledge and practical ability. Therefore, improve the school-enterprise

cooperation evaluation mechanism, pay attention to the industry's requirements for talent training, for teachers to continuously improve the implementation of teaching, and for students to adjust their learning habits through feedback. Combining Table 2 and Figure 6, it can be seen that the scores of the experimental class are higher than those of the nonexperimental class in both the book test and the oral expression test. In the aspect of oral expression, the gap between the two classes is larger, and the scores of the experimental class are 21.4 points higher than those of the non-experimental class. This shows that the business English teaching system based on wearable motion sensor, speech recognition designed in this paper can promote students to learn business English better and has a better effect in oral expression, which can make up for the shortcomings of oral expression in our country.

5. Conclusions

As an applied multiple discipline, the Business English major includes linguistics, psychology, sociology, economics, management, law, education, computer science and many other disciplines. Therefore, general comprehensive English textbooks may not fully meet the needs of business English majors for professional knowledge acquisition. With the changes of times, the monotonous college English textbooks have lost their vitality, and it is an inevitable trend to develop specialized English textbooks suitable for all walks of life. Considering the importance of comprehensive English courses in this major, it is particularly important to study whether the selection of materials and themes of the course materials are suitable for business English majors. Nowadays, with the rapid development of science and technology, the traditional business English teaching can no longer meet the needs of today's students, and the reform of business English teaching has become increasingly urgent. Wearable motion sensor is a kind of technology which can improve life and learning effectively with the development of technology. This is probably a good request. Based on the advantages of laundry motion sensors, this paper applies them to teaching English business so that students can better learn business English. This article specifically collects the student's reading speech through a vibration sensor and then extracts them. LPC and MFCC speech feature parameters through speech preprocessing with the relative benefits of HMM and WNN. The collection, based on the HMM / WNN, is a hybrid speech regulation model that allows students to feel the success of speech reading. By analyzing the recognition rate of speech recognition methods, it shows that the speech recognition wording in this article is good performance and resistance performance and can be used to recognize English business reading speech. Through questions, this paper illustrates students' views on vibration sensors in applied English business learning and through experimental classroom experiments, also shows the effect of vibration sensors on applied business English learning. It can promote business learning.

References

1. Barshan B, Yurtman A. Classifying daily and sports activities invariantly to the positioning of wearable motion sensor units. *IEEE Internet of Things Journal*, 7(6):4801-4815. (2020)
2. Huang H, Li X, Liu S, et al. TriboMotion: a self-powered triboelectric motion sensor in wearable internet of things for human activity recognition and energy harvesting. *IEEE Internet of Things Journal*, 5(6):4441-4453. (2018)
3. Fallahzadeh R, Ghasemzadeh H. Trading off power consumption and prediction performance in wearable motion sensors: an optimal and real-time approach. *ACM Transactions on Design Automation of Electronic Systems (TODAES)*, 23(5):1-23. (2018)
4. He Z, Liu T, Yi J. A wearable sensing and training system: towards gait rehabilitation for elderly patients with knee osteoarthritis. *IEEE Sensors Journal*, 19(14):5936-5945. (2019)
5. Davide M, Leonardo B, Michele C, et al. Profiling the propagation of error from PPG to HRV features in a wearable physiological-monitoring device. *Healthcare Technology Letters*, 5(2):59-64. (2018)
6. Bichitra, Nanda, Sahoo, et al. Superhydrophobic, transparent, and stretchable 3D hierarchical wrinkled film- based sensors for wearable applications. *Advanced Materials Technologies*, 4(10):1900230-1900230. (2019)
7. Ostaszewski M, Pauk J. Estimation of ground reaction forces and joint moments on the basis of plantar pressure insoles and wearable sensors for joint angle measurement. *Technology & Health Care*, 26(2):605-612. (2018)
8. Deqing Zhuoma. The presentation and improvement of business English teaching mode in Colleges and Universities under the background of innovation and entrepreneurship%. *Science and Technology Entrepreneurship Monthly*, 032(006):101-103. (2019)
9. Si J. The practicality of ELF-informed teaching: attitudes and perceptions of Chinese business English teachers. *Journal of English as a Lingua Franca*, 8(2):269-296. (2019)
10. Li J. Design and implementation of project-based business English teaching based on culture of core professional competencies. *Journal of Zhejiang Communications Vocational and Technical College*, 020(001):50-53, 59. (2019)
11. Chen X. The application of flipped classroom in Business English Teaching -- Taking Business English reading for example. *Journal of Heilongjiang Institute of Education*, 038(003):130-132. (2019)
12. Qi C. Research on business English Teaching in minority areas under the training target of applied talents. *Guizhou Ethnic Studies*, 39(6):239-242. (2018)
13. Cui B Y. Research on the characteristics, orientation, development and teaching of business English. *Overseas English*, 000(1):244-245. (2019)
14. Zhao J. Research on business English listening and speaking learning reform from the perspective of ESP teaching theory% ESP, 029(001), 154-156. (2019)
15. Hu Y. Study on the effectiveness of task driven block teaching in Business English. *Journal of Xinyu University*, 023(004):123-127. (2018)
16. Liu R. Research on the business English translation teaching in the sight of frame semantics. *Journal of Chongqing Vocational College of Electronic Engineering*, 27 (3):110-112, 116. (2018)
17. Sun H. Project based teaching of business English translation based on micro class. *Heilongjiang Science*, 9(4):118-119. (2018)
18. Zhan W, Shen Z. Research on business English translation teaching in Ethnic Colleges under the circles of economic integration. *Guizhou Ethnic Studies*, 039(008):236-239.
19. Chen W, Department, F. Research on innovation and strategy of teaching paradigm of business English Major under the vision of information. *Journal of Changchun University (Natural Science Edition)*, 28(4):82-86. (2018)

20. Feng J. Effectiveness of SPOC based blended teaching for Higher Vocational Business English. *Vocational Technology*, 017(011):52-55. (2018)

Dan Lu was born in Yiyang, Hunan, P.R. China, in 1977. She received the Master's degree from Hunan University, P.R. China. Now, she works in School of Foreign Languages, Shunde Polytechnic. Her research interests include Business English and English teaching. E-mail: rachellu2009@126.com

Fen Guo was born in Yiyang, Hunan, P.R. China, in 1979. She received the Ph.D. degree in Computer Application Technology from the School of Software Engineering, South China University of Technology (SCUT), Guangzhou, China, in 2015. She has been working as a teacher in SCUT since 2005. Her current research interests include data mining, data security and cloud computing. E-mail: csguofen@scut.edu.cn

Received: March 20, 2021; Accepted: March 08, 2022.

Construction of Innovative Thinking Training System for Computer Majors under the Background of New Engineering Subject

Guoxun Zheng¹, Xiaoxian Zhang^{2,*}, Ruojin Wang³, Liang Zhao⁴, Chengming Wang⁵,
and Chunlai Wang⁶

¹ Science and Technology Research Division, Changchun Institute of Technology,
Changchun 130012, Jilin, China
zhengguoxun@ccit.edu.cn

² School of Computer Technology and Engineering, Changchun Institute of Technology,
Changchun 130012, Jilin, China
zhangxiaoxian@ccit.edu.cn

³ Enrollment and Employment Division, Jilin Communications Polytechnic,
Changchun 130012, Jilin, China
260908900@qq.com

⁴ Secondary Education Enrollment and Examination Department, Jilin Education
Examinations Authority, Changchun 130012, Jilin, China
105009162@qq.com

⁵ Higher Education Enrollment and Examination Department, Jilin Education Examinations
Authority, Changchun 130012, Jilin, China
729998967@qq.com

⁶ Enrollment and Employment Division, Jilin Jianzhu University,
Changchun 130012, Jilin, China
63625851@qq.com

Abstract. Computer major has trained a large number of computer related talents for the society. The graduates of this major are an important force of social development, and also make a significant contribution to the development of the national economy. Paying attention to the new demand of social development for high-quality computer talents, targeted training is the key to the development of scientific and technological innovation. Firstly, the paper points out the main problems affecting the cultivation of talents in this major. Then, based on the basic idea of new engineering subject, it discusses how to renew the basic educational concept of computer major, strengthen the cooperation between industry and university, reform according to the requirements of new engineering subject, and realize incremental optimization, stock adjustment and cross-integration from various aspects.

Keywords: New engineering subject, Computer Major, Personnel Training, Educational Idea, Industry-University Cooperation, Teaching Reform.

* Corresponding author

1. Introduction

According to statistics, up to 2016, the number of computer majors in China, including 6 basic majors and 7 special majors, has reached 2956. In addition, in recent years, the salary of computer graduates, whether half a year after graduation or three years after graduation, ranks first in all majors. The rapid development of Internet and "Internet +" (mobile payment and so on has been the world's leading), the rapid development of mass innovation and entrepreneurship has made it one of the new driving forces of social development, which has created great demand for computer professionals. It can be predicted that this momentum will continue for quite a long time in the future. Of course, this phenomenon not only illustrates the importance of computer professionals, but also shows that the society has a greater, more vigorous and long-term sustained demand for these professionals, and there is a big gap between the talents trained in this profession and these needs, especially the new ones in the future. Where is the gap? What is the key to the problem? How to carry out the reform? It is worth our in-depth thinking and exploration.

2. Related Work

According to the research progress at home and abroad, different scholars also have a certain degree of cooperative research on the innovative thinking training system of computer majors. Although hardware and software technology is running faster, there is a doubt that all older gamers who use high-speed computers can work faster with today's speed or functional computers. In this context, Shafiulla S proposed a model to simulate the analysis of the interaction between the cognitive responses of the brain to gaming at different ages. The model in turn adheres to the actual reaction speed of the brain, allowing people to perceive work at different speeds and identify different ways of working [1]. Using the principle of blended learning, Wang R proposed a blended learning model combining online assessment and computer programming courses [2]. The Rizk N study aimed to understand the impact of metacognitive strategies on the development of creative thinking in primary school students. They were defined as students with IQ scores of 120 and above according to the Wakslar Children's Intelligence measure and were selected by their teachers [3]. The Dessie W M study increases the innovation point of companies by examining whether training encourages small companies to be more creative and innovative. It does this by investigating the extent of creative and innovative activities of small enterprises (SEs) trained with the support of the Ethiopian Technical and Vocational Education and Training System [4]. The aim of Garad A is to propose an organizational learning model that can help organizations transform into learning-driven organizations; the model considers the entire ecosystem and its subsystems, and takes into account the importance of technology, digitization, and dataism [5]. The purpose of Biletska EM research is to develop an innovative system of educational means for the formation of preventive thinking of students in higher medical education institutions, involving professional-oriented doctoral undergraduate training [6]. Cao K takes the application of "Internet +" in sports training and monitoring as the research object, and uses the literature method,

model parameter construction method and application demonstration method to carry out innovative research on the service system. In this service system, the new thinking of "Internet +" uses Internet technology and platform as the medium [7]. However, these scholars did not discuss the construction of the innovative thinking training system for computer majors based on the background of new engineering, but only discussed its significance unilaterally.

3. Major Problems Affecting the Quality of Education for Computer Majors

Computers are relatively young. In 1956, Harbin University of Technology and Tsinghua University took the lead in setting up computer specialty in accordance with the needs of the construction of China's "Vision Plan for the Development of Science and Technology for the Twelve Years 1956-1967". By 1960, 14 colleges and universities in China had set up computer specialty; from 1978 to 1993, 123 universities had increased to 137; and by 2012, 931 universities had set up computer specialty. In this period of time, because of the continuous innovation and development of science and technology, some new computer majors are gradually set up in universities. In 1998, the figure of network engineering major appeared in the professional catalogue issued by the Ministry of education. By 2001, there were 12 major of network engineering in China. With the continuous development of emerging industries, the country's demand for talents is also increasing. Therefore, software engineering, Internet of things engineering and information security have been established. The major includes six basic majors and a number of ad hoc majors. Computer Science and Technology is one of the majors. By 2016, there were 2956 computer specialty points, and 985, 598, 503 and 427 computer science and technology specialty, software engineering specialty, Internet of Things engineering specialty and network engineering specialty respectively ranked the top four. These developments have obvious characteristics of "extension development", which can be regarded as a stage of the development of this major. After the development of denotation, we are now turning to the stage of "connotative development" which focuses on improving quality. The construction of new science and technology is an opportunity to promote the strategic transfer of development. Therefore, we must solve the problems that affect the connotation professional development. In the author's opinion, there are three main problems.

3.1. Basic Ideas of Education Need to be Renewed

First of all, we should change from traditional teaching-centered to student development-centered. In the stage of elite education, specialties are subject-oriented. Specialized education emphasizes the systematic and comprehensive coverage of the corresponding backbone disciplines. Graduates are assigned to certain posts according to the needs of the country. They rely on solid foundation to gradually adapt to work. This elite education mode cannot adapt well to the reality of talent training in the stage of popular education. In the stage of popular education, education must be designed and

implemented according to how the educated can better meet the needs of society. That is to say, in the stage of elite education, specialty aims at the problem space of the whole discipline and specialty when planning and implementing personnel training, while in the stage of popular education, it should implement classified training, and divide the problem subspace adapted to the students' specialty points in the problem space of the whole discipline and specialty, so as to deal more effectively with the contradiction between the limitation of students' time in school and the infinity of knowledge. As well as the contradiction between the fundamentality of education and the futurity of exploration and creation, we should make better and more effective use of limited time and face the future so as to enable students to better develop their specialties, provide better services for the society and realize the goal of "promoting their strengths for excellence". Therefore, we need to focus more on social needs and student development.

Secondly, closely related to the first point, we should shift from curriculum-oriented education (CBE) to output-oriented education (OBE). The key difference between output-oriented education and curriculum-oriented education lies in whether output (ability) or input (knowledge/curriculum) is pursued. Course-oriented education pays attention to what kind of courses should be learned, which are the main courses in these courses (so some professional norms and standards specify what courses must be taken by the specialty). Course teaching pays attention to how well the teachers are doing, and course examination pays attention to the coverage of the knowledge points of the course by the examination papers. In this mode, there will be the key knowledge coverage of examinations (determining the key points before examinations, etc.), a hotbed for examination-oriented education, and opportunities for the existence of specialties that are not qualified (as long as the listed "name" courses are available, as for the understanding of the specialty, whether there are basic professional conditions is not so important). Output-oriented education focuses on students' ability to solve problems. According to the international equivalent standard, two-year college graduates only need to solve narrow engineering problems, three-year college graduates will solve broad engineering problems, and undergraduates will solve complex engineering problems. Therefore, the degree to which the undergraduates majoring in computer science should acquire knowledge is sufficient to support the solution of complex computing problems. They should learn to analyze and design, and take into account the corresponding social, ethical, moral, legal and other social factors. In this way, it is difficult to achieve the goal only by experience. What we need is more scientific and more elaborate design and teaching implementation: design training objectives, design suitable graduation requirements, decompose graduation requirements, design appropriate curriculum system and carry out appropriate education according to the needs of graduation requirements.

Thirdly, establish a perfect continuous improvement (CQI) system. Perfect continuous improvement system is different from the traditional quality monitoring system currently in operation in most schools. The traditional quality monitoring system serves curriculum-oriented education [8]. Its focus is to monitor teachers and monitor how well they teach. The perfect continuous improvement system is to fully reflect the needs of student-centered and output-oriented education, its quality monitoring mechanism is to promote students to effectively achieve the curriculum objectives and graduation requirements embodied in the ability requirements; the graduate tracking feedback mechanism and social evaluation mechanism is to evaluate the rationality of

training objectives and promote their realization, taking into account the evaluation and reform of graduation requirements. Enter. Especially in the perfect continuous improvement system, the requirements are not isolated evaluation and simple feedback, but feedback based on system evaluation and improvement based on evaluation.

3.2. Standard of Talent Training Needs to be Accelerated

As an expert organization hired and led by the Ministry of Education, the Professional Teaching Steering Committee is entrusted by the Ministry of Education to carry out research, consultation, guidance, evaluation and service of undergraduate teaching in Colleges and universities. In recent three years, the Teaching Steering Committee of computer specialty has developed and issued professional norms, standards and relevant guidance, and has carried out a large number of extensive publicity and promotion. They have played an important role in promoting the construction, reform and development of computer specialty. For example, "Strategic Research Report and Professional Specification on the Development of Computer Science and Technology Specialty in Colleges and Universities", "Public Core Knowledge System and Courses of Computer Science and Technology Specialty in Colleges and Universities", "Implementation Plan of Core Course Teaching for Computer Science and Technology Specialty in Colleges and Universities", "Composition and Training of Professional Ability of Computer Science and Technology Specialty in Colleges and Universities", "General Higher Instructions on Computer Science and Technology for Undergraduates of the University, National computer professional teaching quality standards (to be officially released). In addition, other documents are instructive. The document has not been officially released by the Ministry of education. Therefore, it can be considered that the computer professional teaching quality is still lack of national standards. In addition, although in the guidance documents listed above, we have begun to guide all majors to get rid of the constraints of the curriculum, giving the recommended areas of knowledge, basic requirements for training, and the ability composition of computer professionals, etc., we have not yet clearly and thoroughly guided education from curriculum-oriented education to output-oriented education. In addition, people come from the traditional curriculum-oriented education, which makes the training of talents based on appropriate standards just start, and there is still a long way to go from full implementation. It is for these reasons that the orientation of undergraduate teaching is not clear or even deviated. Many specialty-oriented education of engineering degree has not yet realized that it should focus on cultivating students'ability to solve complex engineering problems. There are also some specialty points, such as emphasizing theoretical teaching (not required theoretical teaching), neglecting practical teaching, and "scientific engineering education".

3.3. Quality consciousness of talent training needs to be strengthened

Quality is the lifeline of professional development [9]. However, in the case of denotative development complex, people's quality consciousness is still relatively

indifferent. The curriculum-oriented education concept not only makes it difficult for professional education to meet the requirements, but also opens the door for examination-taking and simple 60-point "meeting the standard".

In addition, as mentioned earlier, the traditional quality assurance system has been unable to meet the current needs of personnel training, and the new continuous improvement system needs to be established and improved [10].

3.4. Cultivation of Innovative Thinking Ability

(1) Adjust the teaching content.

To cultivate the innovative thinking of college students majoring in computer science, it is necessary to add a link to cultivate students'innovation ability in the daily teaching plan, so that students can understand and master the process characteristics and forms of the formation of innovation ability in the usual learning process, and cultivate innovation purposefully Awareness, and constantly stimulate students'enthusiasm for innovation.

(2) Reform teaching methods.

The premise of innovative thinking is to have a wealth of professional knowledge and extensive relevant professional knowledge. This knowledge depends on students' independent learning on the one hand, and teachers on the other hand. Computer professional teaching has the generality of general teaching, but it also has special professional characteristics based on its particularity [11].

(3) Create an atmosphere of innovation.

Cultivating students'innovative thinking, stimulating students'interest in innovation, and creating a good campus innovation atmosphere is a basic task [12-13]. Organize and carry out popular innovation activities through all-round, multi-level and wide channels to attract students'attention and stimulate their enthusiasm for participating in innovation activities, and strive to create a strong innovation atmosphere on campus, infect and nurture a large number of students, and promote They participate in scientific research and innovation activities and gradually cultivate students'innovative thinking and innovative spirit [14-15].

(4) Cultivate the spirit of innovation.

During the teaching period, teachers should encourage students to ask questions, do not easily deny students' ideas, do not rashly judge students' innovative thinking viewpoints, and encourage and praise innovative and exploratory student thinking models [16].

Reference evaluation is not only limited. In textbooks, standard answers are not set, and students are encouraged to think about the answers to questions in multiple ways and perspectives, and not blindly believe in authority [17].

(5) Strengthen the teaching staff.

Teachers are leaders. To cultivate computer professionals with innovative thinking, teachers must first be innovative [18]. If a computer professional teacher has no development experience and research ability himself, and only teaches textbooks, then there is no way to cultivate students' innovative thinking and ability. Teachers who have scientific research and innovation capabilities and rich practical experience can use examples and vivid examples in the course of teaching; they can analyze and think sharply in practical links such as graduation design and curriculum design [19].

4. Basic Thought and Path of New Science Construction

4.1. Basic Thought and Path of New Science Construction

It is an inevitable requirement for the development of China's higher engineering education to adapt to the construction of engineering education at a new historical starting point [20]. Wu Aihua and others pointed out that: China's economic development is entering a critical period of structural adjustment, transformation and upgrading, and the momentum of new and old growth is changing [21]. A new round of technological and Industrial Revolution centered on the Internet awaits development. Innovation has become a new arena of International competition, which not only provides strategic opportunities for later countries to catch up and surpass, but also provides strategic opportunities for them. Further intensify the international talent competition. Engineering education and industrial development are closely linked and mutually supportive. The development of new industries depends on engineering education to provide talent support. Especially to meet the challenges of international competition in new technologies and industries in the future, it is necessary to actively lay out the training of Engineering Science and technology talents, accelerate the development and construction of new engineering specialties, transform and upgrade traditional engineering specialties, and enhance the ability of engineering education to support the development of service industries. It can be said that the active layout of engineering education and the deepening of reform in place will have a positive role in promoting economic transformation and upgrading; on the contrary, the lagging reform of engineering education will delay the process of industrial upgrading.

The construction of new engineering subjects should "set up new ideas of innovative, comprehensive and full-cycle engineering education", construct a new structure of engineering specialty through incremental optimization, stock adjustment and cross-integration, clarify the ability system of Engineering talents, construct

curriculum system according to engineering logic, cultivate students'innovative spirit, entrepreneurial According to the thinking logic of engineering discipline, the course teaching system is constructed, the students' innovative consciousness and entrepreneurial ability are cultivated, and the quality standard system of engineering talents training is established and improved. Form classification culture and improve quality. From the discipline orientation to the industrial demand orientation, from the professional division to the cross-border integration, from the service adaptation to the support orientation, we explore a new engineering development paradigm.

4.2. The Construction of Innovative Thinking Training Mode for Computer Majors

Through the analysis of the status quo of three representative vocational schools, compared with other vocational schools that have been more successful in the construction of computer majors, combined with the actual situation, based on the market-oriented, for the training of computer professionals, a long-term class "1 The training mode combining +1+1" and short classes is shown in Figure 1.

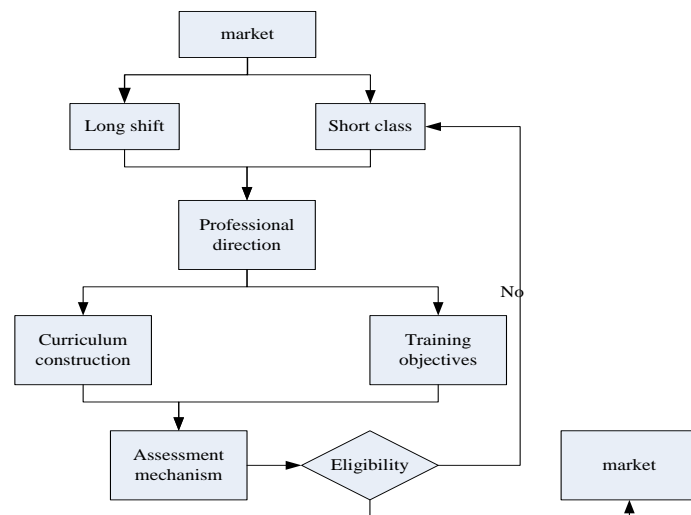


Fig 1. Cultivation model of innovative thinking for computer majors

The long-term "1+1+1" training mode for computer professionals in secondary vocational schools is only for three years, that is, the school system is three years, the first "1" refers to the first year of entering the school (the first year of high school)) In this year, we mainly offer moral education courses, cultural courses and basic computer science courses. At the end of the first academic year, students are divided into classes. Students choose the corresponding professional application direction according to their own interests and hobbies, and regroup the classes. The second "1" refers to the second year of entering the school (the second year of high school). During this year, students

mainly study professional application courses, job courses, etc. The last "1" refers to the third academic year (senior year), during which students are organized to work as internships in enterprises.

"Short class" refers to a short learning time, usually 1 to 3 months, mainly to meet the needs of in-service and non-employed personnel with learning needs, mainly targeted training, such as the use of typists, computers and peripherals Training in the use of office software.

Finally, if there are students who fail to pass the examination upon graduation, as long as the students are willing, they can participate in the school's "short class" training and obtain the corresponding skills and certificates, so as to achieve employment and entrepreneurship.

5. Promoting the Reform of Computer Specialty with the Opportunity of New Engineering Construction

Computer science has a close relationship with strategic emerging industries, especially the reform and innovation of related majors and disciplines. With the basic status of undergraduate education, the construction and development of computer majors has become very important to the development of a school, and it is no longer a problem of the profession itself. To meet the new needs, we must vigorously promote the construction and reform of computer specialty in an all-round way.

5.1. Cultivation Weight of Innovative Thinking of Computer Majors

Based on the "Questionnaire for the Evaluation of Innovative Thinking Cultivation of Undergraduates Majoring in Computer Science", the research topics are further explained to the participants in the questionnaire survey, and their questions are answered. A total of 120 questionnaires were issued to management personnel of undergraduates majoring in computer science in Shanxi universities, teachers engaged in cultivating creative thinking of undergraduates majoring in computer science, various university managers, and school-enterprise cooperative enterprises, and 100 copies were eventually recovered.

The Delphi method was used to summarize the tables filled out by 100 experts according to factors, and the average value of each grading factor calculated by the scoring results divided by 120 was used as their weight. The weight value of the cultivation of innovative thinking in computer science is shown in Table 1, and Figure 2:

Table 1. Weights for training innovative thinking in computer science

	Average value	Standard deviation	Weights
Culture condition	38	3.1	0.38
Subject relevance	43	3.6	0.44
External environment	18	4.9	0.20

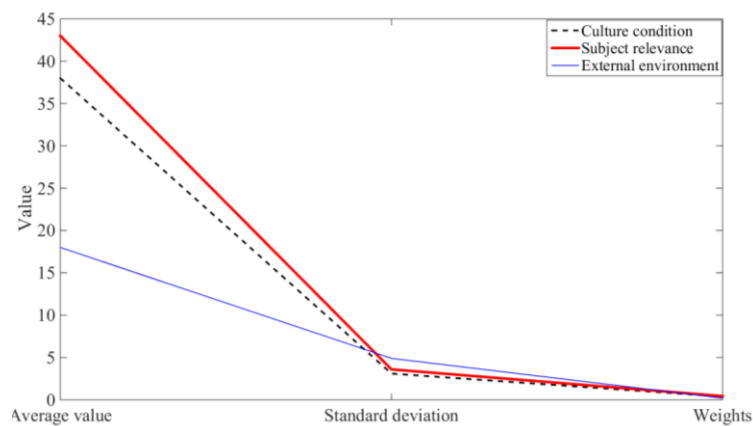


Fig. 2. The results of the training weight of innovative thinking in computer science

It can be seen from the results of the Delphi method that each evaluation index has a certain level of influence on the cultivation of innovative thinking of undergraduates majoring in computer science (that is, the order of importance). Among them, the cultivation conditions are similar to the two secondary indicators of cultivation subjects and their relevance, with weights of 0.38 and 0.44 respectively, and the external environment is 0.20. This sequence reflects that the training conditions and the relevance of the training subject play a key role in the process of developing innovative thinking in universities.

5.2. Current Situation of the Cultivation of Innovative Thinking in Computer Science

The purpose of training computer majors in ordinary colleges and universities in China is mainly to cultivate general-purpose graduates for all industries in the society, focusing on the cultivation of theoretical knowledge and learning methods in the teaching process, while computer industry-related companies need a variety of practical talents , Not only need to have relevant theoretical knowledge, but also need to have innovative and innovative thinking methods, good communication skills, teamwork and self-

discipline ability, etc. The current training methods lack innovative thinking methods training and professional quality education. Professor Tan Haoqiang, president of the National Institute of Computer Basic Education Research in Institutions of Higher Learning, pointed out that the current teaching of computer majors in colleges and universities is too focused on theory and has broken away from the actual needs of the industry. Students lack practical experience, which ultimately results in students unable to meet the actual requirements of the company after graduation.

This paper investigates the teaching situation of computer majors in three undergraduate universities. This time we surveyed the degree of satisfaction of computer students with the current teaching situation. As shown in Table 2, Figure 3:

Table 2. Student satisfaction with the current situation of innovative thinking training

Satisfaction level	Very satisfied	Satisfaction	Basically satisfied	Not satisfied	Very dissatisfied
Satisfaction	40%	23%	10%	20%	7%

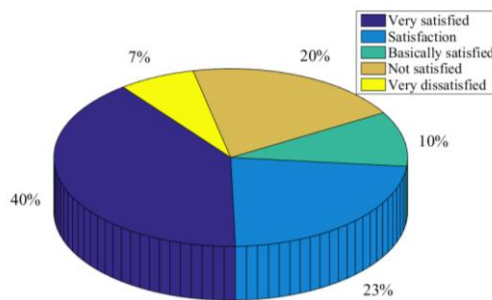


Fig 3. Student satisfaction with the current innovative thinking training situation

According to the data analysis in the figure, it can be seen that 40% of the students are very satisfied, 23% are satisfied, 10% are basically satisfied, 20% are not satisfied, and 7% are very dissatisfied. The above figure shows that the current training model is not conducive to talent training and not conducive to student development.

5.3. Incremental optimization

As one of the core of information technology, computer technology is developing rapidly. Internet of Things, big data, artificial intelligence, mobile computing, cloud computing and so on are changing the production and life style of human beings. The society urgently needs a large number of relevant talents, especially application and

development talents. The development of computer specialty is the response to this new demand.

Some special majors added to computer majors have become new growth points. Originally, there were only two special majors: Intelligent Science and Technology (080907T) and Spatial Information and Digital Technology (080908T). Since 2016, five new specialties have been added, including data science and big data, cyberspace security, new media technology and film production. Basic specialties such as Internet of Things (080905), Software Engineering (080902), Digital Media Technology (080906) will be another major growth point. Whether the addition of basic majors or special majors should meet the forthcoming National Standard for Teaching Quality of Computer Majors, especially the following requirements.

Firstly, we should clarify the social needs of our major. Generally speaking, the construction of new disciplines should aim at training graduates who meet the specific needs of the society, ask for the specialty of industrial demand construction, ask for the content of technological development and reform, and intensify the cooperation between industry and university. As mentioned earlier, the graduates of any major have their own problem subspace, which reflects the main service orientation of the students trained by the professional point and their own advantages. Therefore, the problem space subspace is neither the whole problem space nor the problem subspace of the same major in other schools. This requires the applicants to really understand and aim at the specific social needs of the students they have trained, rather than the general social needs, let alone the "needs" that can be downloaded online. And make it clear that I really want to, and be able to cultivate qualified personnel for this subspace.

Secondly, a reasonable talent training program. First of all, we should have a training goal that reflects the students' specialty and subject advantages and is in line with the "problem subspace". We should make clear the students' professional expectations about five years after graduation, instead of the general goal of "senior talents" who can do everything in the specialty such as "scientific research, engineering development, application maintenance, teaching and management" when they graduate. Secondly, there should be professional graduation requirements to support the realization of training objectives. The graduation requirement of specialty should not be lower than the graduation requirement of National Standard for Teaching Quality of Computer Specialty, and it has its own characteristics. With clear and accurate objectives, the graduation requirements of the major will reflect the characteristics of the talents trained by the major. Thirdly, the curriculum system can effectively support the professional graduation requirements. It is necessary to decompose the professional graduation requirements into indicators, and then assign the tasks of achieving these indicators to the corresponding courses (theoretical courses and practical courses). It is worth noting that the newly-built specialty should have its own characteristics, not a "direction" of computer science and technology specialty, nor simply adding or replacing several new courses on the basis of the original specialty to form a "curriculum system" of the new specialty.

Of course, as a professional education, we must have the basic conditions of classroom, laboratory, experimental equipment and so on.

5.4. Stock adjustment

As mentioned earlier, in 2016, there are nearly 3,000 specialty points in computer specialty, especially about 95% of them have not had more than 25 years' educational history. With the rapid development of higher education in recent years, the following aspects need to be adjusted.

(1) Renewing the Basic Idea of Undergraduate Talents Training

Renewal of the basic undergraduate talent training concept, in order to improve the efficiency and level of personnel training to provide security. First of all, we should return to the original idea of "talent training", firmly establish the basic status of undergraduate education, clarify the responsibilities of teachers, increase the energy input of undergraduate education, consolidate the foundation and avoid "shaking the ground". Secondly, it is to promote the implementation of the three advanced educational concepts. Establishing undergraduate education is not the input of basic knowledge, nor the basic curriculum requirements, but to pursue the output of students' basic ability to solve complex engineering problems in the future.

(2) Constructing a complete continuous improvement system

According to the theory of total quality management, we should establish a system that can be continuously improved and finally achieve perfection. To ensure its effective operation, we should establish three mechanisms: real-time quality monitoring, graduates tracking survey feedback and social overall evaluation. Pay attention to the improvement of training objectives, adjust the graduation requirements according to the background of the times, improve the curriculum system, and constantly improve the rationality and "achievement degree" of these three aspects. Three points need to be paid attention to.

First, evaluation is the foundation. Continuous improvement requires evaluation as the basis, the accuracy of evaluation, the pursuit of data reasonableness and analysis in place. To clarify the object (objective), criteria (basis), executors, appropriate and effective evaluation methods and appropriate evaluation cycle, we should base on effective collection and in-depth analysis of the original data that truly reflects the evaluation object.

Second, mechanism is guarantee. Mechanisms need to be used to ensure the effectiveness of evaluations and the sustainability of improvements. Only by constantly improving the mechanism can we effectively and continuously promote the reform. In view of quality monitoring, graduates' follow-up feedback and social evaluation, a set of standardized processing procedures should be established, and the relevant personnel involved in these processes and their respective roles should be clearly defined.

Finally, improvement is the goal. Applying the evaluation results to the improvement and sticking to the improvement based on evaluation, the improvement is well-founded and effective. "Improvement" based on evaluation emphasizes avoiding "change" based on feeling, experience and personal opinions; without in-depth analysis

of evaluation information and decision-making based on analysis, "change" is likely to be "blindly change"; to pursue "improvement", while "change" is not equal to "improvement".

(3) Define the basic orientation of training the ability of focusing on solving complex engineering problems

It is necessary to determine and implement the basic orientation of undergraduate education for engineering related majors. In order to cultivate students' ability to solve engineering problems with innovative thinking, the talent training plan is designed and implemented, and to evaluate learning output according to this requirement. Gradually rectify the problems of unclear orientation and degraded requirements of some specialty points.

It is especially emphasized here that the cultivation of students' ability to solve complex engineering problems is not simply to let students participate in one or even several complex engineering projects, but to decompose the cultivation of this ability into all aspects of teaching, which can be referred to in reference.

(4) Strengthen and implement systematic design of training program

The design and implementation of the training program must be carried out within the framework of the National Standard for Teaching Quality of Computer Specialty.

As mentioned above, first of all, to determine an appropriate training goal, then to design the corresponding graduation requirements according to the needs of supporting the realization of the goal, to divide the graduation requirements into a series of indicators according to the standards (measurable) which are easy to implement and evaluate the achievement degree, to construct the curriculum system according to the needs of supporting indicators, and to teach the actual theoretical and practical courses. Specific implementation of the corresponding indicators, through the evaluation of the results of teaching activities to prove that the indicators of graduation requirements have indeed been achieved, and through these evaluations found that the need for improvement, forward feedback, to guide improvement.

(5) Steadily moving towards scientific teaching

We should completely get rid of curriculum-oriented education and really move towards output-oriented education. Knowledge should be used as a carrier to teach students the ideas, methods and professional skills of solving complex engineering problems of computer science. These ideas and methods are the contents of computational methodology, including 12 core concepts in the sense of methodology, typical mathematical methods and systematic methods, abstraction of problems, theory and design solving process.

It must be clearly stipulated in the syllabus. Define the curriculum objectives, and the curriculum objectives are related to the graduation requirements (indicators)

supported by them (so, it is no longer just the requirements of knowledge points). The curriculum content should not only include knowledge, but also include the ideas and methods of problem solving. Therefore, the traditional syllabus has been unable to meet the new requirements, especially the "catalogue syllabus of textbooks", far from the new requirements.

(6) Strengthen the Cooperation between Industry and University, enhance the Consciousness and Ability of Innovation and Entrepreneurship

We should further strengthen the awareness of education in training talents for industrial development. In addition to "asking industry needs to build specialty and constructing new structure of Engineering specialty", we should actively promote industry-university cooperation to educate people, keep up with the development of technology, train students to pay attention to social needs, actively guide exploration, and constantly enhance students' innovative, entrepreneurial consciousness and ability. The purpose of the computer contest in Chinese universities is to guide students to select topics for social needs through cooperation between industry and university, to stimulate their intelligence, to exercise their comprehensive use of what they have learned, and to create a service for society.

5.5. Impact of Innovative Thinking Training on Computer Students

This paper conducts surveys and statistics on 200 computer students, and compares their results before and after innovative thinking training. The results are shown in Figure 4:

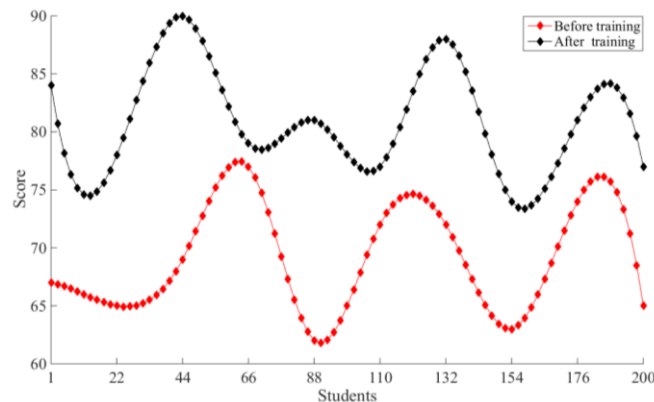


Fig 4. The impact of innovative thinking training

According to the trend and changes of the data in the figure, the academic performance after the innovative thinking training has been significantly improved compared with the performance before the innovative thinking. This also shows that the

computer majors of the innovative thinking training team have a promoting effect and it is their performance improvement. , Stepping stones to successful employment.

6. Conclusion

The new subject reform of computer specialty is not accomplished overnight. It should be carried out step by step. It needs to focus on three tasks: teaching and teaching, practice and innovation, entrepreneurship, localization and internationalization, strengthen discipline construction and change teaching ideas. Engineering technology makes the world more advanced, and science and technology can improve the living standard of the people, and it is the need of the times to train senior computer talents to meet the requirements of new subjects. It is also the mission of our computer professional teachers.

Acknowledgment. This work was supported by the 2019 Vocational Education and Adult Education Teaching Reform Research Project of Jilin Provincial Department of Education under Grant nos. 2019ZCY384, 2019ZCZ067, 2019ZCY413 and 2019ZCY414.

References

1. Shafiulla S, Nawaz G. Impact of Age Level Intelligence and Components of the Game on Simulated Response Measure in Human Computer Interaction. *International journal of computational intelligence research*, 2017, 13(2):225-241.
2. Wang R. Design and practice of the blended learning model based on an online judge system. *International Journal of Continuing Engineering Education and Life-Long Learning*, 2017, 27(1-2):45-56.
3. Rizk N, Attia K, Al-Jundi A. The Impact of Metacognition Strategies in Teaching Mathematics among Innovative Thinking Students in Primary School, Rafha, KSA. *International Journal of English Linguistics*, 2017, 7(3):103-103.
4. Dessie W M, Ademe A S. Training for creativity and innovation in small enterprises in Ethiopia. *International Journal of Training & Development*, 2017, 21(3):224-234.
5. Garad A, Gold J. The learning-driven organization: toward an integrative model for organizational learning. *Industrial and Commercial Training*, 2019, 51(6):329-341.
6. Biletska E M, Holovkova T A, Antonova O V. Practice of implementation of innovative means of teaching in forming of preventive thinking in students of higher medical educational establishments. *Medicini perspektivi (Medical perspectives)*, 2017, 22(1):20-24.
7. Cao K. Design and implementation of Internet plus mode in sports training and monitoring. *Revista de la Facultad de Ingenieria*, 2017, 32(16):811-817.
8. Dym, Clive L., et al. "Engineering design thinking, teaching, and learning." *Journal of engineering education* 94.1 (2005): 103-120.
9. Harasim, Linda. "Shift happens: Online education as a new paradigm in learning." *The Internet and higher education* 3.1-2 (2000): 41-61.
10. Chau, K. Y., Law, K. M., & Tang, Y. M. (2021). Impact of Self-Directed Learning and Educational Technology Readiness on Synchronous E-Learning. *Journal of Organizational and End User Computing (JOEUC)*, 33(6), 1-20.
11. Savić, M., Ivanović, M., Luković, I., Delibašić, B., Protić, J., Janković, D.: Students' Preferences in Selection of Computer Science and Informatics Studies - A Comprehensive

- Empirical Case Study. *Computer Science and Information Systems*, Vol. 18, No. 1, 251–283. (2021).
12. Hong, J. Y. , Ko, H. , Mesicek, L. , & Song, M. B.. (2019). Cultural intelligence as education contents: exploring the pedagogical aspects of effective functioning in higher education. *Concurrency and Computation Practice and Experience*.
 13. Romero, C.A.T., Ortiz, J.H., Khalaf, O.I., Ortega, W.M."Software Architecture for Planning Educational Scenarios by Applying an Agile Methodology".*International Journal of Emerging Technologies in Learning*, 2021, 16(8), pp. 132–144.
 14. Wang, Q., & Lu, P. (2019) "Research on Application of Artificial Intelligence in Computer Network Technology", *International Journal of Pattern Recognition and Artificial Intelligence*, 33(5), 1959015.
 15. N. Jamal, C. Xianqiao, F. Al-Turjman, F. Ullah, "A Deep Learning–based Approach for Emotions Classification in Big Corpus of Imbalanced Tweets", *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 20, no. 3, pp. 1-6, 2020.
 16. Man, N., Wang, K., & Liu, L. (2021). Using Computer Cognitive Atlas to Improve Students' Divergent Thinking Ability. *Journal of Organizational and End User Computing (JOEUC)*, 33(6), 1-16.
 17. H Wei, & Kehtarnavaz, N. (2019). Semi-supervised faster rcnn-based person detection and load classification for far field video surveillance.
 18. S Wan, L Qi, X Xu, C Tong, Z Gu. Deep Learning Models for Real-time Human Activity Recognition with Smartphones, *Mobile Networks and Applications*, 1-13, 2019.
 19. Z. Lv, H. Song, P. Basanta-Val, A. Steed and M. Jo, "Next-Generation Big Data Analytics: State of the Art, Challenges, and Future Research Topics," in *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 1891-1899, Aug. 2017.
 20. Ivanaj, S., Nganmini, G., & Antoine, A. (2019). Measuring E-Learners' Perceptions of Service Quality. *Journal of Organizational and End User Computing (JOEUC)*, 31(2), 83-104.
 21. Chen, S., Xiao, H., He, W., Mou, J., Siponen, M., Qiu, H., & Xu, F. (2021). Determinants of Individual Knowledge Innovation Behavior: A Perspective of Emotion, Knowledge Sharing, and Trust. *Journal of Organizational and End User Computing (JOEUC)*, 33(6), 1-24.

Guoxun Zheng, born in Dehui, Jilin Province in 1982, obtained a master's degree from Changchun University of Science and Technology. Now he works in Changchun Institute of Technology, associate researcher, engaged in higher education research and vocational education research. E-mail: zhengguoxun@ccit.edu.cn

Xiaoxian Zhang, born in Changchun, Jilin Province in 1980, obtained a master's degree from Changchun University of Science and Technology. Now she teaches in Changchun Institute of Technology, lecturer, engaged in higher education research and vocational education research. E-mail: zhangxiaoxian@ccit.edu.cn

Ruojin Wang, born in Songyuan, Jilin Province in 1982, obtained a bachelor degree from Jilin Normal University. Now he works in Jilin Communications Polytechnic, lecturer, engaged in higher education research and vocational education research. E-mail: 260908900@qq.com

Liang Zhao, born in Siping, Jilin Province in 1983, obtained a bachelor degree from Jilin Normal University of Science. Now he works in Jilin Education Examinations

Authority, associate researcher, engaged in higher education research and education examination research. E-mail: 105009162@qq.com

Chengming Wang, born in Songyuan, Jilin Province in 1980, obtained M.A. and PhD from Jilin University. Now he works in Jilin Education Examinations Authority, associate researcher, engaged in higher education research and education examination research. E-mail: 729998967@qq.com

Chunlai Wang, born in Dongliao, Jilin Province in 1981, obtained M.A. and PhD from Northeast Normal University. Now he works in Jilin Jianzhu University, associate professor, engaged in higher education research. E-mail: 63625851@qq.com

Received: June 08, 2021; Accepted: April 08, 2022.

Multimedia Teaching System Based on Art Interaction Technology

Xiaozhong Chen

Hainan Normal University Academy of Fine Arts,
Haikou 571100, Hainan, China
cxz15248935533@163.com

Abstract. With the development of the times, traditional classroom education has gradually failed to meet the needs of teaching. Now, with the application of computers in modern education, hybrid learning has developed rapidly in the field of education. How to promote the better development of hybrid learning has become a new research hot spot. Therefore, this paper aims at improving the MOOC platform, which is the largest application of hybrid learning. It integrates animation technology and multimedia technology, and designs a multimedia-teaching platform based on art interaction technology, which effectively improves the attraction of MOOC platform to learners. Firstly, this paper introduces multimedia technology, animation technology and interactive animation technology in detail, and applies them to MOOC platform. Secondly, according to the analysis of the research results of teaching platform requirements, the design principles and system framework of this paper are given. Finally, the information processing system of B/S architecture mode is built to make the improved platform have high response speed and data processing ability. In addition, this paper constructs a small-scale multimedia hybrid learning platform for testing, and finds that the multimedia teaching platform based on art interactive technology designed in this paper can well promote students' autonomous learning and improve the effect of students' learning.

Keywords: mixed learning, MOOC platform, multimedia teaching, art interaction.

1. Introduction

Since the beginning of the 21st century, the theoretical system of E-learning has been gradually established and improved, but the development of E-learning is not ideal. In 2001, the American Society for Training and Development reported that only 20% of corporate training was conducted in the form of E-learning, and E-learning was at low ebb. People began to reflect on this new way of learning, which led to the emergence of Blended Learning [1].

The main characteristics of MOOC curriculum are: large scale, openness, flexibility, easy to use, multi-source resources, participation autonomy, course audience, not limited by space and time, etc. Specifically, it is: (1) the scale of online learning is relatively large, and the number of people participating in curriculum learning is relatively large. (2) Openness: Regardless of the location, there is no threshold, as long as you want to

learn the interest of learning can participate in. (3) Flexibility: learning time is not single or gathered at a certain point, want to learn. We can use the fragmentary time to learn, and we don't need a fixed or specific time to learn. (4) Easy to use: the development of the teaching activities of "network course" can break through the traditional teaching requirements of time and space, connect the people who need to learn in various countries and regions of the world through the way of network connection, and learn the relevant courses they want to learn on the network. (5) Resource diversification: MOOC curriculum combines various forms of digital resources and various social network tools to form diversified and abundant curriculum resources and learning tools. (6) Participation autonomy: MOOC course not only has a higher choice rate, but also has a higher learning rate, which is inseparable from the need to learn a stronger sense of autonomy and a strong sense of honor, so as to complete the course learning content on time and learn the corresponding knowledge. (7) Course audiences: there is no restriction on the number of teaching staff, which can meet the needs of a large number of learners at the same time. The starting point of MOOC is people-oriented, which believes that everyone should have the right to learn. Every person with learning motivation and a person with learning ability in the world should acquire the survival skills they need in their life to improve their living environment, make their life better and make more contributions to their society [2-4].

Based on the advantages of MOOC platform, this paper introduces multimedia technology and animation interaction technology to improve the artistic quality of MOOC platform and strive to enhance the attraction of MOOC courses to learners. Multimedia, with the characteristics of text, picture, sound, image, and interaction, makes teaching intuitive and visual, greatly improves the transmission efficiency of knowledge and information, making teaching and learning is no longer a boring work, but a kind of enjoyment of art. Multimedia has interactive functions and hypermedia characteristics, making learning an autonomous behavior, truly reflecting the dominant position of students, while the leading role of teachers is to cut into the core of the teaching process [5]. The development and application of network technology, remote communication technology and virtual reality technology make multimedia teaching develop further. There will be rapid changes in modern educational technology. This topic will reveal the principles of multimedia teaching from the perspective of educational theory, study the practical functions of teaching system development from the perspective of teaching application, and provide a train of thought for the theoretical research and application research of multimedia teaching [6-8]. The development of art interaction technology has met the development needs of various fields. At the same time, the development of art interaction technology has broadened the new direction and field for the development of multimedia, which can enable students to learn more intelligently from multiple directions and angles. In the teaching process of many professional disciplines, vivid animation can make students feel high emotions in class, more attentive learning and thinking, and excellent works produced by animation can also shock students' hearts and make students feel the joy of creation and success. From the point of view of teaching effect, mature art design does not need teachers to teach too much. It can let students do their own work and brains, and let students interact and communicate with art. Let teachers teach knowledge into students' own exploration of knowledge, and let students actively explore knowledge in the true sense of realization [9-11].

In summary, on the basis of the advantages of MOOC platform, this paper introduces multimedia technology, animation technology and interactive technology, analyses the hierarchical structure and development process of the system, constructs information processing system using B/S architecture mode, and designs a multimedia teaching platform based on art interactive technology.

2. Related work

Hybrid learning mode is a new learning mode, which combines the traditional face-to-face learning mode and E-learning learning mode. It combines the advantages of traditional learning mode and E-learning learning mode. It not only avoids the neglect of students' autonomy in traditional teaching, but also alleviates the low utilization rate of E-learning because it can't completely replace traditional teaching. Mixed learning involves many fields, including school education, enterprise training, teacher training, adult education and so on. There are also increasing number of academic conferences, project topics and papers on hybrid learning. The Horizon Report of the New Media Alliance (2016 Higher Education Edition) points out that among the six key trends that may affect the reform of higher education in 2016-2020, the widespread use of blended learning design will be increasingly concerned as a short-term trend (within one year). In China, some universities have incorporated the results of blended learning into the credit system, adding the credits obtained by blended learning to the credits obtained by traditional classes as the total credits of the course for learners. The vigorous development of blended learning mode provides a new direction for the teaching mode of open universities.

In 2008, the concept of MOOC (Massive Open Online Course) was formally proposed by Canadian scholars Dave Cormier and Bryan Alexander, that is, large-scale open online courses. In 2012, MOOC got a breakthrough development. Udacity, Coursera and edX, three major MOOC website platforms, were established and operated one after another. Millions of learners registered for online courses. MOOC swept the educational circles with a trend of giving up others and attracted wide attention from all walks of life. The New York Times called it "the first year of MOOC". With the entry of Harvard University, Stanford University, MIT and other world top universities, nearly 100 famous universities around the world have also invested in MOOC platform, and there are more than 500 online courses. In May 2013, the curriculum projects of six famous Asian universities, including Peking University, Tsinghua University, Hong Kong University and Hong Kong University of Science and Technology, were launched on edX platform. In July 2013, Shanghai Jiaotong University and Fudan University signed a contract with Coursera, one of the troikas. According to Coursera, nearly 130,000 Chinese learners registered on the Coursera platform in 2013, while in 2014, the number of registered learners increased fivefold to 650,000. In September 2016, the Ministry of Education issued Opinions on Promoting Credit Recognition and Conversion in Higher Education, pointing out that all kinds of college students can obtain credits not only by learning their own courses, but also by learning out-of-school courses and converting non-academic achievements. Encouraging students to take elective courses in universities or on the basis of Internet learning platform and

promoting credit recognition and conversion in higher education will adhere to learner-centered, university-centered, substantive equivalence and standardized and orderly, and the credit recognition and conversion system in higher education will be more perfect by 2020.

3. Key Technology Overview

3.1. Multimedia Technology

Multimedia network teaching breaks the limitation of region and time and space, realizes many functions such as bidirectional interaction, real-time multi-point communication, wide dissemination, fast data and information collection, and relies on network communication technology and multimedia technology. Therefore, in the multimedia network teaching, students can share learning resources, discuss and communicate problems together, and through a variety of media (such as audio, video, image, text, etc.) to enable students to strengthen memory and improve understanding, in order to avoid the original monotonous and mechanical learning, improve students' learning initiative. It can also enable students to exchange information with computers actively and frequently, and get timely feedback. It can be said that the interactive teaching of network multimedia has changed the process of students' understanding of things in the teaching process, changed the traditional teaching principles, changed the teaching content and the form of teaching materials.

Streaming media is simply a multimedia file (audio, video, animation or other multimedia files) transmitted over the network by streaming technology. Streaming technology is a network transmission technology that compresses the continuous video and audio information and places it on the server. Users can download it while watching, without downloading the entire compressed file to the local computer [12-14]. Media in streaming media can be audio, video, animation or other multimedia files.

Streaming transmission has two modes: sequential streaming transmission and real-time streaming transmission [15,16]. Generally speaking, real-time streaming transmission is used when video is transmitted in real time, or when streaming media servers or real-time protocols such as RTP and RTCP are used. RTP is a real-time transmission protocol developed by IETF [17,18]. If an HTTP server is used, the files are sent through a sequential stream. There are some differences between real-time streaming transmission and sequential streaming transmission. Special streaming media server and transmission protocol are used for real-time streaming transmission. Real-time streaming has the advantages of realizing live broadcast, broadcasting and multicast, random access to material, saving user's disk space and not wasting bandwidth. The disadvantage of real-time streaming is that dedicated streaming servers are needed, and lost packets will be lost permanently.

Real-time transport protocol (RTP) is a transmission protocol for multimedia data streams over the Internet. It is defined as one-to-one or one-to-many transmission. Its purpose is to provide time information and synchronize multimedia data streams. RTP

usually transmits data on the basis of UDP protocol. However, RTP does not provide any mechanism to guarantee the transmission quality, which is accomplished by RTP.

3.2. Animation Technology

Animation is a series of continuous playing pictures, these pictures are static, each adjacent two pictures have slightly different, when these slightly different static pictures continue to pass through the eyes, we feel that the scene in the picture is moving, so the essence of animation is movement, is the continuous movement of multiple pictures [19].

The basic principle of animation is "visual pause". Human vision has a temporary characteristic. It means that when the human eye sees a picture or an object, the image of the picture or thing will stay in the human vision for a short time and will not disappear for at least 0.25 seconds. Animation is to use this visual principle to produce a series of continuous changes in the picture, so that the previous picture in the human vision has not disappeared before playing the next picture, thus causing a continuous change in the visual effect. Animation can't be played too slowly. If a single picture stays in the human vision for more than 0.25 seconds, we can feel the incoherence of the picture. What we see is a static picture.

In the 1950s and 1960s, with the birth and promotion of computers, people began to use computers to produce animated films, thus stepping into the era of computer animation. It not only pushed the film and television technology to a climax, but also played an irreplaceable role in scientific research, health care, education and teaching.

Computer animation is generally divided into two categories: two-dimensional animation and three-dimensional animation. There is much software for making two-dimensional animation. Flash is a rising star, but also the current mainstream [20]. Flash is widely used because it generates vector graphics, has small files and fast propagation speed. Flash does not have the function of three-dimensional modeling, but it can import the three-dimensional animation created by other software into Flash to synthesize. Two-dimensional animation production software also includes Ulead GIF Animator, COOL 3D, Firework and so on. They have different functional characteristics, so the animation style is different, which can improve the artistry of courseware.

Because animation has the characteristics of image, continuity, and narrative, the animation design teaching platform will be able to play the advantage of animation to excel and vividly shape objects and storytelling. The visual beauty of animation narrative and intuitive beauty will be Close the psychological distance between the platform and the students, so that students have higher acceptance, faster acceptance and better learning results.

For courseware and other teaching materials, PPT and Flash can meet the requirements of teachers. But Flash tends to create animation, and can develop independent courseware integrating video, sound, image, text and animation. PPT tends to demonstrate graphics and text. Although Flash can express the same content, it is inferior to Flash in form. In the actual survey and analysis results, we can also see that Flash animation courseware can attract students' attention better than PPT demonstration courseware, and in the application of animation courseware in the classroom learning atmosphere is more active, the interaction between teachers and students is better. Based

on the simplicity of Flash operation, high compatibility and independence, and strong interaction ability, this paper chooses Flash as the development software of this artistic interaction-teaching platform.

3.3. Interactive Animation Technology

Interactive animation is the product of the combination of computer graphics and art. It is a new subject with the rapid development of computer hardware and graphics algorithms. It makes use of the knowledge of computer science, art and other related disciplines to create colorful and continuous virtual reality pictures on computers, providing people with a new world to fully display their imagination and artistic talent. Computer graphics is widely used in fine arts and business arts. Artists use a variety of computer methods, including dedicated hardware, commercial software packages, symbolic mathematics programs, CAD software packages, desktop publishing software and animation software to design the shape of objects and describe the movement of objects [21-23].

Interaction is the basic feature of interactive animation. Interaction is a term in computer technology that allows users to exchange information with computer systems. Interactive is to realize the interaction between human and computer through program. The way of human-computer interaction is actually an interactive process. When designing a program to achieve a certain function, it intends to set some links in the middle of the program that users can choose or decide according to their own needs. When running the program, all the operation methods of the computer are displayed on the monitor. As long as the user acts on the computer according to the prompts or suggestions presented on the monitor, the computer can automatically process the corresponding work according to the user's operation, so that the program can run according to the user's needs.

The application of interaction technology makes people's communication reach unprecedented breadth and depth. The types and quantities of information that digital media can carry are unparalleled by traditional media. The development of interactive technology makes people digitize their various behaviors. Digital information can be controlled by these digitized instructions. The application of interactive technology makes our interactive design more intuitive and convenient. Interactive animation will guide students step by step into the learning content of platform design, learning will become an autonomous behavior, let students feel the dominant position in learning, stimulate students' desire for exploration and curiosity, so that the original boring learning content becomes lively and interesting.

4. Design of Multimedia Teaching \platform based on Art Interactive Technology

4.1. Platform Design Principles

In order to design the teaching platform more reasonably, this paper adopts the following principles to design the teaching platform:

1. With business as the core, business modules are reasonably divided. The teaching system platform is designed according to the characteristics of the education industry and from the perspective of educational and teaching needs to meet the needs of daily teaching management, teachers' teaching, teacher training and students' self-learning of the education management departments.

2. Define the interface clearly, improve the reusability of service, define the interface of service clearly, and distinguish the boundary between service interface and internal implementation. Service design should be reusable, defining abstract service interface, stabilizing abstract interface, changing its specific implementation class when business changes, so as not to affect the call between systems, realizing the user's requirements for complex work changing with demand, and adapting to business process changes.

3. Expansibility and compatibility: The system adopts standardized interface design, and business modules are independent of each other to meet the expansion needs of future business changes. The message communication system with standard XML format is compatible for data exchange.

4. Emphasis is placed on practicality and effectiveness: the system design emphasizes practicality, practicality and effectiveness, and meets the actual needs of teachers' digital and networked lesson preparation and teaching. The interface of the system is simple, the layout is reasonable, and the operation is simple.

5. Emphasis on interactivity and fun: The system design emphasizes the reasonable layout of the interface, has strong interactivity, is interesting and easy to operate, and improves the effectiveness of student platform learning.

6. Advancement and stability: Emphasis is laid on service-oriented design method and SOA architecture, J2EE system architecture, following the industry's open standard system, to ensure the portability and stability of the system.

7. Reliability of implementation: Using multiple clusters and high availability mechanisms, it has the ability of fault tolerance and disaster tolerance, and can meet the needs of applications in a variety of environments.

4.2. System Frame Structure

The whole system can be divided into software support part and hardware support part. The software part includes multimedia teaching program, animation resources, and the hardware part includes computer, display and so on. Based on the research and analysis of the requirements of the teaching system, this paper divides the platform into six subsystems: resource system, lesson preparation system, teaching system, test system,

interactive system and information system. Figure 2 shows the system framework of this paper. The system is built on the basis of hardware devices. The UI interface is used for teachers and students. The local or network teaching resources, including text information, image information and animation information, are used through other controls such as communication control and call control.

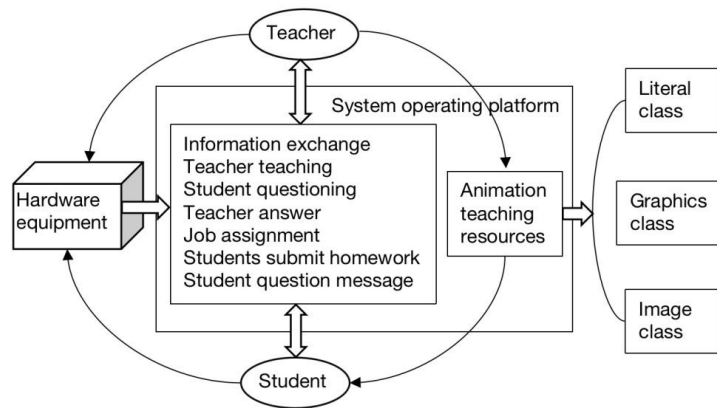


Fig. 1. System framework diagram

4.3. Teaching Platform Architecture Design

Nowadays, the most popular software architecture patterns are C/S structure and B/S structure [24, 25]. C/S (Client/Server) structure is client and server structure. This structure can make full use of the advantages of hardware environment at both ends, and assign tasks to Client and Server to realize reasonably, which reduce the communication overhead of the system. However, both Client and Server need specific software support. The software of this structure is not easy to transplant. Different versions of the system need to be developed for different operating systems. It is difficult to adapt to the simultaneous use of more than 100 local area network users. Moreover, it is expensive, inefficient and difficult to upgrade and maintain. B/S (Browser/Server) structure, browser and server structure, is a kind of structure that changes or improves C/S structure with the rise of Internet technology. There are many advantages in B/S structure system:

1. Client

In B/S structure, browser implements user interface, and very few things logic is implemented in Browser. Transaction logic is mainly implemented in Server, which can simplify the computer load of client. The operation and design of the network-teaching

platform should be considered for the client, to minimize the maintenance costs of the client and to reduce the requirements for the software and hardware equipment of the client.

2. Easy system maintenance and upgrade

Because server side mainly implements transaction logic, all clients are browsers. The software of B/S architecture only needs to manage the server. The client does not need any maintenance. The administrator can easily monitor the operation of the server and control the number of users accessing the server and using the server application. This can simplify the administrator's management of the system, reduce the workload of system maintenance, and ensure the reliable operation of the system. If it's different, it only needs to connect the server to the private network to achieve remote maintenance, upgrade and sharing.

3. Reducing development cost and improving system security

Based on the current technology, the local area network establishes the B/S structure of the network application, and through the Internet/Intranet mode database application, it is relatively easy to grasp and the cost is low. Since the B/S architecture software has no dependency on the server operating system, it can be installed on a free Linux server with high security. Because all operations are only for the server, it can effectively protect the data platform and manage access rights, and the server database is also safe. Combined with cross-platform language programming, B/S architecture management software is more convenient, fast and efficient.

4. Multi-selectivity of network hardware environment

B/S structure does not need special network hardware environment, such as telephone access, renting equipment, information management by itself; generally as long as there is an operating system and browser, because it is a one-time development, can achieve different personnel, from different locations, access and operate the common database in different ways (such as LAN, WAN, Internet/Intranet, etc.).

Therefore, the system development can adopt three-tier software architecture mode, namely, presentation layer, logic layer and storage layer, and B/S structure as the solution of the system. The system architecture is shown in Figure 2.

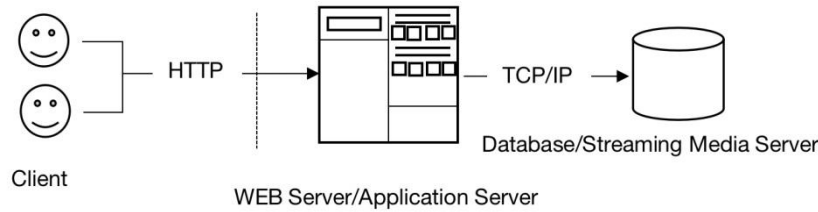


Fig. 2. System architecture

In the three-tier software architecture, the presentation layer is user interface, and the logic layer is divided into application layer and data interface layer. The storage layer is mainly used to store data, that is, physical database layer. The system is divided into three parts: database, application server and browser client. All users access the application server through the browser, and the application server and the database server interact to obtain the required data. User side adopts browser form, which can satisfy the usability requirement of zero client side.

5. Realization of Multimedia Teaching Platform based on Art Interactive Technology

5.1. Database Implementation

The basic functions of the teaching platform include students' question answering, transmission of teaching materials and notification of teaching information. Therefore, the key point of the platform construction is how to realize data information operation and processing and access of large object data files. Of course, connecting SQL Server database through MFC in VS is the most basic problem.

Accessing databases through data sources requires a driver engine. Common engines include ODBC, OLE DB and ADO [26,27]. This platform uses ADO (ActiveX Data Objects) control. ADO is an optimized set of dedicated objects, providing a complete solution for site database access. Users can use SQL instructions to add, modify and clear the data of the background server in the interactive interface. The ADO interface uses RecordSets objects. It not only supports multiple languages to access databases and output query results, but also links ODBC databases, such as SQL Server, Access, Oracle and so on [28, 29]. In MFC, statements are needed to refer to type libraries that support ADO components. Type libraries can be positioned as part of executable programs in their own affiliated resources, and initialization components are added to realize initialization components, and the components are closed at the end. In this way, ADO control can be used directly to complete the connection to the database, and SQL statements can be used for data processing. The basic processing methods mainly include query, addition, modification and deletion.

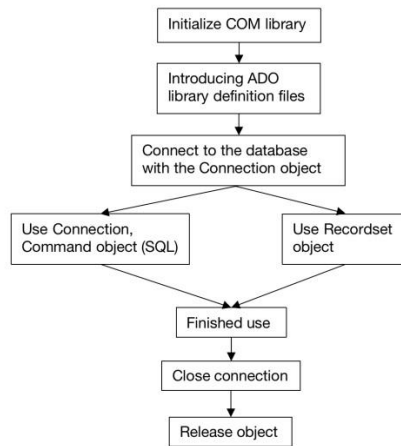


Fig. 3. System database development process

5.2. Interface Effect Diagram of Teaching Platform

Based on the advantages of MOOC platform, this paper improves and designs a multimedia-teaching platform based on art interactive technology. When the platform runs, it first enters the landing interface, as shown in Figure 4. As shown in the figure, users can login to the platform by entering their own username and password. The system will judge whether the landing user belongs to a teacher, a student or an administrator according to the database data, and then enter the corresponding interface.

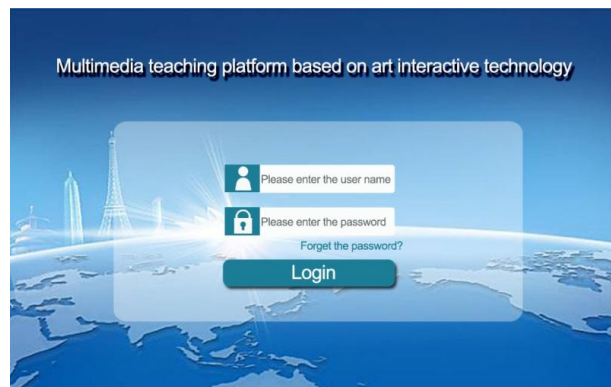


Fig. 4. User login interface

If students log in, they can enter the user interface, which is shown in Figure 5. As can be seen from the graph, the test system has two columns: the homework subsystem and the examination subsystem. Students can test in the system to verify what they have

learned. The test results will be retained. Teachers can inquire about the results so as to understand the students' learning situation. In the lower part of the interface, two interactive keys can be seen from the last question to the next question. Clicking can realize the function of jumping from the top to the bottom question.

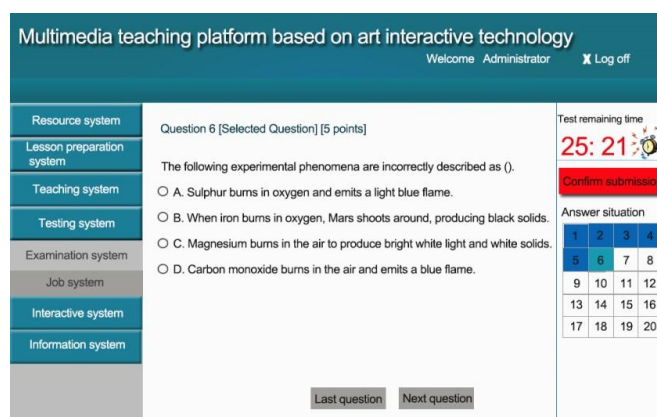


Fig. 5. Test system usage interface

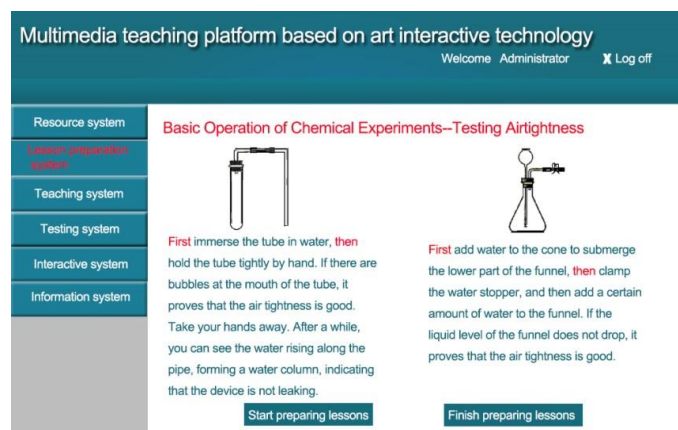


Fig. 6. Use interface of lesson preparation system

If the teacher logs on to the system, he can choose the lesson preparation system. The interface diagram of the lesson preparation system is shown in Figure 6. From the figure, we can see that after entering the interface, we can enter the course preparation process by clicking the interactive button to prepare lessons. Through animation technology, Flash software is used to design and process the material, so that it can be presented in different media forms such as text, image, picture-story book, video, audio and so on. Finally, Flash is used to organize and integrate all the contents, to make more

vivid/artistic and interactive teaching courseware with picture-text collocation and combination of dynamic and static cases; after the courseware is made, the interactive button for preparing lessons can be clicked and the courseware can be withdrawn. Courseware with vivid animation effect is undoubtedly more attractive for students to learn, and vivid animation, so that students can consciously think about the teaching content, improve the learning effect.

6. Test and Analysis of the Effect of Platform Implementation

Before the test, this paper conducts a questionnaire survey among 129 students to analyze the current situation of mixed learning using MOOC platform. The results of the survey are shown in Figure 7. From the figure, we can see that most of the students surveyed have the habit of using computer to study, and willing to use interactive media and other new media means, and only 3.5% of them do not use it. This data shows that today's learning is not confined to traditional classroom learning. Mixed learning using MOOC platform has become an important part of learning.

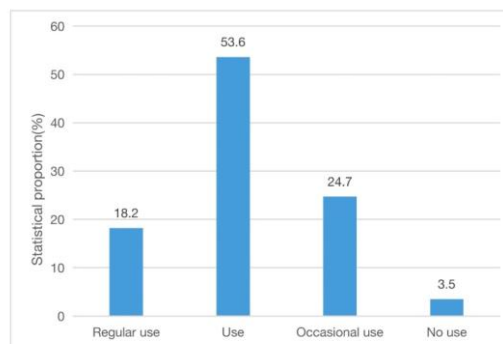


Fig. 7. Whether to learn using the MOOC platform

Secondly, 100 users of the small-scale teaching platform constructed in this paper have been surveyed. All 100 users have used MOOC platform to learn. This paper compares the learning effect, operation effect, interface design and transmission speed with MOOC platform, and the comparison results are shown in Figure 8. From Figure 8, we can see that compared with MOOC platform, users are more satisfied with the learning effect and operation effect of the teaching platform designed in this paper, and the satisfaction is more than 70%. Among them, the unsatisfactory aspects are interface design and transmission speed, which are 21.9% and 16.3% respectively. This shows that compared with MOOC platform, these two aspects need to be strengthened. Of course, this paper only builds a small-scale test platform. Compared with the mature MOOC platform, it is reasonable that interface design and transmission speed are insufficient. However, the introduction of interactive technology and animation technology in this paper can effectively increase the attraction of MOOC courses to learners.

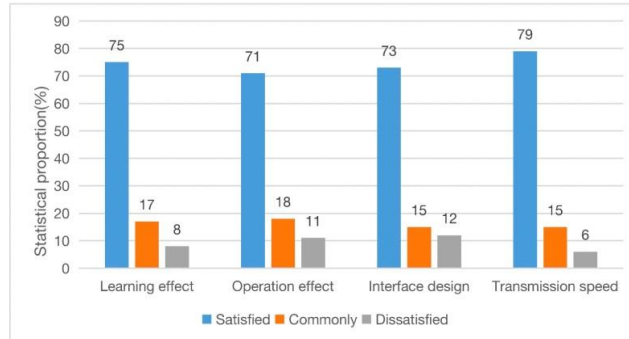


Fig. 8. User experience survey results

Finally, this paper uses the experimental class and non-experimental class to test the effect of use, in which the experimental class uses the teaching platform designed in this paper to guide learning. After a semester, the test scores are shown in Table 1. From Table 1, we can see that the results of the experimental class are better than those of the non-experimental class. This shows that the multimedia-teaching platform based on art interactive technology designed in this paper can promote students' learning and improve the effect of students' learning.

Table 1. Platform effect test data

Test class	Average score of test scores			
	Mathematics	English	Chemistry	Physics
Non experimental class	71.7	70.9	73.5	74.1
Experimental class	82.3	80.7	88.6	89.3

7. Conclusion

Nowadays, in the era when traditional classroom education can't meet the needs of teaching, hybrid learning has gradually risen, and the MOOC platform, which is more representative of hybrid learning, has led the learning trend and become the latest research hot spot. Based on the advantages of MOOC platform, this paper studied the advantages of multimedia technology, animation technology and animation interaction technology in MOOC platform design, and designs a multimedia platform based on art interaction technology. The platform designed in this paper gives the system framework after analyzing the research results of the teaching platform requirements, and designs the teaching platform with B/S mode. In addition, the idea of database development and the interface effect diagram of the small-scale teaching platform constructed in this paper are given. Through the questionnaire survey and the test results of the

experimental class, we can see that MOOC platform has become the mainstream of online learning for students. Compared with the past, the MOOC platform designed in this paper uses art interaction technology. Improve the artistry, fun and operability of the platform, enhance its appeal to learners, and entertain and educate, so that students can actively learn, feel the joy of creation and success, and improve learning.

Data Availability. Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Conflict of Interest. The author states that this article has no conflict of interest.

Funding. The author received no financial support for the research, authorship, and/or publication of this article.

References

1. Porter W W, Graham C R, Bodily R G, et al. A qualitative analysis of institutional drivers and barriers to blended learning adoption in higher education. *Internet & Higher Education*, 2016, 28(1):17-27.
2. Aksela, Maija|Wu, Xiaomeng|Halonon, Julia. Relevancy of the Massive Open Online Course (MOOC) about Sustainable Energy for Adolescents.. *Education Sciences*, 2016, 6(4):40.
3. Engle, Deborah|Mankoff, Chris|Carbrey, Jennifer. Coursera's Introductory Human Physiology Course: Factors That Characterize Successful Completion of a MOOC.. *International Review of Research in Open & Distributed Learning*, 2015, 16(2):46-68.
4. Zhang, Min|Yin, Shuaijun|Luo, Meifen|Yan, Weiwei. Learner Control, User Characteristics, Platform Difference, and Their Role in Adoption Intention for MOOC Learning in China.. *Australasian Journal of Educational Technology*, 2017, 33(1): 114-133.
5. Solorio R, Nortonshepk P, Forehand M, et al. Tu Amigo Pepe: Evaluation of a Multimedia Marketing Campaign that Targets Young Latino Immigrant MSM with HIV Testing Messages.. *Aids Behav*, 2016, 20(9):1973-1988.
6. Lv J. Research of Japanese Translation Teaching Based on Multimedia Network Technology. *Journal of Computational & Theoretical Nanoscience*, 2016, 13(12):10375-10379.
7. Kollmann T. Attitude, adoption or acceptance? Measuring the market success of telecommunication and multimedia technology. *International Journal of Business Performance Management*, 2004, 6(2):133-152.
8. Maples-Keller J L, Bunnell B E, Kim S J, et al. The Use of Virtual Reality Technology in the Treatment of Anxiety and Other Psychiatric Disorders. *Harv Rev Psychiatry*, 2017, 25(3):103-113.
9. Scherzer J, Buchanan M F, Moore J N, et al. Teaching veterinary obstetrics using three-dimensional animation technology.. *Journal of Veterinary Medical Education*, 2010, 37(3):299-303.
10. [10]Xiong X, Yao B, Ouyang K, et al. Study on virtual animation technology of sheet metal bending. *Forging & Stamping Technology*, 2014, 39(3):15-19.
11. Zhao W, Xie X F. Development of Virtual Human Technology and Its Engineering Application. *Journal of System Simulation*, 2009, 21(17):5473-5476.
12. Wang B, Sen S, Adler M, et al. Optimal Proxy Cache Allocation for Efficient Streaming Media Distribution.. *IEEE Transactions on Multimedia*, 2004, 6(2):366-374.

13. Guo L, Chen S, Zhang X. Design and Evaluation of a Scalable and Reliable P2P Assisted Proxy for On-Demand Streaming Media Delivery. *IEEE Transactions on Knowledge & Data Engineering*, 2006, 18(5):669-682.
14. Pan D, Xin Z, Xiong L, et al. Buffer management for streaming media transmission in hierarchical data of opportunistic networks. *Neurocomputing*, 2016, 193(C):42-50.
15. Wei S, Zhuang W. Performance Analysis of Probabilistic Multipath Transmission of Video Streaming Traffic over Multi-Radio Wireless Devices. *IEEE Transactions on Wireless Communications*, 2012, 11(4):1554-1564.
16. Ma C, Yang Y. Battery-Aware Routing for Streaming Data Transmissions in Wireless Sensor Networks. *Mobile Networks & Applications*, 2006, 11(5):757-767.
17. Eken M M, Dallmeijer A J, Doorenbosch C A, et al. Assessment of muscle endurance of the knee extensor muscles in adolescents with spastic cerebral palsy using a submaximal repetitions-to-fatigue protocol. *Archives of Physical Medicine & Rehabilitation*, 2014, 95(10):1888-1894.
18. [18]Yu C M. RTCP: reliable topology construction protocol of bluetooth hybrid single-hop and multi-hop networks. *Iet Communications*, 2018, 12(2):136-143.
19. Mcghee J. 3-D visualization and animation technologies in anatomical imaging. *Journal of Anatomy*, 2010, 216(2):264-270.
20. Grasso S, Saunders T, Porwal H, et al. Flash Spark Plasma Sintering (FSPS) of alpha and Beta SiC. *Journal of the American Ceramic Society*, 2016, 99(5):1534-1543.
21. Kakinuma A, Nagatani H, Otake H, et al. The effects of short interactive animation video information on preanesthetic anxiety, knowledge, and interview time: a randomized controlled trial. *Anesthesia & Analgesia*, 2011, 112(6):1314-8.
22. Dan C, Tejera M, Guillemaut J, et al. Interactive Animation of 4D Performance Capture. *IEEE Transactions on Visualization & Computer Graphics*, 2013, 19(5):762-773.
23. Hui L, Deng S, Jian C, et al. Semantic framework for interactive animation generation and its application in virtual shadow play performance. *Virtual Reality*, 2018, 22(2):149-165.
24. Guimaraes T, Igarria M. Client/Server System Success: Exploring the Human Side. *Decision Sciences*, 2010, 28(4):851-876.
25. Wang S, Chao H, Liu S, et al. Establishment on Space Objects Database Management System Using Browser/Server Mode. *Procedia Engineering*, 2012, 29:1071-1074.
26. Huang B H, Wang T J, Ma Y, et al. Schema of Enhancing User Authentication and Encrypting Lob Data in ODBC Driver for Database. *Applied Mechanics & Materials*, 2014, 543-547:3688-3691.
27. Burger K R. Using activeX data objects to publish an Excel grade book on the World Wide Web. *Journal of Computing Sciences in Colleges*, 2001, 16(16):341-352.
28. Aguirre, C. C., González-Castro, N., Kloos, C. D., Alario-Hoyos, C., Muñoz-Merino, P. J.: Conversational agent for supporting learners on a MOOC on programming with Java. *Computer Science and Information Systems*, Vol. 18, No. 4, 1271–1286. (2021).
29. D. Kavadi, F. Al-Turjman, K. Reddy, R. Patan, "A Machine Learning Approach for Celebrity Profiling", *International Journal of Ad Hoc and Ubiquitous Computing*, vol. 38, no. 1/2/3, pp. 111-126, 2021.

Xiaozhong Chen was born in Changchun, Jilin Province, China, in 1969. Obtained a master's degree in design from Jilin Art Institute of the People's Republic of China. He is currently studying at the University of San Carlos in the Philippines, pursuing a Doctor of Philosophy in Education. The work unit is the Academy of Fine Arts of Hainan Normal University. His research interests include computer digital media technology, film and television animation, industrial modeling, educational philosophy, and art education. E-mail: cxz15248935533@163.com

Received: April 05, 2022; Accepted: May 15, 2022.

The Research and Implementation Feasibility Analysis of an Intelligent Robot for Simulating Navigational English Dialogue under the Background of Artificial Intelligence

Wei Sun

Foreign Language College, Anhui Xinhua University,
Hefei 230000, Anhui, China
sunwei@axhu.edu.cn

Abstract. The rapid development of artificial intelligence and robots, the research and development of intelligent dialogue robots are very necessary for today's society. However, a robotic algorithm system that simulates navigational English conversation has not yet been developed. In order to find a suitable algorithm system for dialogue robots, this paper uses the test data set to test the analytical model of navigational English dialogue instructions. The experimental results show that the conditional random field (CRF) + domain dictionary + ambiguity resolution method has the highest segmentation effect. The calculated percentages of the analytical model are correct rate: 76.85%; recall rate: 80.36%; F-value: 88.46%. This paper implements a robot teaching and reproduction method based on simulated navigational English conversation and human-computer interaction under the background of artificial intelligence, and designs robot motion realization experiments and speech recognition experiments. The three-dimensional error after fine-tuning the voice is between 1.6798mm and 2.9968mm. This article constructs a simulation navigational English dialogue robot system. The FAQ component has up to 79.2%; others have a lower accuracy rate of only 59.03%.

Keywords: Artificial Intelligence, Simulated Navigational English Conversation, Dialogue Robot, Teaching and Reproduction, Speech Recognition, Two-Dimensional Error, Conditional Random Field (CRF).

1. Introduction

As AI advances, robots have become the mainstream direction of robot development, and they have increasingly received the attention and attention of researchers. Therefore, the research and development of intelligent dialogue robots are very necessary for today's society.

At present, most of the simulation dialogue robots are still pre-set voice instructions, and then perform specific tasks according to the set instructions, which does not achieve the real human-computer intelligent interaction that people desire [1]. When people express natural language, some words will be omitted because of the context environment, which leads to the text information after speech recognition cannot be

correctly converted into the control instructions that can be executed by the simulation dialogue robot. It is unrealistic to simply use the knowledge stored in the knowledge base itself to include all possible situations. This requires the simulation dialogue robot to use a dialogue management technology combined with knowledge base to judge and inquire these unknown or missing information, to complete the missing information by reasoning, and to store the unknown words. As an unstructured way of expression, natural language can be understood by human beings in line with human habits, but simulation dialogue robots tend to understand structured language [2]. How to take efficient methods to process navigational English instructions and convert them into executable instructions of simulation dialogue robots is a subject that needs further study.

In this paper, the test data set is used to test the established model of navigational English dialogue instruction analysis. In this paper, the robot teaching and playback method based on the simulation of navigational English conversation and human-computer interaction is realized, and the robot action realization experiment and speech recognition experiment are designed to verify the method proposed in this paper. This paper constructs a simulated navigational English dialogue robot system which can use knowledge to understand problems on a large scale.

2. Related Work

Many research teams at home and abroad have conducted in-depth research on the scheme research of simulated navigational English conversation under the background of artificial intelligence. In [3], the authors evaluated whether playing a dialogue game could motivate participants to participate in an advance care plan (ACP). The study found that people who played dialogue games had a higher ACP behavioral execution rate within 3 months. In [4], the author used a linear regression model to analyze the random intercept of the speaker. The results also show that the VOT of Hawaiian words ($\beta = -0.01$, $t(289) = -2.0$ and $p < 0.05$) is significantly shorter. In [5], an experienced language teacher is tested in a weekly one-to-one navigational English conversation under the background of artificial intelligence. The results show that all three recasting forms can effectively help learners improve the accuracy of navigational English past tense. In [6], the results showed that the intervention reduced the auditory recognition of the random number formula. The accessible, universal and multilingual nature of popular SNSs such as Facebook has inspired many scholars of second language teaching [7]. In [8-9], the research results show that the interaction mode of navigational English teacher educators is both heterogeneous and homogeneous. In [10], the author studied conversation excerpts from 5 academic conversations, exploring the different ways of metaphorical ideation and the reactions they elicited in conversation partners. In [11], the author checks the acoustic changes of vowels in the speaker during the whole speech task. The overall goal of this study is to understand the differences among speakers as an indicator of the range of normal vowel movement in American navigational English. In [12-13], the author studies the extended use of multiple phone corpora based on three languages. In [14-15], the author used dialogue analysis to study the repair sequence corpus related to pronunciation among Chinese and Japanese students in a Japanese

university. The research claims that three segmental repair strategies are used in the interaction process to maintain mutually understandable pronunciation. In [15], as many as 93% of the endings were released by Putonghua speakers, 41% of which were not released by Cantonese speakers. In addition, it is found that Mandarin speakers do not consume human voice, but suck the end with 58% strong voice, which is contrary to most previous studies. In [16-17], the author explores how learners can take collaborative planning tasks as local emergency activities in class. The analysis shows that group planning is essentially a non-linear, social and practical activity, in which students' management participants work together to achieve effective task completion.

Aiming at the algorithm research and system development of intelligent robot, many research teams at home and abroad have carried out in-depth research. In [18], the author evaluates the commercial robots that have been deployed in suburban houses. In [19], the author considers the application of artificial agent in the navigation of semi-automatic mobile robot in the environment with obstacles. In [20], the author puts forward a new type of articulated cantilever sensor structure, analyzes and compares the simple cantilever sensor and articulated sensor. In [21], the author also analyzes the "kidnapping robot problem" and puts forward practical solutions. In [22-23], the author puts forward a new strategy of people to people cooperation and interaction. The proposed strategy is based on data fusion between inertial measurement unit and laser rangefinder to obtain human gait parameters. Finally, the controller shows how the walkers' orientation follows the human orientation in the actual experiment [24]. The unique performance of intelligent materials greatly improves the performance of soft robots [25]. In [26-27], the author proved the feasibility of rotating sensors and smart phone brain for mobile robots. In [28], from the system point of view, the author reviewed the latest work of underwater robots supporting IPMC from the perspective of modeling, manufacturing and biologically inspired design. In [29-31], the author's research results demonstrate that cultural genetic algorithms can effectively solve delays and problems, avoid falling into a peak and guide the mobile robot more effectively.

3. Method

3.1. Navigational English Word Segmentation Algorithm Model Based on Conditional Random Field

Set $G = (V, E)$ as undefined, V is a node in the graph, and E is an undirected edge in the graph. $Y = \{Y_v | v \in V\}$. Let us define an observation sequence X . If each random variable Y_v satisfies the following formula, then:

$$p(Y_v | X, Y_o, \omega \neq v) = p(Y_v | X, Y_o, \omega \approx v) \quad (1)$$

(X, Y) is called a conditional random field, where the parameter $\omega \approx v$ represents two adjacent nodes in the graph. Its advantage is that it only needs to take into account

the characteristics of the current observed state, and there are no special requirements for independence. In natural language processing, there is:

$$p(Y_i | X, Y_1, Y_2, \dots, Y_n) = p(Y_i | X, Y_{i-1}, Y_{i+1}) \quad (2)$$

Given a given observation sequence X with a value of x and a random variable Y with a value of y, the parameter expression is shown as:

$$P(y | x) = \frac{1}{Z(x)} \exp\left(\sum_{i,k} \lambda_k t_k(y_{i-1}, y_i, x, i) + \sum_{i,l} \mu_l s_l(y_i, x, i)\right) \quad (3)$$

Where Z(x) is the normalization factor, λ K and μ L are the corresponding weights, TK and SL are the characteristic functions, K and l represent the number of characteristic functions, and I represent all possible values of Y. The conditional probability defined by CRF is expressed by formula:

$$p(Y | X, \lambda) = \frac{1}{Z(X)} \exp(\lambda_1 F_j(Y, X)), \quad Z(X) = \sum_Y \exp(\lambda_j F_j(Y, X)) \quad (4)$$

The conditional random field model, like other probability graph models, also needs to solve three standard problems: feature selection, parameter training and decoding. Viterbi algorithm is used in the process of label prediction.

Probabilistic combinatorial category grammar is to apply probabilistic knowledge to combinatorial category grammar, and use the distribution of probability to carry out the optimal solution, so as to eliminate the ambiguity of analytical results, and to eliminate the ambiguity by introducing probability and finding the maximum value of probability. If x is a natural language sentence, Z is a semantic form, and Y is a syntax structure that guides generation, the model representation is as shown in the formula:

$$P(y, z | x, w, \Phi) = \frac{\exp^{w\varphi(x,y,z)}}{\sum_{(y',z')} \exp^{w\varphi(x,y',z')}} \quad (5)$$

Where, Φ is the dictionary, $\varphi(x, y, z) = \{f_1(x, y, z), \dots, f_n(x, y, z)\}$ is the eigenvector of N dimension, $w \in R^n$ is the parameter vector, corresponding to the eigenvector, $w \cdot \varphi(x, y, z) = \sum_{j=1}^n w_j \cdot f_j(x, y, z)$. If there are N training samples

$\{(x_i, z_i), i = 1 \dots n\}$, the parameter estimation is the optimal w that maximizes the value of the corresponding log likelihood function by solving, as shown in

$$L(w) = \sum_{i=1}^n \log\left(\sum_y P(z_i, y | x_i; w, \Phi)\right)$$

For the method of solving the log likelihood estimation function, we can use gradient descent algorithm or inward outward algorithm. Relatively speaking, the efficiency of inward outward algorithm is higher.

3.2. Answer Sorting Algorithm based on Robot Conversation System

(1) Listwise sort learning algorithm. Create a feature vector $x_j^i = \psi(q^{(i)}, d_j^{(i)})$, $i = 1, 2, \dots, m$; $j = 1, 2, \dots, n^i$ a for each (query document) pair. Each eigenvector list $x^{(i)} = (x_1^i, \dots, x_n^i(i))$ and its score set $y^{(i)} = (y_1^i y_2^i, \dots, y_n^i)$ constitute a training example.

Given a eigenvector list $x^{(i)}$, we can get a fraction list $z^{(i)} = (f(x_1^{(i)}), f(x_2^{(i)}), \dots, f(x_n^{(i)}))$.

$$\sum_{i=1}^m L(y^{(i)}, z^{(i)}) \quad (6)$$

When a document list needs to be sorted, given a new document, we can build a feature vector $x^{(i)}$ and adopt sort parameter to give a score to this document. Finally, we sort the candidate documents from high to low based on their scores.

(2) Permutation probability calculation. We denote the arrangement as $\pi = \langle \pi(1), \pi(2), \dots, \pi(n) \rangle$, where $\pi(j)$ represents the object at the j -th position in the arrangement.

Suppose there is a sorting function that can score each object. Use s to represent the score set $s = (s_1, s_2, \dots, s_n)$, we believe that there is uncertainty when using a sort function to predict a sorted list. In other words, any permutation and combination is possible, but the probability of different permutations and combinations is different for a given sorting function. It has some desired properties in the process of representing the likelihood of permutations and combinations. At this time, given the score list s , the probability of permutation and combination π is:

$$P_s(\pi) = \prod_{j=1}^n \frac{\phi(s_{\pi(j)})}{\sum_{k=j}^n \phi(s_{\pi(k)})} \quad (7)$$

$s_{\pi(j)}$ is the value of the object at the j -th position of the permutation combination π .

(3) Top k probability. Its formula is shown as:

$$\wp_k(j_1, j_2, \dots, j_k) = \{ \pi \in \Omega_n \mid \pi(t) = j_t, \forall t = 1, 2, \dots, k \} \quad (8)$$

It can be seen that the number of elements here is much less than the number of Ω_n elements. The constraints of these permutations and combinations are: the object list (j_1, j_2, \dots, j_k) must be at the top k position of the permutation.

3.3. Teaching Reproduction Technology of Dialogue Robot

(1) Robot teaching reproduction based on gesture. The basic principle of the Kalman is to use the state equation of the linear system, the input observation data of the current system and the last-minute estimated system state, and finally obtain the optimal system state estimation function through iteration. The iterative model of state estimation of standard Kalman filter includes system state equation and observation equation:

$$\begin{aligned} x_k &= F_k x_{k-1} + B_k u_{k-1} + w_{k-1} \\ z_k &= H_k x_k + v_k \end{aligned} \quad (9)$$

Each iteration of the Kalman filter is divided into two steps: the prediction process and the update process. The observation data and prediction are fused by the product of independent Gaussian distributions. The idea of maximum likelihood is used to suppress noise, thereby obtaining the optimal estimation of the system state. The iterative process is as follows:

Forecasting process:

1) Forecast system status: $\hat{x}_k = F_k \hat{x}_{k-1} + B_k u_{k-1}$

2) Prediction system error covariance matrix: $\tilde{P}_k = F_k \tilde{P}_{k-1} F_k^T + Q_{k-1}$

Update process:

3) Update the Kalman gain matrix: $K_k = \tilde{P}_k H_k^T \left[H_k \tilde{P}_k H_k^T + R_k \right]^{-1}$

4) Estimate system covariance: $\hat{P}_k = [I - K_k H_k] \tilde{P}_k$

5) Estimate system status: $\hat{x}_k = \hat{x}_k + K_k (z_k - H_k \hat{x}_k)$

(2) Speech-based robot teaching and reproduction. Assuming that the acquired position data of the gesture is represented as $X = (x, y, z)^T$, the model of the adaptive double-exponential smoothing filter can be expressed as:

$$\begin{aligned} b_n &= \beta (\hat{X}_n - \hat{X}_{n-1}) + (1 - \beta) b_{n-1} \\ \hat{X}_n &= \alpha X_n + (1 - \alpha) (\hat{X}_{n-1} + b_{n-1}) \end{aligned} \quad (10)$$

In the formula, X_n represents the position measurement value of the gesture at time n , \hat{X}_n is the gesture position value output by the filter at the current moment, \hat{X}_{n-1} is the gesture position value output by the filter at the previous moment, α and β are adaptive weight factors, The value range is $\alpha, \beta \in (0, 1)$, b_n represents the trend of input data at the current moment, and b_{n-1} represents the trend of the previous moment. From the two equations of the filter model, it can be seen that the trend b_n is calculated by exponential filtering of the difference between the first two outputs of the filter, and then

the current trend b_{n-1} and the previous output \hat{X}_{n-1} of the filter are used to calculate the output of the filter [32]. The trend b_n reduces the delay of the filter. The weighting factor β that weights the input data is used to calculate the trend b_n . Therefore, the weighting factor β controls the sensitivity of the trend b_n when the input data changes.

4. Experiment

4.1. Data Source

In order to make the results objective, a mature annotated corpus, people's daily, was selected as the training set. The corpus was developed by the Institute of linguistics of Peking University and provided to the researchers. Because it is a simulated navigational English dialogue oriented to the professional field, 150 sentences were collected as the test set for the evaluation experiment.

This experiment uses three different ways to simulate navigational English conversation: conditional random field, conditional random field and domain dictionary, conditional random field, domain dictionary and ambiguity elimination. We use the Python extension package provided by Harbin University of technology. The language model of the extension package includes: Modules such as unknown word recognition, navigation English, etc. In this experiment, only the Chinese word segmentation module is used. First, the module is trained through the corpus of people's daily to establish a language model, and then tested in different ways. The format of the domain dictionary used in the experiment is shown in Table 1.

Table 1. Example of domain dictionary

Terms	Candidate part of speech	Terms	Candidate part of speech
Coffee table	nz	Refrigerator	nz
Wash basin	n	Fruit knife	nz
Color TV	ns	Wooden dining table	nz

4.2. Description of Experimental Scene

The real scene of teaching reproduction experiment based on simulated navigational English dialogue and human-computer interaction is as follows: the operator stands at the table, wears Hololens glasses, and observes the movement of the robot. Before the teaching, the robot has completed three-dimensional registration and overlapped on the real robot. Kinect placed on the table is used to capture the operator's voice commands,

and the standard camera beside the tool board is used to observe the movement of the robot's end axis. The hardware involved in the teaching experiment is shown in Table 2:

Table 2. Hardware equipment of analog navigational English conversation

Serial number	Name	Number	Features
1	Kinect sensor	2	Gesture tracking and speech recognition
2	Camera	1	Observe the end axis movement of the robot
3	Dialogue robot	1	Simulation dialogue experiment tool
4	HoloLens glasses	1	Augmented reality display
5	Experiment Tool Board	1	Simulation dialogue experiment tool
6	Computer	2	navigational English dialogue reproduction program
7	Wireless Router	1	HoloLens communicates with a PC

4.3. Experimental Verification

(1) **Speech recognition.** During the experiment, the tip of the left index finger of the operator moves along the center line of the S-shaped groove on the tool board, the Kinect of the hand eye position tracks the gesture posture of the operator, which is converted into the joint angle required for the movement of the dialogue robot, and sent to HoloLens through the wireless network. The operator wears HoloLens glasses to observe the situation of the dialogue robot moving with the fingertip. When the operator speaks the voice command "follow the fingertips," an navigational English conversation is started and the action is realized. When the action is completed, the operator only needs to say the voice command "dialogue completed" to stop the dialogue and movement, and then the dialogue robot moves to the initial position. This experiment mainly verifies the performance of speech recognition and action realization.

(2) **Action realization.** During the experiment, the operator simulated the navigational English conversation of the robot through two natural human-computer interaction modes, gesture and voice. The operator first uses gestures to guide the end of the dialog robot to the target hole. Since the diameter of the robot end axis is 16mm, the diameter of the target hole is 20mm, and the end axis of the robot is 2mm away from the center of the target hole, the jack cannot be completed. The voice instructions are fine-tuned and inserted into the target hole.

4.4. Evaluation Criteria

According to the characteristics of the word segmentation system, the performance of each system is mainly measured by using three indicators of accuracy (P), recall (R), and

F-measure as evaluation indicators. The calculation formulas are as shown in the formulas:

$$\begin{aligned}
 Accuracy &= \frac{\text{Number of correct results}}{\text{Total number of results}} \\
 Recall &= \frac{\text{Correct results in word output}}{\text{Correct results in the test set}} \\
 F &= \frac{2 \times P \times R}{P + R}
 \end{aligned} \tag{11}$$

5. Results and Discussions

5.1. Analysis of Navigational English Word Segmentation and Speech Based on CRF

Word segmentation based on different combinations of CRF models yields different results, and calculations of various performance indicators are performed according to evaluation criteria. This article summarizes and compares the data obtained from the evaluation of various models. The results are shown in Table 3 and the corresponding histogram 1. It is found through the chart that combined with navigational English literature knowledge, a dictionary of navigational English conversation has been added, and the word segmentation after conditional random field combined with the field dictionary has significantly improved accuracy, recall, and F value compared with the word segmentation using conditional random field. Combined with the ambiguity resolution method, 6.0%, 6.2%, and 5.3%, the corresponding evaluation index after the experiment is the highest among the three methods, which are 9.4%, 8.4%, and 10.2% higher than using the CRF model alone.

Table 3. Comparison of segmentation results of different combinations

Word segmentation model	Accuracy	Recall	F value
CRF model	0.804	0.799	0.802
CRF Model + Domain Dictionary	0.864	0.861	0.855
CRF model + domain dictionary + ambiguity resolution	0.898	0.883	0.904

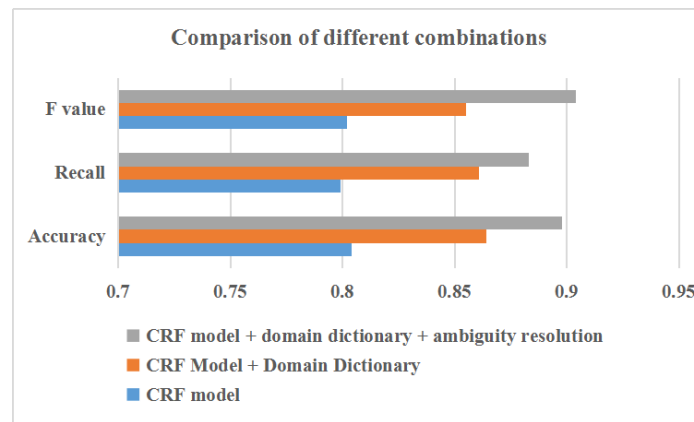


Fig 1. Comparison results of different combinations

Through the above experimental steps, a semantic analysis model for dialog robots is initially established. The analytical model is tested through the test set. Three major indicators are evaluated based on the test. The percentages of the analytical model are calculated as correct rate: 76.85%; Recall: 80.36%; F-value: 88.46%. Experiments show that the semantic parsing model can parse navigational English natural language instructions after word segmentation, and nearly 77.95% of sentences can be parsed and output correctly, which means that the semantic parsing model can convert navigational English into an intermediate representation. Finally, internal instructions that the robot can recognize are generated.

It is known from experiments that the CRF model can significantly improve the accuracy of word segmentation by combining with the domain dictionary and ambiguity resolution. The main reason is that the domain dictionary can identify proper nouns and unregistered words in the domain, and can correct the word segmentation results in the domain, and the use of ambiguity removal technology can further solve the problem of result ambiguity on this basis.

5.2. Analysis of Teaching Results of Dialogue Robots

In the simulated navigational English dialogue teaching and reproduction, the teaching accuracy is completely dependent on the accuracy of the gesture pose data acquisition, so the corresponding filtering algorithm is used to promote the precision. Gesture and posture data is filtered by Kalman filter. The filtering effect is shown in Figure 2. The blue dotted line in the figure is the gesture posture information (observed value) directly obtained by the gesture tracking system. The green solid line is after Kalman filtering. It can be seen from the figure that the attitude data directly acquired by the sensor has a lot of random noise, which fluctuates greatly. After Kalman filtering, the noise is effectively filtered, making the data smoother and more stable, and improving the stability of the teaching system. Because the tool board is placed horizontally and both are within two degrees.

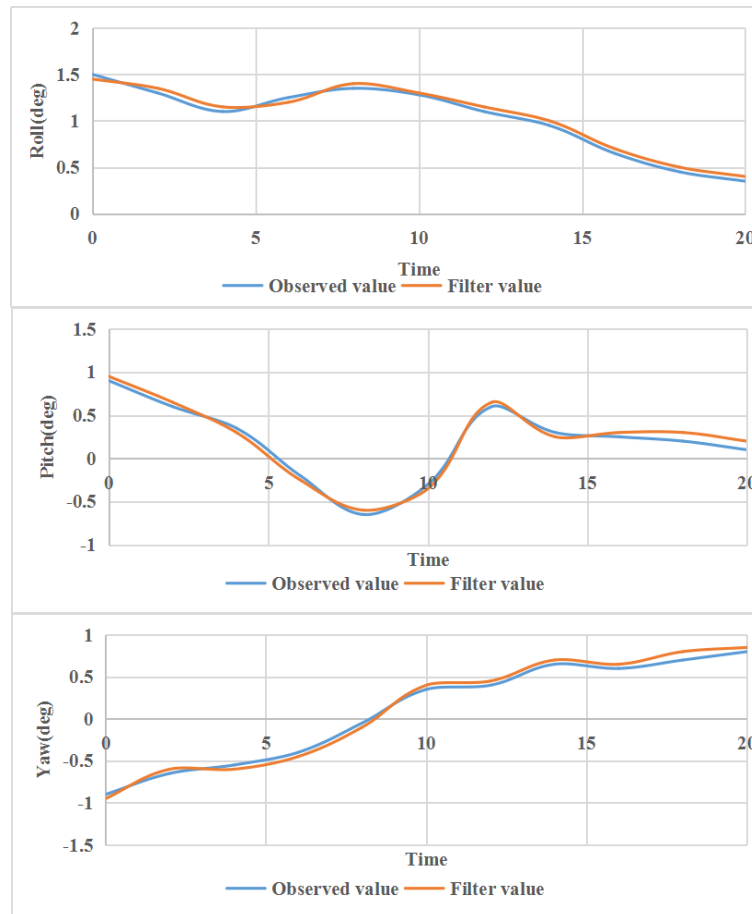


Fig 2. Filter comparison of gesture and pose data

The comparison is shown in Figure 3. The blue dotted line in the figure is the gesture position information (observation value) obtained directly by the gesture tracking system, and the green solid line is the estimated value (filter value) after the adaptive double exponential smoothing filter. Because the robot end moves perpendicular to the tool plate placed horizontally, only the X and Y direction motion data are considered. It can obtain a stable and smooth trajectory.

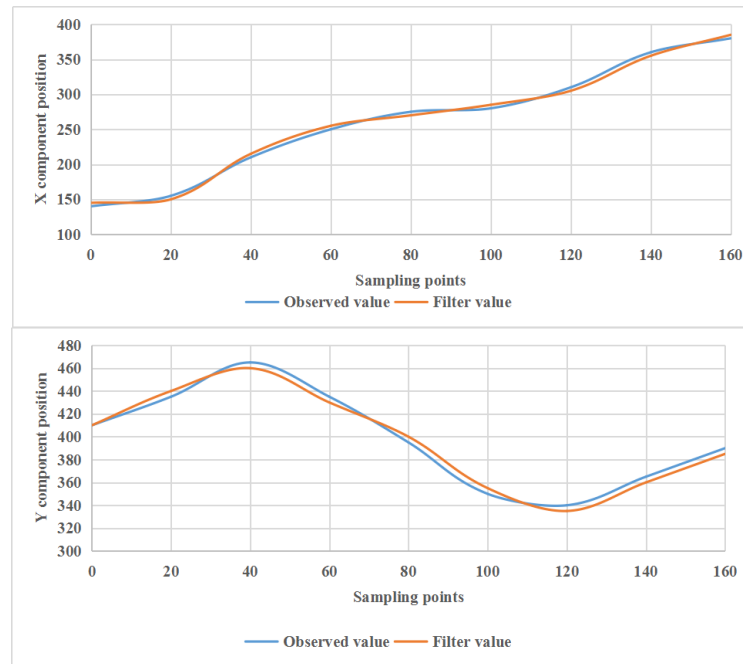


Fig 3. Filtering comparison of gesture position data

In the navigational English conversation and action realization reproduction experiment, first guide the robot end axis to the top of the target hole, because the accuracy of gesture teaching is 5mm, and the robot end axis can not complete the jack as long as it is 2mm away from the hole center, so gesture teaching can not meet the requirements of the jack, and can not carry out the actual jack, so only analyze the two-dimensional error, as shown in Table 4. Then use the voice command to fine tune the robot end axis after hand potential teaching to complete the jack. Table 5 shows the experimental error data after speech fine teaching, including two-dimensional error and three-dimensional error. The change of two-dimensional error after gesture teaching and speech fine teaching is shown in Fig. 4.

Table 4. Two dimensional error of gesture Teaching

Test number	1	2	3	4	5	6
X-axis error	2.6513	-1.0006	0.9913	3.6438	2.9956	1.1124
Y-axis error	-1.5688	3.2511	1.2293	3.5551	-1.8864	1.6643
Two-dimensional error	2.6781	3.4437	1.6606	5.0007	3.8511	1.8888

As shown in Figure 4, the error range of gesture teaching is relatively large. and those with small error can be directly inserted into the hole without the need of voice teaching for adjustment, and those with large error can be successfully inserted into the target hole within 2mm after voice fine-tuning. The three-dimensional error of speech fine tuning is between 1.6798mm and 2.9968mm. The errors of gesture teaching and speech teaching are in millimeter level. There are two factors that affect the robot's end trajectory. One is that there is a small error in the modeling process of dialogue robot.

The simulation results show that the error is 0.1684mm; The second is that the accuracy of the registration of the dialogue robot on the real robot is not high, which is limited by the 3D registration technology based on vision, and its registration accuracy is within 1.25 mm, thus affecting the accuracy of the simulation dialogue of the robot.

Table 5. Experimental error after fine speech teaching

Test number	1	2	3	4	5	6
X-axis error	0.3316	-1.0058	0.6687	0.2175	0.6799	0.7361
Y-axis error	-1.4431	1.2221	.11633	0.1999	1.3467	1.3491
Z-axis error	2.2655	1.4389	3.2451	2.1137	2.8134	2.6643
Two-dimensional error	1.9934	1.2964	1.2473	0.2946	1.0673	1.3450
Three-dimensional error	2.8115	1.8873	3.0521	2.0455	2.9431	2.9910

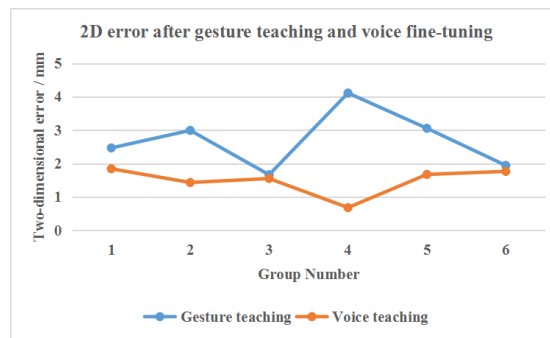


Fig 4. Two dimensional error after gesture teaching and speech fine tuning

5.3. Result Analysis of Answer Sorting of Dialogue Robot

(1) Sorting accuracy analysis. In order to compare the advantages and disadvantages of different sorting methods, three sorting methods were used in the experiment: (1) sorting purely according to the correlation degree of BM25, without considering any other sorting factors, this method is called rawrank; (2) Listnet sorting learning method we use; (3) Ranknet is the representative of pairwise sorting learning method and the most effective method in pairwise, so ranknet method is selected as the contrast object in this paper. We divide the training data set into five sub sets on average and carry out 5 fold cross validation. In each round of test, 3 data are used as training data, 1 data is used for verification, and the remaining 1 data is used for test. For Ranknet and Listnet, verifying which data is used to determine the number of iterations. Figure 5 and Table 6 show the accuracy.

Table 6. Ranking accuracy (map metrics)

Algorithm	ListNet	RawRank	RankNet
Result	0.394	0.267	0.202

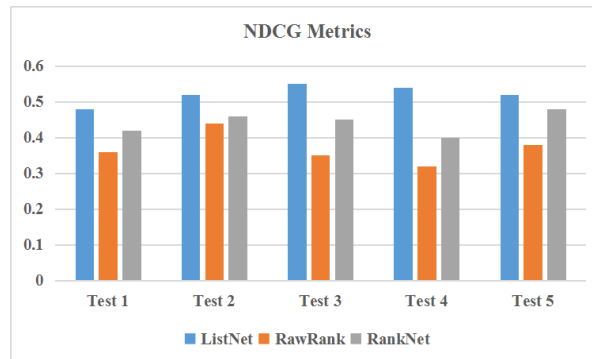


Fig 5. Sorting accuracy (NDCG metrics)

(2) Performance analysis of dialogue robot prototype system. The NLP part can solve 430 questions. In the FAQ section, 312 input problems can be successfully found, But in general, there were still no answers to the 219 input questions, so the recall rate was 78.1%. As can be seen from the second column of Table 7, the FAQ component has up to 78.13%.

Table 7. Evaluation results of FAQ and NLP components

	Return times	Precision
FAQ	312	78.13%
No-wiki-NLP	399	72.43%
wiki-NLP	31	60.98%
ALL	802	73.62%

Analyzing the experimental data, we found that in the first two months of the system's launch, the regular matcher component solved 1.1% of the problems raised by users. Later, in the first four months of the system's launch, we found that 2.01% of the problems had been resolved in the regular matcher. This shows that the regular matcher component works as we wish.

6. Conclusions

In this paper, a test data set is used to test the established navigational English dialogue instruction analysis model. The results of this work proves to be efficient, and the CRF + domain dictionary + ambiguity analysis method has the highest word segmentation results. The calculated percentages of the analytic model are correct rate: 76.85%; recall rate: 80.36%; F-value: 88.46%. Nearly 77.95% of sentences can be parsed and output correctly.

In this paper, a robotic teaching and reproduction method based on simulated navigational English conversation and human-computer interaction is implemented, and robot motion realization experiments and speech recognition are used to test methods proposed in this paper. The gesture teaching with large error can be smoothly inserted into the target hole after the two-dimensional error can be within 2mm after fine-tuning

the voice. The three-dimensional error after fine-tuned speech is between 1.6798mm and 2.9968mm.

This paper constructs a simulated navigational English dialogue robot system that can use knowledge to understand problems on a large scale. It consists of a three-tier system. The FAQ component has up to 79.2%.

References

1. Šoić, R., Vuković, M., Ježić, G. (2021). Spoken Notifications in Smart Environments Using Croatian Language. *Computer Science and Information Systems*, Vol. 18, No. 1, 231–250.
2. Rakesh kumar S, Muthramalingam S, F. Al-Turjman, (2021). "Multimodal News Feed Evaluation System with Deep Reinforcement Learning Approaches", *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 20, no. 1, pp. 8 – 20.
3. Yin, S., & Yuschenko, A. S. (2019). Dialogue system of controlling robot based on the theory of finite-state automata. *Mekhatronika Avtomatizatsiya Upravlenie*, 20(11), 686-695.
4. Drager, K., Grama, J., Gonzalez, L., & Copeland, C. (2016). Voice onset time and closure duration of word-initial voiceless plosives in pidgin conversation. *Acoustical Society of America Journal*, 140(4), 3052-3052.
5. Jain, D., Jakhalekar, I. R., & Deshmukh, S. S. (2017). Navigational strategies and their neural correlates. *Journal of the Indian Institute of Science*, 97(4), 511-525.
6. Sara, D., Steve, S., Ha Phuong, V., Lan-Xi, D., & Ana, J. (2017). Rendering website traffic data into interactive taste graph visualizations. *Big Data and Information Analytics*, 2(2), 107-118.
7. C. Warner, & H.-I. Chen. (2017). Designing talk in social networks: what facebook teaches about conversation. *Language Learning & Technology*, 21(2), 121-138.
8. Edgar Lucero, & Jeesica Scalante-Morales. (2018). navigational English language teacher educator interactional styles: heterogeneity and homogeneity in the elite classroom. *HOW*, 25(1), 11-30.
9. James F. navigational English, & Ted Underwood. (2016). Shifting scales: between literature and social science. *Modern Language Quarterly*, 77(3), 277-295.
10. Azmuddin, R. A., Nor, N., & Hamat, A. (2017). Metacognitive online reading and navigational strategies by science and technology university students. *Gema Online Journal of Language Studies*, 17(3), 18-36.
11. Christina Kuo, & Gary Weismer. (2016). Vowel reduction across tasks for male speakers of american navigational English. *Journal of the Acoustical Society of America*, 140(1), 369-383.
12. Lee, Heeju, Su, Danjie, & Tao, Hongyin. (2018). A crosslinguistic study of some extended uses of what-based interrogative expressions in chinese, navigational English, and korean. *Chinese Language & Discourse*, 8(2), 137-173.
13. Pearson, M. (2016). Staging power in tudor and stuart navigational English history plays: history, political thought, and the redefinition of sovereignty by kirstin m. s. bezio. *Theatre Journal*, 68(4), 698-699.
14. Brenda Henry-Offor. (2017). Early modern women in conversation by katherine r. larson. *Esc navigational English Studies in Canada*, 43(1), 110-112.
15. Cicienia, J., Avasarala, S. K., & Gildea, T. R. (2020). Navigational bronchoscopy: a guide through history, current use, and developing technology. *Journal of Thoracic Disease*, 12(6), 3263-3271.
16. Lu Wanling. (2016). The examination of acoustic feature of navigational English obstruent coda by mandarin and cantonese speakers. *Acoustical Society of America Journal*, 140(4), 3336-3336.

17. Shu, W., Dong, Z. Y., F Luo, Ke, M., & Zhang, Y. (2018). Stochastic collaborative planning of electric vehicle charging stations and power distribution system. *IEEE Transactions on Industrial Informatics*, 14(99), 321-331.
18. Paul Gilmore. (2017). Charles Brockden Brown's romance and the limits of science and history. *ELH navigational English literary history*, 84(1), 117-142.
19. Oh, Y. H., Kim, J., & Ju, D. Y. (2019). Investigating the preferences of older adults concerning the design elements of a companion robot: analysis on type, weight and material of companion robot. *Interaction Studies*, 20(3), 426-454.
20. C. Ton, Z. Kan, & S. S. Mehta. (2017). Obstacle avoidance control of a human-in-the-loop mobile robot system using harmonic potential fields. *Robotica*, 36(4), 1-21.
21. Lisheng Kuang, Yunjiang Lou, & Shuang Song. (2017). Design and fabrication of a novel force sensor for robot grippers. *IEEE Sensors Journal*, 18(4), 1.
22. Cottrill, C. ., Pereira, F. C., Fang, Z., Dias, I. F., Lim, H. B., & Ben-Akiva, M. E., et al. (2018). The future mobility survey: experiences in developing a smartphone-based travel survey in singapore. *Transportation Research Record*, 2354(1), 59-67.
23. Cifuentes, C. A., Rodriguez, C., Frizeraneto, A., Bastosfilho, T. F., & Carelli, R. (2017). Multimodal human-robot interaction for walker-assisted gait. *IEEE Systems Journal*, 10(3), 933-943.
24. H Su, Lallo, A. D., Murphy, R. R., Taylor, R. H., & Krieger, A. (2021). Physical human-robot interaction for clinical care in infectious environments. *Nature Machine Intelligence*, 3(3), 184-186.
25. Wang, Q., Yang, X., Huang, Z., Ma, S., Li, Q., & Gao, D. W., et al. (2018). A novel design framework for smart operating robot in power system. *IEEE/CAA Journal of Automatica Sinica*, 5(2), 531-538.
26. T. Li, G. Li, Y. Liang, T. Cheng, & Z. Huang. (2016). Review of materials and structures in soft robotics. *Chinese Journal of Theoretical & Applied Mechanics*, 48(4), 756-766.
27. K Wang, Yang, C., & Wang, T. (2020). A smart robot training data acquisition and learning process recording system based on blockchain. *Open Access Library Journal*, 07(9), 1-5.
28. Amir Firouzeh, & Jamie Paik. (2017). An under-actuated origami gripper with adjustable stiffness joints for multiple grasp modes. *Smart Material Structures*, 26(5), 1-17.
29. Zheng Chen. (2017). A review on robotic fish enabled by ionic polymer-metal composite artificial muscles. *Robotics & Biomimetics*, 4(1), 24.
30. Yong-feng Dong, Hong-mei Xia, & Yan-cong Zhou. (2016). Disordered and multiple destinations path planning methods for mobile robot in dynamic environment. *Journal of Electrical and Computer Engineering*, 2016(6), 1-10.
31. Luo, D., Chen, D., Wang, J., Zhu, G., & Xu, W. (2020). The smart robot crafting approach to computing materials. *Construction Robotics*, 4(2), 1-11.
32. Li, H., Han, D. (2021). Multimodal Encoders and Decoders with Gate Attention for Visual Question Answering. *Computer Science and Information Systems*, Vol. 18, No. 3, 1023-1040.

Wei Sun, was born in Chuzhou, Anhui, P.R. China, in 1982. She received the master's degree from South Central University for Nationalities, P.R. China. Now, she studies in Lyceum of the Philippines University for doctor's degree. Her research interests include linguistics, application of computational intelligence into translation. E-mail: sunwei@axhu.edu.cn

Received: August 20, 2021; Accepted: May 10, 2022.

Interactive and Innovative Technologies for Smart Education

Babatomiwa Omonayajo¹, Fadi Al-Turjman², and Nadire Cavus³

¹ Department of Computer Information Systems, Near East University
99138 Nicosia, Cyprus
20212976@std.neu.edu.tr

² Artificial Intelligence Engineering Department, Research Center for AI and IoT, AI and Robotics Institute, Near East University
Nicosia, Mersin 10, Turkey
fadi.alturjman@neu.edu.tr

³ Department of Computer Information Systems, Computer Information Systems Research and Technology Centre, Near East University, 99138 Nicosia, Cyprus, Mersin 10, Turkey
nadire.cavus@neu.edu.tr

Abstract. New concepts and ideas have emerged in the process of obtaining and disseminating cognitive, ethical, and public knowledge. In the current state of education, learners, tutors, and the knowledge being transferred are all present, and smart education has made the process of acquiring knowledge more flexible. This concept is accomplished through the use of smart devices and technologies that are interconnected to access digital resources. Smart education refers to a new way of learning that has gotten a lot of attention, notably during the 2020 Covid-19 Pandemic. This article examines the technologies that have aided smart education in achieving its educational goals. With smart technological solutions, modern technologies are enhancing the teaching - learning process in today's education. It is with great hope that the use of modern technologies in smart education will improve educational quality while also making teaching and learning more convenient.

Keywords: smart education, technologies, innovation, interaction, e-learning.

1. Introduction

For the successful structuring of students' educational processes, the system of education necessitates the employment of new teaching technology. In today's digital era, smart education is a form of education which is becoming increasingly generally recognized and enthusiastically embraced adopted by younger generations. This highlights the fact that education based on current technology allows for more efficient and convenient knowledge transfer to students [22]. Smart education can be referred to the process of learning adapted by the new age of digital orientation, providing a more interactive, collaborative and visual method for the purpose of increasing learners engagements and allows tutors to understand student skills and learning preferences. Smart education is a means of learning that has been adopted by the modern era of digital orientation to provide a more dynamic, interactive, and visible way for enhancing students' engagement and allowing instructors to identify student talents and student motivation. Nonetheless, the growing diversity

of data collection is posing a substantial challenge to information systems (IS) and instructional development research. These issues inspire the creation of new ideas, which leads to new technology advancements that allow for the advancement of how improved educational activities are offered [30].

Considering rapid technology breakthroughs, anything can be implemented, networked, and integrated with intelligent and innovative design, and education is no exception. In terms of innovative methods of education, learning, and instructional practices, comprehensive and innovative technology have offered unique chances for academic institutions [29]. Regardless of the fact that key technical innovations have been overlooked in the past, events such as the 2020 pandemic demand quick adaptation of the usage of smart devices and technologies to promote smart learning and education.

Innovative and interactive technology is employed for many core skills, allowing learners to experience smart education via the use of various interactive technologies [18]. Some innovative technologies, such as augmented reality (AR), virtual reality (VR), internet of things (IoT), metaverse, immersive 3D etc., have rattled the education system with their amazing abilities. but concerns are still raised about human alteration or laziness in the adaptation of these technologies, which is related to the fear that prompted the late adaptive response of these technologies in the educational systems.

2. New Technological Approaches For Smart Education

Education is indeed an area where one might anticipate innovations to take hold as soon as they become accessible. But in actuality, other sectors are faster to adjust to new technological ideas than the education industry [9]. However, since the discovery of mobile devices and the internet, technological advancement has gone decades beyond the conventional instructional system. AR, VR, IoT, metaverse, Immersive 3D, and other emerging approaches for smart education are further discussed.

2.1. Augmented/Mixed/Virtual Reality

Although AR and VR are similar, they are two separate technologies. AR is an immersive experience in which computer-generated facts and features are connected to the actual world; examples of AR include Global Positioning System (GPS) and cameras [21]. In comparison to AR, VR tends to take place in a controlled environment in which users may engage with it and modify computer-generated elements in a digital world using sensory devices. Examples of VR include contemporary game consoles such as the PlayStation and the Oculus Rift [26]. It can be seen in Figure 1 that AR and VR are combined in mixed reality (MR).

According to [6], augmented reality and virtual reality offer significant possibilities for supporting students in improving their abilities and expertise. Furthermore, integrating AR/VR into education may provide engaging and effective education experiences.

In recent years, technological advancements have simplified the usage of VR and AR, making it more accessible. According to a [19] report, many people already own VR/AR compatible smart gadgets like the ARCore and Oculus Quest 2 (VR headset). Thus, there's no need for a huge and pricey headgear to carry around. In addition, one of the most recent practical applications of this technology is for showing instructions or

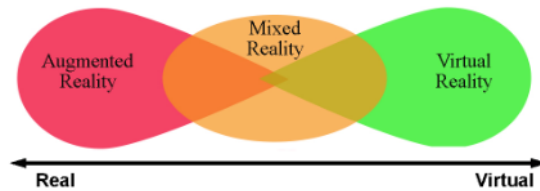


Fig. 1. A representation of AR, MR, VR classification [6]

information in big indoor spaces such as shopping malls and airports. Furthermore, it has been employed in education and other fields like healthcare, remote support, shopping convenience, automotive manufacturing, and so on

2.2. Internet of Things

The internet of things (IoT) is a technological evolution in which items connected to sensing devices, controllers, and processors connect with one another to accomplish a specific goal [16]. Using the internet of things concept in any educational setting would improve the quality of the educational system since learners develop faster and teachers will do their jobs better. As a result, a smart educational system is developed. It includes various communication technologies to stimulate the learning experience and adapt to the demands of various learners [1].

2.3. Metaverse

Metaverse is a three-dimensional virtual reality world in which any user with access to a terminal, from anywhere on the planet, may engage in everything from trading to entertainment. Much of human day-to-day communication has been transferred to the virtual world as a result of the metaverse's advent, which has had a significant impact on human societies and culture in the physical world [12]. It is now being marketed as the technology of the future since it has been incorporated into a variety of industries. The potential of the metaverse as a new educational environment is considered to be a place to build interpersonal contact, greater flexibility to create and distribute, and the introduction of better perceptions and rising participation through virtual machines. As seen in Fig 2, a classroom session conducted in the metaverse.

The current evolution of the metaverse in the educational system enables students to conduct experiments such as examining the anatomy lab on the interior of the body seen in fig 3. As face-to-face interaction becomes more difficult as a result of the Covid-19 pandemic, events that were previously considered to be only feasible outdoors are being transformed to virtual worlds and are fast growing into industries like as medical services, education, and entertainment [17].

2.4. Big Data Technology

Mostly associated with smart education, big data technology has grown at an enormous speed, and incorporates information retrieved from students' interactions with technology



Fig. 2. Classroom map in Zepeto [17]



Fig. 3. Metaverse avatars watching a surgical scene in the smart operating room [15]

as well as their personal and academic profiles. One of the most perplexing features of big data in academia is the lack of standardization [11], as it is in other industries, is how to draw meaning from the data gathered. Researchers and practitioners are steadily publishing and disclosing additional evidence of the advantages of big data technologies.

In smart education, the phrase “big data” is not yet universally defined. All behavioural data obtained from human beings’ everyday educational actions, which contains hierarchy, sequence, and contextual properties, is referred to as big data technology in smart education. Furthermore, it relates to data collected from student activations, which is mostly generated through student monitoring systems, interactive learning systems, and curriculum management systems, among other sources [28].

2.5. Blockchain Technology

Due to its distinguishing properties such as decentralization, trustworthiness, security, and integrity of data, blockchain technology has received considerable attention lately. Many industries are investigating the prospect of incorporating blockchain technology into their respective fields in order to fully use blockchain’s potential. Despite its fast expansion, there is little information available about blockchain’s current cutting-edge technology in the education system [25]. Beyond credential administration and success evaluation; blockchain technology can be used in education in a variety of innovative ways. Blockchain technology offers a lot of possibilities for both teachers and students in terms of formative assessments, teaching session design and execution, and tracking the progress of the entire learning experience [10].

2.6. Cloud Computing

In the education sector, cloud technology has gotten a huge interest as a method to offer more stable and secure quality education. [14] introduced a cloud-based smart system of education for e-learning digital experiences with the purpose of distributing and sharing advanced forms of educational material such as text, pictures, images, videos, and three-dimensional (3D) objects, among other things. Digital learning has always been a part of smart education, dating back to instructive TV programmes. Electronic-learning, mobile-learning, and now smart-learning have all grown from it. With the use of cutting-edge technology such as IoT and cloud computing, academies are getting smarter. It improves the typical classroom setting to help students learn more effectively [5].

2.7. Artificial intelligence Technologies

Artificial intelligence plays a variety of functions in education today, making it smarter. Artificial intelligence advancements have opened a new era of computer-assisted education. Computer systems with cognitive capacities can not only operate as intelligent teachers, resources, or students, but they can also help with strategic planning in the context of education. The combination of artificial intelligence with education opens up new avenues for considerably improving educational and learning quality. Intelligent systems assist instructors in testing, data collection, academic performance improvement, and the development of new strategies. Asynchronous learning and smart educators assist students enhance their academic performance [13]. Furthermore, the implementation of artificial intelligence with education is a revolution not just in smart education but also in human understanding, cognitive, and civilizations.

3. Application Of Artificial Intelligence Techniques In Smart Education

As the base technology used in the evolution of smart education, Artificial Intelligence (AI) techniques are important mechanism because it creates and mimic people's decision-making processes. Additional explanation of some artificial intelligence model such as Fuzzy Logic, Decision Trees, Neural Networks, Bayesian Networks, Genetic Algorithms, and Hidden Markov Models are used in smart educational reforms etc., to develop a suitable AI integration in education [7].

Fuzzy Logic Fuzzy logic is an artificial intelligence strategy for processors to use the “degrees of fact” instead of the traditional “true or false” (1s or 0s) Binary logic that underlies current computers.

Decision Tree For classification and regression in AI machine learning, Decision Tree (DT) is a nonlinear supervised training technique for developing algorithms for defining target classes by learning fundamental decision rules from data characteristics.

Neural Network The neural network technique is a collection of algorithms that attempt to find hidden patterns in a data stack by employing approaches that mirror how the brain operates. Neural networks in this sense are made up of a collection of naturally or artificially neurons.

Bayesian Network The Bayesian network (BN) is a completely comprehensive model for describing data about an uncertain domain, with each node representing a random price and each chord showing the transition options for those unknown parameters.

Genetic Algorithm A genetic algorithm (GA) is an artificial intelligence (AI) approach for addressing both limited and uncontrolled classification problem that uses a natural selection process similar to evolutionary biology.

Hidden Markov Model The Hidden Markov model (HMM) is a statistical Intelligence technique commonly used to represent biological events. In reality, a pattern is represented as the outcome of a continuous random process that progresses through a set of “secret” levels from the viewer.

Some performance indicators (e.g., specificity, accuracy, area under the curve, sensitivity, and so on) are used in measuring the approaches employed in the implementation of machine learning and artificial intelligence in the educational system and other domains [8].

Accuracy Accuracy matrices specify the amount of accurate classification model across all types of assumptions made. *Accuracy* is an excellent measure to utilize whenever the target factor categories in the data are nearly equal, but it is not the ideal measure as well when the point factor categories in the data are a bulk of one category. $Accuracy = (TP + N) / (TP + FP + FN + TN)$

Sensitivity For a given model and prediction issue, sensitivity analysis provides a method for assessing the link between model performance and dataset size. Sensitivity is not so much about accurately obtaining instances as it is about obtaining all cases that match. So there is 100 per cent *sensitivity* if there is always matching data in a dataset. $Sensitivity = (TP) / (TP+FN)$.

Specificity The fraction of true negatives properly predicted by the model is known as *specificity*. $Specificity = (TN) / (TN+FP)$. Specificity is the polar opposite of Sensitivity.

Note in the above, TP (True Positive), TN (True Negative), FP (False POSITIVE) and FN (False Negative). As further explained by [31], these indicators or metrics are used after developing a model to determine the model’s effectiveness. The measures used to evaluate the AI model are significant. The measurements are used determine how machine learning techniques’ performance is assessed and evaluated.

From the standpoint of instructional methods shown in figure 4, AI in education can serve as an intelligent tutor, tutee, learning tool/partner, or policy-making adviser [13].

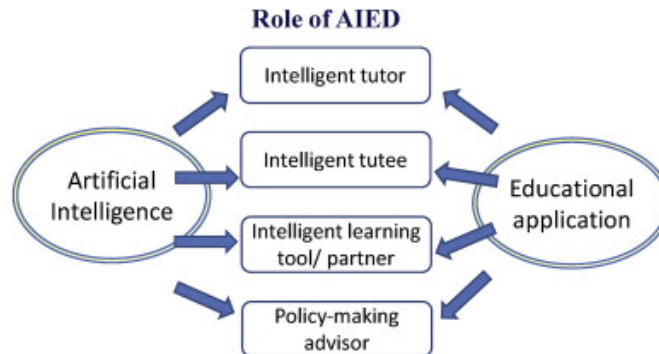


Fig. 4. Structure for AI's roles in educational systems [13]

4. Smart Education Environment

The integration of innovative technology is often limited when there are two main components, competition and skill level of technology. The smart educational environment represents a step forward in the application of innovative and interactive technologies to the traditional library, transforming it into a completely different system that can be called smart education. With the incorporation of new technologies, an environment can be considered a smart environment for smart education in which the educational approach uses technologies such as IoT as part of new innovation with which it can collaborate to enhance the quality of services by offering a customized smart educational environment [20].

In the proposed intelligence library system framework by [4], The framework, according to the authors, provides libraries with an integrated database that combines three basic components: the electronic book shelf (EBS), which provides complete access to price levels; the virtual white space (VWS), which allows users to discuss data library resource; and the Innovation and Social Networking Database (ISNB), which allows users to share and store innovative concepts.

The benefits of a smart educational environment come from leveraging the knowledge that exists in the library and the way advanced wireless technologies are used and applied in the library. Students can directly benefit from participating in research and development using a smart educational environment. Smart education environment will provide an academic environment with intelligent educational services through the framework proposed by [4].

The interaction of humans and computers in the educational environment should be taken into account during the implementation of a smart education environment. Computer systems, when used appropriately, can be a useful tool for enhancing and assisting learning and teaching. They make students absorb information more easily during the learning process. Combining computer-based systems with practice programs, learning tutorials, and Internet-based learning can help achieve this. When students are enthusiastically immersed in the use of technology, they learn more quickly. Both students and teachers should be responsible for ensuring that continuous progress in learning and technology adoption is maintained. As a result, for the successful implementation of human-

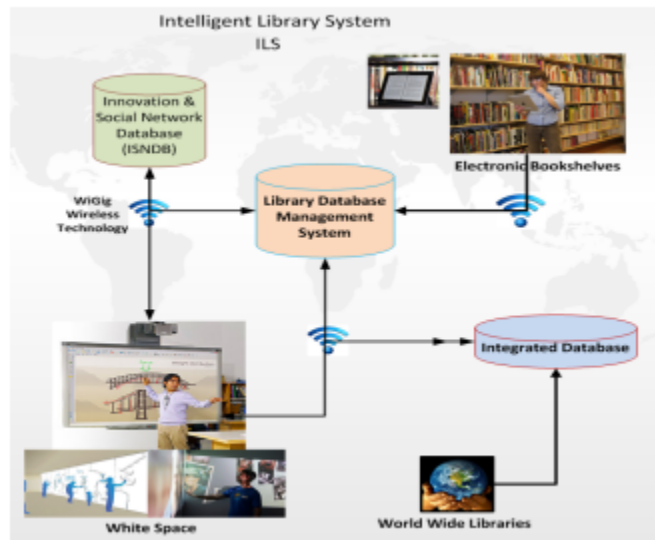


Fig. 5. Intelligence library system framework [4]

computer interaction in schools, every entity participating in the educational system must be devoted and supportive. Both students and instructors must accept and embrace technological advancements. Everyone needs to learn about information systems, and it's inescapable as the world transforms into the internet of things. New things are discerned by creativity, and new things are performed by inventing [23].

5. Smart Education Technology Features And Challenges

Undoubtedly, smart education technology has its features and challenges. In Table 1, the technologies and their challenges and features are compared in terms of complexity, increase, reduction, moderate, and support.

Innovative technologies in smart education can be seen in Table 1 combine with technological features and challenges like [24]:

5.1. Agility

The flexibility of smart education technologies to operate rapidly and effortlessly varies between the complexity of traditional education, the growing capabilities of artificial intelligence, augmented reality, virtual reality, blockchain and the best of 5G performance, learning management systems and applications.

5.2. Adaptability

Traditional education is complicated in its capacity to adapt to changing situations, which is modestly controlled for augmented reality, artificial intelligence, and virtual reality.

Technology/ Features & Challenges	Agility	Adaptability	Integration	Interoperability	Reuse	Reliability	Quality	Learning Expe.	Cost	Safety	Proof Of Work	Op. Efficiency
Traditional Education	C	C	C	C	C	C	C	C	C	C	C	C
AI	I	M	I	M	M	M	I	I	R	-	N	R
AR and VR	I	M	I	M	M	M	I	I	R	-	-	-
Big Data	-	I	I	I	I	I	I	I	R	R	-	-
Blockchain	I	I	I	I	I	I	I	I	R	I	B	B
Cloud Computing	-	I	I	-	I	I	I	I	R	I	-	I
Data Science	-	-	I	-	-	M	I	I	R	-	-	M
IoT	-	I	I	C	R	I	I	B	R	R	-	S
ML and DL	-	I	-	-	I	-	I	I	R	I	-	-
5G	B	B	B	-	-	B	-	-	R	-	-	-
LMS & App.	B	B	B	B	B	B	I	I	R	-	B	I
STEM	Experiment, Discover, Thinking, Collaboration, etc.											
<i>C-Complexity, I-Increased, R-Reduced, B-Better, M-Moderate, S-Support</i>												

Table 1. A comparison of smart education innovative technologies [24]

In terms of adapting to new developments, big data, blockchain, cloud computing technologies, the internet of things, machine learning, and deep learning are all gaining in popularity, while 5G performance, learning management systems, and apps are improving.

5.3. Integration

Traditional education's ability to integrate with existing or emerging technologies is still a slow process, while augmented reality, virtual reality, artificial intelligence, big data, cloud computing, blockchain technologies, the internet of things, and data science are all being moderately integrated. The performance of 5G networks, learning management systems, and applications are getting better as they integrate with current and future technology.

5.4. Interoperability

Interoperability is a feature of smart education technology that allows them to communicate with other systems. The capacity of conventional education and the internet of things to interact with other systems or networks outside of their respective sectors is difficult. Artificial intelligence, augmented reality, and virtual reality have limited capacity to integrate with other systems. Big data and blockchains are increasingly being used in conjunction with other platforms, while learning management systems and apps, work better with other educational systems.

5.5. Reuse

For artificial intelligence, augmented reality, and virtual reality, the reusability of smart educational technology is modest. Big data, machine learning, deep learning, cloud computing, and then blockchain are all being utilized more and more often. Learning management systems and apps are better utilised than the internet of things.

5.6. Reliability

Artificial intelligence, data science, augmented reality and virtual reality are efficient technologies. Cloud computing, the internet of things, big data and blockchain are becoming increasingly trustworthy. 5G networks, as well as learning management systems and apps, are more dependable.

5.7. Quality

The quality of smart education technology is improving generally. Every time there is a new technology, it is normal for it to be tried in every possible applicable area. This has prompted for a steady quality assurance in improvement of smart education technology.

5.8. Learning Experience

Learning experience is a procedure that allows a student to accomplish a desired academic achievement in a human-centred manner, most often through technological advances. According to table 1, the learning opportunity of smart technology in education is typically growing, with the exception of the internet of things, which has a greater learning experience.

5.9. Cost

In education and other areas, the cost of innovative and interactive technology for smart education is decreasing and becoming more affordable. Cost has always been a major factor to be considered in any sector and technological invention and innovation. If a certain technical product is overpriced or too expensive, it can become a problem for both the inventor and the users because if customers are unable to afford the product, the inventor or producer will be unable to make required amount for advancement or further researches

5.10. Safety

The security of artificial intelligence, augmented reality, data science, and virtual reality, as well as 5G networks, learning management systems, and apps, has yet to be determined. But it is low in huge data and the internet of things, as well as strong performance in blockchain technology, cloud computing, machine learning, and deep learning.

5.11. Proof of Work

Proof of work is a type of encryption proof whereby one group demonstrates to another that a specified amount of processing effort may be invested. In this situation, most smart education technology has no evidence of work, according to table 1, excluding blockchain technology and learning management systems and applications, which have greater proof of work in educational settings.

5.12. Operation Efficiency

Performance measures the intelligent allocation of educational resources between the increased efficiency of cloud technology, learning management systems, and applications. Artificial intelligence is reduced to a minimum; the blockchain is more efficient; data science is modest, and the internet of things is highly favourable.

5.13. Comparison of Modern and Traditional Educations

The enormous contrast between conventional education and current smart education is clearly shown in Table 1. In terms of holistic growth, connection and response, cognition, attention, delivery techniques, persistence, comprehensibility, and accomplishment of the students. Table 2 shows a feature-by-feature comparison of modern and traditional educations. Leveraging the power and advantages, effective technology adoption creates a scalable and cost-efficient smart education system. Educational administrators and end-users should be aware of how new technologies are being used in the classroom today, as well as instances of how these technologies are being used in entrepreneurial solutions [3].

Table 2. Feature wise comparison of modern and traditional educations [24] [3].

S/N	Parameter(s)	Traditional Education	Smart Education
1	Academic Independence	Classroom only	Through technology
2	Attainment Capability	Lower	Higher
3	Attention span	Very short	Fairly Large
4	Cognitive Ability	Limited	Enhanced
5	Evaluation	Prefixed	Continuous
6	Feedback	No provision	Evaluation with a feedback mechanism
7	Interaction	Limited	Enhanced
8	Learning Time	Fixed	Anytime & anywhere
9	Delivery	Teacher	Learner centric
10	Motivation	Teachers	self-motivated
11	Retention	Lower	Higher
12	Study type	Not promote	Promote. Group/ Collaborative
13	Understanding the ability	Limited	Much better

6. Ethical Consideration in Technology Usage For Smart Education

Avoiding the misuse or misconduct in the learning process, ethical usage of technology should be addressed in the process of learning through technology. According to [2], humans must first talk about morality in a form that computers can comprehend before giving robots a sense of morality. This means that moral and ethical algorithms must be written in a style that allows them to be codified. Regardless of any ethical guidelines or regulations governing the use of technology in smart education, the establishment of trust is a necessary requirement for the mutual advantage of smart and intelligent education systems [27]. Students and instructors should always have mutual point of view towards the usage of technology for smart education and penalties in place for those taking unethical steps towards smart education

7. Advantages and Disadvantages of Smart Education Technologies

As with any other innovation or advancement, there are always upsides and downsides [3]. The advantages and disadvantages of smart education are explained.

7.1. Advantages

a. Smart technology encourages students to stay engaged in their studies. Students may use a learning management tool, for example, one may look up more information about the topic being studied, play game based learning to supplement lectures, or focus on more advanced topics.

b. It encourages instructors and parents to communicate more. Teachers may utilize applications and software to report on a student's actions in real time, alerting parents of what is happening throughout the day.

c. The cost of using smart education technology in education systems is quite low. The expense of implementing new technology throughout the municipality can be significantly high, but student computers, tablets, and class materials are inexpensive.

d. It gives today's modern student new methods to learn. Students who grow up outside of the typical educational setting can realize their full potential even if they have access to technology.

e. Smart education technology enables educational administration to provide students with data access from a single point. A learner could use technologies to get all of the resources they need for a project in one place.

f. It improves access to student behavioural data. Technology assists the education system in identifying outstanding kids and continuing to push them toward increasingly challenging tasks in order to pique their interest in the academic setting.

g. Collaboration is encouraged in the classroom through the use of these technologies. When the teacher gives a lecture from a textbook, the student retains just a little amount of information. Students will never forget anything if they can instantly put what they have learnt into practice.

h. When lecturers employ technology in the classroom, they gain credibility. Parents who are hesitant to allow their children additional screen time for learning purposes may also criticize.

7.2. Disadvantages

a. The presence of technology can cause students to get distracted. When implementing reward-based activities to boost learning in an educational setting, students may focus on what they obtain through software or applications rather than what they learn.

b. During tests, students can text themselves. It also necessitates strict restrictions for the use of technology in tests and examinations that need an accurate evaluation of the student's knowledge in order to evaluate the student's overall growth.

c. Some individuals may be unable to distinguish between trusted and unsafe websites. There are many phony and exaggerated things on the internet these days, yet they all appear true.

d. Rather than implementing modern learning technology, some institutions are unable to pay instructors' wages on a yearly basis. Prioritizing the use of technology in the classroom puts individuals at the bottom of the pay scale at a significant disadvantage.

e. Today's active learning classes are so successful that the programs may function as the instructor rather than being hands-on; technology allows the educator to take a more passive role.

f. Annually, a substantial number of people are victims of some form of identity theft. We put our pupils' identities at jeopardy every day by bringing technology into the classroom. Even if effective privacy filters are built into programs, computers, portable devices, and operating systems to lower the risk of data theft, there is no means to ensure that all dangers have been eradicated until the equipment is never connected to the internet.

g. When you look at a computer display for a lengthy amount of time, you get eye strain. This ailment causes back ache, eye pain, neck pain, fatigue, decreased vision, and difficulties focusing.

h. Educators devote as much time to their pupils as parents do to their children. As a result, the classroom will become an integral part of each life of a student. That is why, if feasible, any school that incorporates technology into education should simultaneously promote up to 30 minutes of good physical exercise

8. Conclusion and Recommendations

With smart gadgets, modern technologies are enhancing today's learning with increased teaching and learning. This technology enhances the classroom experience by transforming the education platform to allow sophisticated real-time search, sharing, collaboration, and communication. Smart education, smart training, smart instructors, smart analytic, smart reporting, and smart learning spaces are all examples of smart education solutions. The differences between conventional and new technology continue to exist, as do the gaps between traditional and modern education, which is merely the conventional hesitation to adapt to modern means. Smart education, with the assistance of advanced technologies, simplifies the tasks of educating, studying, connecting, and cooperation, and makes rapid alerts more productive.

New technologies are finding their footing in the learning industry. More research should be conducted to offer a stronger link between technology and education, so that the application of rising technological trends can lead to a revolution in educational models and architecture, radically re-imagining how people approach learning in general.

References

1. Abdel-Basset, M., Manogaran, G., Mohamed, M., Rushdy, E.: Internet of things in smart education environment: Supportive framework in the decision-making process. *Concurrency and Computation: Practice and Experience* 31(10), e4515 (2019)
2. Aberšek, B., Aberšek, M.K., Sik-Lanyi, C., Flogie, A.: Ethical issues associated with the use of smart and intelligent learning environments. In: *Pannonian Conference on Advances in Information Technology (PCIT 2019)*. p. 155 (2019)
3. Ahad, M.A., Tripathi, G., Agarwal, P.: Learning analytics for ioe based educational model using deep learning techniques: architecture, challenges and applications. *Smart Learning Environments* 5(1), 1–16 (2018)
4. Al-Majeed, S., Mirtskhulava, L., Al-Zubaidy, S., et al.: Smart education environment system. *Computer Science and Telecommunications* 44(4), 21–26 (2014)
5. Alam, M.A., Saiyeda, A.: A cloud based solution for smart education. *International Journal of Smart Education and Urban Society (IJSEUS)* 11(2), 28–37 (2020)
6. Ardiny, H., Khanmirza, E.: The role of ar and vr technologies in education developments: opportunities and challenges. In: *2018 6th RSI International Conference on Robotics and Mechatronics (IcRoM)*. pp. 482–487. IEEE (2018)
7. Bajaj, R., Sharma, V.: Smart education with artificial intelligence based determination of learning styles. *Procedia computer science* 132, 834–842 (2018)
8. Brownlee, J.: *Machine learning algorithms from scratch with Python*. Machine Learning Mastery (2016)
9. Burns, M.: The internet of things in education : Tendencies and assumptions: The college puzzle (2017), <https://collegepuzzle.stanford.edu/the-internet-of-things-in-education-tendencies-and-assumptions/>
10. Chen, G., Xu, B., Lu, M., Chen, N.S.: Exploring blockchain technology and its potential applications for education. *Smart Learning Environments* 5(1), 1–10 (2018)
11. Chen, N.S., Yin, C., Isaias, P., Psotka, J.: Educational big data: extracting meaning from data for smart education (2020)
12. Collins, C.: Looking to the future: Higher education in the metaverse. *Educause Review* 43(5), 51–63 (2008)
13. Hwang, G.J., Xie, H., Wah, B.W., Gašević, D.: Vision, challenges, roles and research issues of artificial intelligence in education (2020)
14. Jeong, J.S., Kim, M., Yoo, K.H., et al.: A content oriented smart education system based on cloud computing. *International Journal of Multimedia and Ubiquitous Engineering* 8(6), 313–328 (2013)
15. Koo, H.: Training in lung cancer surgery through the metaverse, including extended reality, in the smart operating room of seoul national university bundang hospital, korea. *Journal of educational evaluation for health professions* 18 (2021)
16. Kuppusamy, P.: Smart education using internet of things technology. In: *Emerging Technologies and Applications in Data Processing and Management*, pp. 385–412. IGI Global (2019)
17. Kye, B., Han, N., Kim, E., Park, Y., Jo, S.: Educational applications of metaverse: possibilities and limitations. *Journal of Educational Evaluation for Health Professions* 18 (2021)
18. Liu, T., Zheng, H.: A study of digital interactive technology and design mode promoting the learners' metacognitive experience in smart education. *International Journal of Information and Education Technology* 11(10) (2021)
19. Makarov, A.: 10 augmented reality trends of 2022: A vision of immersion (Apr 2022), <https://mobidev.biz/blog/augmented-reality-trends-future-ar-technologies>
20. Mbombo, A.B., Cavus, N.: Smart university: A university in the technological age. *TEM Journal* 10(1), 13–17 (2021)

21. Milgram, P., Takemura, H., Utsumi, A., Kishino, F.: Augmented reality: A class of displays on the reality-virtuality continuum. In: *Telemanipulator and telepresence technologies*. vol. 2351, pp. 282–292. International Society for Optics and Photonics (1995)
22. Norbutaevich, J.T.: Use of digital learning technologies in education on the example of smart education. *Journal La Edusci* 1(3), 33–37 (2020)
23. Nyabuga, D.O., Nyasani, E.I.: Human-computer interaction enhancement in learning and teaching in schools (2018)
24. Palanivel, K.: Emerging technologies to smart education. *International Journal of Computer Trends and Technology (IJCTT)* 68(2) (2020)
25. Sathya, A., Panda, S.K., Hanumanthakari, S.: Enabling smart education system using blockchain technology. In: *Blockchain Technology: Applications and Challenges*, pp. 169–177. Springer (2021)
26. Schnabel, M.A., Wang, X., Seichter, H., Kvan, T.: From virtuality to reality and back (2007)
27. SHAIKH, N., KASAT, K., SHINDE, M.: Trust among faculty and students as an essential element of smart education system. *Journal of Contemporary Issues in Business and Government— Vol 27(3)*, 1569 (2021)
28. Shi, W., Liu, X., Gong, X., Niu, X., Wang, X., Jing, S., Lu, H., Zhang, N., Luo, J.: Review on development of smart education. In: *2019 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI)*. pp. 157–162. IEEE (2019)
29. Shoikova, E., Nikolov, R., Kovatcheva, E.: Conceptualising of smart education. *Electrotech. Electron. E+ E* 52 (2017)
30. Singh, H., Miah, S.J.: Smart education literature: A theoretical analysis. *Education and Information Technologies* 25(4), 3299–3328 (2020)
31. Sunasra, M.: Performance metrics for classification problems in machine learning (2019), shorturl.at/ceiKY

Babatomiwa Omonayajo is a Ph.D. student at Near East University's Department of Computer Information Systems. Graduated with high honors for his master's degree in 2021 from the department of Computer Information Systems at Near East University, and was then offered a scholarship to pursue a PhD degree. Artificial intelligence, Blockchain networks, programming languages, and information technology are among Babatomiwa's interests.

Fadi Al-Turjman is currently the Head of the Artificial Intelligence Engineering Department of the AI and Robotics Institute, a research center for AI and IoT. Prof. Dr. Fadi Al-Turjman earned his Ph.D. in computer science in 2011 from Queen's University in Canada. He is a full professor and the head of the Near East University Research Center in Nicosia, Cyprus. Prof. Al-Turjman is a leading authority in the areas of smart/intelligent IoT systems, wireless, and mobile networks' architectures, protocols, deployments, and performance evaluation in Artificial Intelligence of Things (AIoT).

Nadire Cavus, who is currently working as a lecturer and Head of the Department at the Near East University, Faculty of Economics and Administrative Sciences, Department of Computer Information Systems, graduated from the Near East University, Faculty of Economics and Administrative Sciences, Department of Computer Information Systems AND completed her doctorate education in 2007. Prof. Dr. Nadire Cavus's research interests include distance education, web-based teaching, technology-based teaching, mobile

learning, cooperative learning, instructional management systems, virtual learning environment design, programming languages, databases, data structures, Information systems, mobile technologies, management Information systems, Information technologies, mobile learning environment development, mobile teaching systems, programming logic, online communication, and online commerce.

Received: August 17, 2021; Accepted: May 25, 2022.

The Impact of Digital Transformation in Teachers' Professional Development During The COVID-19 Pandemic

Ayden Kahraman and Huseyin Bicen

Department of Computer Education & Instructional Technology, Near East University,
99138, Nicosia, Cyprus
aydenkahraman@gmail.com
huseyin.bicen@neu.edu.tr

Abstract. This article presents a study conducted to reveal the positive and negative aspects of professional development programs conducted to teachers in distance education during the novice COVID-19 pandemic and investigates whether the programs contributed to the digital transformation competencies the teachers acquired through these programs. The case study was carried out with a total of 30 teachers, who took part in the study voluntarily. Qualitative and quantitative data collection methods were used in order to ensure the validity, reliability, and consistency of the research. Once the teacher-oriented professional development program was completed, the teachers were subjected to an achievement test and a self-assessment questionnaire. A focus group interview was conducted to collect various views of 18 teachers regarding the program. This study also reveals that teacher-oriented professional development programs can be applied efficiently through online education and have a crucial role in strengthening and enhancing the technical competencies of the teachers involved in distance education.

Keywords: COVID-19; pandemic; teacher development, online teaching, distance education, digital transformation

1. Introduction

COVID-19 virus that initially emerged at the end of 2019 and later was declared a global pandemic by the World Health Organization has affected the entire world in a considerably short time. This pandemic has emerged with different scenarios in many ways. Education is one of these scenarios. The COVID-19 pandemic has changed our perspective on education and the way we interpret it. A pandemic that is effective on a global scale can be considered as a disaster by nature, and while this disaster can be the end of some things, it can also be interpreted as a sign of new beginnings. Within the scope of these thoughts, it is possible to say the impacts that Covid-19 has had on the world has led to the new normal and this will lead to a new world order [1].

With the onset of the pandemic, schools were closed and face-to-face education was suspended due to the global pandemic. 1.5 billion students, which corresponds to nearly half of the student population from all education levels were unable to proceed with

their education. In Northern Cyprus, approximately 50 thousand students at the primary and secondary education levels were affected by this problem [2].

“The COVID-19 pandemic has led to new inequalities that can push us further back. In order to make up for the interruption of education and recover the lost time caused by the pandemic, rapid emergency distance education applications have been implemented by many educational institutions all over the world. Education stakeholders have faced the unexpected challenges of providing emergency distance education. Significant efforts have been made to leverage technology to support distance learning, distance education, and online learning during the COVID-19 pandemic” [3].

Being able to have access to uninterrupted to the Internet in countries and regions is certainly beneficial to have online education, enabling schools to deliver their courses online and where teachers can manage control and share their teaching resources. The transition to home learning and working from home have contributed to a significant increase of time children and their parents shared. Consequently, parents have actively participated in supporting their children's online learning experiences. Moreover, distant education has also increased the participation of societies in creating a possible suitable learning environment [4].

Due to the COVID-19 pandemic, education almost in all schools have been discontinued, and face-to-face education at all levels has been suspended. This has made countries and schools deploy distance education modalities through the internet, online platforms, mobile apps, and television to ensure that students are not completely disconnected from their education processes.

In this process, teachers' academic support became necessary and they were involved in the distance education process. Countries that were able to integrate distance education into their education systems before the emergence of the COVID-19 pandemic did not experience significant challenges in this emergency distance education process. However, countries that have never used distance education experienced many challenges in this process. The fact that teachers were not ready in terms of technological competencies for this emergency distance education was one of these challenges [5].

Compared to the period in which schools were operational, education was implemented in the classes, and communication technologies were easily accessible, teachers have had to cope with a much more intense workload, demand, and more effort was expected from them while shifting from conventional education to urgent delivery of distance education. The closure of schools and the interruption of face-to-face education not only affected the students, but also their parents. Parents' concerns regarding the learning and development of their children have made them demand more attention and time for their children from the teachers [6].

1.1. Distance Education All Over the World during the COVID-19 Pandemic

Distance education has become one of the most challenging issues teachers faced during the transitional period from the traditional to online education. In this process, educational authorities need to enforce precautions, such as guiding teachers, providing them with the support they need, providing model practices, and creating environments that will enable collaboration between teachers to be carried out remotely. Teachers

have a crucial role in delivering quality education at all levels during and after the COVID-19 pandemic, so it is very important to make sure that teachers are provided with the required support to enable them to overcome the challenges likely to be encountered in this process [7].

The closure of schools and interruption of face-to-face education caused an emergency transformation from traditional to distance learning at all levels of education across the globe. During this process, both teachers and students, and even parents, have been provided with the support required to carry out distance learning. In this section, this study provides a deeper insight and data on how some countries switched from traditional to online learning or distance learning.

Existing infrastructure was used to support distance learning in Southern Cyprus during the Covid 19 process. By providing internet access for students at all levels of education, priority has been attached to receiving online education at home through computers or tablets. A remote synchronized training program was implemented using Microsoft Teams; more than 110,000 teachers and students have been provided with access to this software. Starting with high school teachers, intensive online teacher training courses have been provided by the Cyprus Pedagogical Institute. At the school level, teacher networks have been established to provide peer assistance in the use of distance learning tools. Students have been informed about how to use distance education tools with the help of their teachers [8].

Content repository and supporting educational material for all students have been uploaded to the web page of the Ministry of Education and the web page of each school. Public and private television channels have broadcasted courses and other educational programs especially for primary school students [9].

In North Cyprus, schools were closed as of March 2019, and accordingly, education began to be carried out asynchronously in the Moodle system of the Ministry of National Education. In addition, courses have been scheduled to be broadcasted on national television channels. In order to boost the teachers' competencies in using technology for remote learning, in-service training was held through online teaching. As of September 2019, schools started their new semester online supported with live classes [2].

1.2. Recent Studies

Many teachers who were not trained adequately on how to cope with such situations in their pre-service and in-service training, and who had never had such an experience before, were caught off guard for this process. In many countries, teachers were expected to take the initiative and execute their teaching by themselves without the help of others.

Teachers with limited technological competencies in providing remote learning faced challenges in carrying out their tasks effectively. In addition to the different competency levels of technology use between teachers, the resources of schools, students' access the digital tools, and their competency using these tools, the socio-economic profile of schools and students have a significant impact on how teachers conduct distance education [10].

In this context, many online training programs have been tailored and provided for teachers across the globe on emergency distance education. The study conducted by

Fauzi and Khusuma, presents the views of the school teachers in Indonesia regarding the current situation in delivering online education. Based on the data they acquired from their study conducted with 45 Indonesian school teachers, they stated that the teachers had information about online learning but they experienced problems associated with internet access and planning [11].

Johannes König, Daniela J. Jäger-Biela & Nina Glutsch, in their study that they conducted in Germany when the schools were closed, investigated the process regarding the adaptation of teachers to online teaching, and they not only focused on the extent of social contact that the teachers were able to establish with their students, but also the infrastructure and technological competencies required for online education. As a result of the research, it was revealed that training programs were needed to enhance technological and digital competencies of teachers for online education, and inferences were obtained in terms of the scope of this education [12].

In their study, Carrillo and Flores investigated the requirements of teacher training programs-online during the emergency transition from traditional to remote education. By analysing the practices on online teaching and learning, they underlined the significant role of online pedagogy in enhancing the competencies of the teachers involved in distance education [13].

The study conducted by Hebebcı, Bertiz, and Alan presented data based on the analysis of the views collected from teachers and students regarding the distance learning during the COVID-19 process. The study concluded that distance education modalities are effective provided that teachers are competent, content repositories and educational materials are adequate, and uninterrupted internet access is available for online learning. The study emphasized that restricted internet services, limited interaction, lack of materials, and infrastructural problems are factors that affect the learning delivery modality negatively. Additionally, the study drew attention to the necessity of restructuring in-service training [14].

In their study, Donitsa and Ramot investigated the role of collaboration between educational institution and telecom services in ensuring remote learning in a short time during the COVID-19 process. According to the data acquired through the study, they emphasized that in collaboration with schools and respective institutions, universities had successfully transition from traditional to emergency distance education in a relatively short time during the pandemic [15].

1.3. Purpose of the Study

This study has been conducted with the purpose of revealing the digital transformation in teachers by evaluating the effectiveness of in-service training tailored to enable the rapid transition of the teachers to emergency online education in order to prevent the loss of education during the COVID-19 pandemic.

1.4. Research Questions

In order to achieve the above-mentioned goals, the following questions were asked:

- a. What are the situations that work well in the use of distance in-service training application?
- b. What are the problems encountered in the use of the distance in-service training application?
- c. Has online teacher training during the Covid-19 pandemic affected the digital transformation in education?

1.5. Limitations

This study is limited to the use of a mixed research method. The quantitative part of the study is limited to the sample that consisted of 30 teachers who participated in the online professional development training programs voluntarily. For qualitative data, the population was limited to literature regarding distance education modalities during the COVID-19 process and the World Bank resources collected from March 2020 to June 2020, and among these resources, the sample was limited to five countries. The online professional development training program, which was carried out due to the urgent need for distance education, was designed considering that the participants had different technological competencies.

2. Materials and Methods

2.1. Case of the Study

This study has adopted a mixed research method based on acquiring, analysing, evaluating and integrating the data obtained through the qualitative and quantitative phases. The case study was adopted to identify the effect of the online teacher training program and ensure the consistency of the data acquired through the mixed research method.

2.2. Participants and Procedure

The Online New Generation Teacher Training Program tailored for professional development as well as to improve the teachers' competencies required for delivering distance learning was announced through social media channels when schools were closed during the COVID-19 pandemic. A total of 30 teachers from primary and secondary education consented to participate in the event voluntarily.

Table 1. Demographic data of the teachers

		Frequency(f)	Percentage(%)
Gender	Male	6	20
	Female	24	80
Age	20-29	2	7.7
	30-39	10	33.3
	40-49	16	53.3
	50-59	2	6.6
Professional Seniority	1-10	6	23.1
	11-20	11	42.3
	21-30	6	20
Subject	German	1	3.8
	Geography	3	10
	English	3	10
	French	2	6.66
	Mathematics	6	20
	Music	2	6.66
	History	2	6.66
	Turkish	5	16.6
	Biology	2	7.7
	Physics	1	3.8
	Primary teacher	3	10

2.3. Implementation Process

The instructional design model of the New Generation Teacher Training model is based on the ADDIE model, which stands for Analysis, Design, Development, Implementation, and Evaluation. The ADDIE model is an ideal model that can facilitate the e-learning development process in the field of e-learning. At each stage, the e-learning developer can see how far he/she has progressed. Because the ADDIE model is a linear system, it is easy to go back and see what has gone wrong. It is also versatile as you can use it for a variety of e-Learning applications [16].

The participants of the study participated in the application as a single group consisting of 30 teachers. At the end of the study, an achievement test and a self-assessment questionnaire were applied to the participants and a focus group interview was conducted. Figure 1 shows the ADDIE model design process for "Next Generation Teacher Training". The research was carried out in 5 stage within the framework of the determined method.

In order to acquire data on the competencies required for switching to emergency online learning due to the pandemic, a thorough overview of the relevant literature was conducted as the first phase of the study. In the second phase, considering the data acquired through the qualitative method during the initial phase, a questionnaire form was prepared to collect the teachers' views and learn what they knew about the distance education modalities. Then, based on the data acquired through the literature review and survey, a new generation teacher training program was developed to address the requirements. The third phase was the implementation phase of the training program based on the ADDIE model. In the fourth phase, the teachers that participated in the training program voluntarily were subjected to an achievement test and a self-assessment survey. In the last phase of the training program, a focus group interview was held with 18 of the teachers that participated in the program in order to collect their views regarding the teacher training program.

Analysis. With the purpose of tailoring the necessary online teacher training program to enable the teachers to enhance their competencies in delivering distance learning to their students through the internet, online platforms, mobile apps, and television during the emergency distance learning imposed by COVID-19 pandemic, the relevant literature has been reviewed. The data acquired through the literature review were analysed and evaluated to determine the steps of the quantitative phase by taking into consideration certain specifications of the target audience such as technological literacy, age, gender, professional experience and educational level. The roles of education material delivery mode and learning delivery modality in achieving efficiency in distance learning have also been investigated. A special focus has been attached to identifying the physical and institutional limitations. The technical requirements have been identified. Additionally, investigations have been carried out and solutions have been generated to ensure the consistency and compatibility of the learning objectives to be determined with the e-learning media.

Design. Based on the analysis and results, the teacher training program was tailored and scheduled so as to ensure the rapid transformation from traditional education to the emergency distance education. The outcomes and Web 2.0 tools to be used were determined. Training modules based on the content and objectives of the program were created. Google Meet was used as the online learning medium. Google Classroom was used for sharing educational material. The process of evaluating the online teacher training program that would be held was also determined.

Develop. By taking into consideration the requirements of the emergency distance education imposed by the COVID-19 pandemic, the New Generation Teacher training program has been tailored and announced to the teachers via social media. Education materials have been developed in line with the topics to be covered by the training program. The e-mail addresses of the teachers who would participate in the training program voluntarily have been taken and the sample group has been formed. Additionally, instant communication with the group was ensured via WhatsApp. The teachers have been provided with information about the use of the online learning media required for the online New Generation Teacher training program. The training program was scheduled to be held three times a week from 25 March to 22 April.

Implementation. With the participation of 30 teachers, the New Generation Teacher Training Program was initiated as scheduled on 25 March via the Google Meet App. Invitations and course materials relevant to the topic to be covered were mailed to the teachers before the commencement of the program. Google Classroom was used to streamline the process of sharing materials, projects, and exams. During the app process, the teachers were not only provided information regarding the use of Google Drive, Google Documents, and Google Forms, but also the use of Nearpod for interactive lessons. The teachers were also trained on preparing interactive video lessons by using Edpuzzle, and using Jamboard as a collaborative digital whiteboard. In addition, the features of screen recording tools were addressed. The participant teachers were provided adequate information regarding the use of Filmora and YouTube as video editing tools. The Canva design platform was used to create visual contents. In order to evaluate the outcomes of each module, a mini quiz was held or applicable projects were delivered after each lesson. The online lesson recordings were mailed to the teachers and they were provided with the opportunity to watch them again.

Evaluation. During the four-week-long training program, the participant teachers were provided with 24-hour training. Once the New Generation Teacher Training Program was completed, an achievement assessment consisting of 35 questions and a self-assessment questionnaire were applied. Additionally, a focus group interview was held with 18 participant teachers to allow them to express their views regarding the training program.

2.4. Data Collection Tools

The qualitative phase of the mixed-research method was used to collect data by scanning the literature concerning distance learning during the COVID-19 process. Based on the data acquired through the literature review, and the content of the teacher training program, the measurement tool for the quantitative phase was determined. The data of the research were collected through questionnaire, achievement test and focus group interviews. In addition, the opinions of field experts were consulted to ensure the scope and validity of the developed items and their suitability for the research. Consequently, in line with the views of the field experts, some items were revised and others were deleted and the final version of the questionnaire was completed.

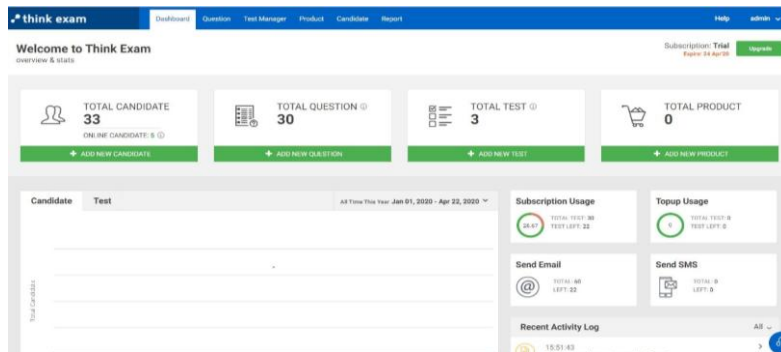


Fig. 1. Evaluation tool

Table 2. Demographic data of the teachers

Data Collection Tools	Female		Male		Total	
	f	%	f	%	f	%
Questionnaire	24	80	6	20	30	100
Achievement Test	24	80	6	20	30	100
Interview Questions	17	94.4	1	5.6	18	100

Questionnaire. An achievement test consisting of 30 items and a self-assessment questionnaire were applied to the teachers who participated in the teacher training program voluntarily. The self-assessment questionnaire used in the research was applied at the end of the training so that the teachers could make their own assessments (N= 35). A five-point Likert scale with the options (1) Strongly disagree, (2) Disagree, (3) Neither agree nor disagree, (4) Agree and (5) Strongly agree was used to enable the participant teachers to express the extent of their agreement or disagreement with the statements in the questionnaire. The achievement test was applied to the participant teachers at the end of the training program as well.

Focus-Group Interview. In addition to the survey questions, a focus-group interview consisting of five open-ended questions was held with 18 participant teachers to explore their views, provide suggestions regarding the New Generation Teacher Training Program and allow them to explain its effect on their accomplishments. The data acquired through the interviews have been provided by using codes instead of names to ensure confidentiality. Teachers' answers for each question were coded according to the frequency values within the determined themes and presented in tables. In addition, sample teacher statements have also been included in order to express the views of the teachers more clearly.

2.5. Data Analysis

SPSS 20 package program was used for statistical analysis of the data acquired through the quantitative phase of the mixed-method. Descriptive analysis (mean, frequencies) was used to calculate and evaluate the data acquired from the teachers' self-assessment questionnaire applied at the end of the training program. Item difficulty analysis was conducted to determine the item difficulty level for the achievement test applied upon the completion of the training program. Furthermore, quantitative data regarding the focus-group interview were analysed, the teachers' answers were coded, themed and tabulated.

Validity and Reliability. Through analysis, the validity and reliability of the data collection tools have been determined. In addition, in order to determine the validity of the scope of the questionnaire, items were revised in line with the views and suggestions of the field experts and the validity of the questionnaire was tested through an internal consistency test. Data from the focus group interview recorded using content were also analysed, the themes were revealed and coded by working with another field expert, and direct quotations from the teachers' answers have been presented in tables to show that the study presents data on what it claims to measure. Cronbach's alpha was used to assess the self-assessment questionnaire score reliability and internal consistency. A Cronbach's alpha value greater than 0.70 indicates that the questionnaire is reliable. The self-assessment questionnaire Cronbach's alpha value was calculated as 0.993, which is a significantly satisfactory coefficient. The KR-20 test was used to calculate the reliability of the achievement test used in the study. The KR-20 coefficient of the achievement test was calculated as 0.72. A KR-20 score above 0.70 is considered a satisfactory level of consistency and reliability [17].

Table 3. Difficulty levels of the items in the achievement test (p)

Difficulty indices			
	Difficult (0.00-0.39)	Moderate (0.40-0.69)	Easy (0.70 – 1.00)
Item numbers	(2,29)	(3,7)	(1,4,5,6,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30)

3. Results

The analysis of the study focused on revealing the positive and negative aspects of the online teacher training program and determining whether it contributed to the professional development of the teachers.

3.1. Questionnaire Results

"New Generation Teacher Training Program", teacher achievement evaluation. In order to prepare the questionnaire for the achievement test to be applied at the end of the online teacher training program within the scope of the study, the objectives and targeted outcomes of the New Generation Teacher Training Program have been determined. During this process, two field experts working in the in-service training unit were consulted for their views and suggestions to ensure the consistency of the questionnaire items in terms of the targeted outcomes. Once the outcomes were determined, a question pool was developed by the researchers so that each outcome would be addressed by at least one question. The achievement test consisting of a total of 30 multiple-choice questions was first presented for expert opinion and necessary corrections or changes were made. Then, at the end of the online teacher training program, the participant teachers were subjected to the achievement test. Item difficulty and distinctiveness indices of each question have been determined and presented in Table 3 and Table 4.

Table 4. Distinctiveness indexes of the items in the achievement test (r)

Distinctiveness indices			
Item Numbers	Difficult (> 0.40)	Quite distinctive ($0.30 - 0.39$)	Low distinctive ($0.19 - 0.29$)
	(3,4,5,6,7,11,12,13,30)	(1,2,4,8,9,10,15,16,17,18,19, 20,23,27,28,29)	(21,22,24,25,26)

As can be seen in Table 4, 9 items have strong distinctiveness, 16 items are quite distinctive and 5 items have a low distinctive index. As a result of the distinctiveness analysis, the mean of difficulty of the items was calculated as $p=0.75$ and the mean of the distinctiveness of items was calculated as $r=0.29$. The reliability coefficient (KR-20) of the 30 items in the achievement test was calculated as 0.72.

3.2. New Generation Teacher Training, Self-assessment Test

Based on the analysis of data acquired from the self-assessment test, it can be concluded that the online teacher training resulted in positive outcomes (total mean $x=4.07$). Considering the self-assessment results, it can be seen that the mean of all items is high. Item 4 and item 5 demonstrate highly significant means ($x=4.55$ and $x=4.51$). These values clearly indicate that almost all teachers who participated in the study gained competencies in using their e-mail accounts and Google Drive. Considering the teachers' answers to items 34 and 35, it is obvious that the teachers consider the online teacher training to be effective and the presentations are comprehensible. However, answers to items 22 and 24 show the lowest rates ($x=3.58$ and $x=3.62$).

Table 5. Self-Assessment Questionnaire

Items	Mean	Std. Deviation
1. I can easily involve in Google Meet communication medium.	4.24	1.37
2. I can easily use the communication medium.	4.17	1.39
3. I can actively participate in the course in the communication medium.	4.24	1.37
4. I can actively use my Gmail account.	4.55	1.24
5. I can upload files to my Google Drive account.	4.51	1.24
6. I can arrange my Google Drive files online.	4.37	1.23
7. I can share my Google Drive files.	4.37	1.23
8. I can share my Google Drive files through Google Classroom.	4.06	1.51
9. I can prepare quizzes by using the Google Forms App.	4.00	1.25
10. I can use different types of questions.	3.93	1.22
11. I can share the exams prepared by using Google Forms in different media.	3.96	1.37
12. I can find videos suitable to the course content through the Edpuzzle app.	4.00	1.33
13. I can add questions to the videos suitable to the course	3.96	1.32
14. I can share the materials prepared through the Edpuzzle app.	4.10	1.34
15. I can prepare e-presentations by using Google slide apps.	4.34	1.23
16. I can add images in the slide.	4.37	1.23
17. I can add a video in the slide.	4.37	1.23
18. I can give links in the slide.	4.06	1.36
19. I can add add-ons to a Google Slide.	4.06	1.38
20. I can use add-ons in a Google Slide.	4.10	1.37
21. I can use the NearPod app.	3.68	1.25
22. I can integrate the Nearpod app into Google Slides.	3.58	1.23
23. I can use Google Jamboard app screen	3.65	1.34
24. I can upload image or pdf file in the Google Jamboard.	3.51	1.35
25. I can edit the uploaded file.	3.75	1.18
26. I can share the edited Jamboard file.	3.62	1.37
27. I can login to YouTube with my Google account.	4.17	1.28
28. I can upload a video to YouTube.	4.10	1.20
29. I can use Edit option from Video Editor tab (Cut. split).	3.96	1.29
30. I can share YouTube videos.	4.31	1.22
31. I know the usage features of screen recording tools.	3.93	1.33
32. I can use the Loom screen recording tool.	3.65	1.39
33. I can record lessons by using the Loom recording tool.	3.75	1.29
34. Online lessons throughout the teacher training program were effective.	4.41	1.29
35. The presentations used in the training were comprehensible.	4.48	1.24
Total Average	4.07	1.30

When the means of the items are examined, it can be said that the online training, which has been tailored to enhance the competencies of the participant teachers on delivering distance learning through remote learning modalities, is effective. Although the means of items regarding the use of NearPod and Google Jamboard demonstrate lower values when compared to the others, the overall results clearly indicate that the content of the online teacher education is highly comprehensible.

3.3. New Generation Teacher Training, Focus-group Interview

The first question of the focus-group interview was intended to collect information regarding the advantages of the online teacher training program that was designed as an in-service training. Based on the teachers' answers to this question, it can be said that training was beneficial in terms of providing the teachers with the opportunity to

improve their technical competencies required for delivering distance learning through learning modalities of the new normal.

Another question of the focus-group interview was designed with the purpose of collecting information about the challenges encountered by teachers during the online in-service training program. Based on the teachers' answers, it can be said that most of them not only experienced internet connection problems, but also challenges in following-up the course. They interpreted these problems as the disadvantages of the in-service teacher training as it requires both uninterrupted internet access and hands-on apps.

Another question of the interview was designed to measure the satisfaction of the participant teachers with the content of the in-service training. Teachers' answers to this question make it clear that the teacher training program-online provided the participant teachers with the opportunity to enhance their technical competencies required for delivering online learning to their students much more effectively during the emergency distance learning process imposed by the COVID-19 pandemic.

Another question was intended to measure whether the teacher in-service training, which was designed to improve the teachers' competencies in using the distance learning modality by using technological or digital environments, contributed to any change in terms of using technology. The teachers' answers to this question make it clear that the online teacher training program provided a platform for them to not only learn about apps that they had never previously encountered but also to use these apps in delivering online distance education to their students through synchronous or asynchronous lessons. They also stated that the online-teacher training program was beneficial as it gave them the opportunity to use different programs actively for different purposes.

The teachers' perceptions regarding the online teacher training program have been revealed through another question. The participant teachers' answers to this question exhibit their satisfaction with the teacher online training in terms of equipping them with teaching technologies to address the requirements of delivering emergency distance learning during the pandemic. They indicated that the knowledge they acquired from the "New Generation Teacher" online training enabled them to use these technologies immediately in delivering their online lessons by turning the existing crisis process into an opportunity.

Table 6. "New Generation Teacher Training", teacher opinions

Main Theme	Codes	f	Selected Teachers' Opinions
Advantages of Conducting In-Service Training Online	Building and increasing technological competency	8	T1: Before this training, I could only search by using Google. Now I can design and deliver online courses.
	Opportunity for self-development (professional development)	6	T7: I have attended many in-service training programs for many years, but I consider this training as one of the rare ones that contributed to my professional development.
	Awareness of Educational Technologies	13	T8: For the first time in my life, I was able to teach online with my students using what I learned in this course. T13: I had the chance to implement programs that I have never applied or used before.

Disadvantages of Delivering In-Service Training Online and Challenges that experienced (Limitations)	Internet Access Problems	6	T8: Internet problems such as poor internet connectivity and interrupted internet access caused me to break away from continuous education. T14: The biggest problem encountered during this training is the poor internet connection.
	Follow-up the course	4	T2: From time to time, I had difficulties in following the lessons. It was a bit problematic to both watch and practice.
Digital Applications Used in Education and Online Teaching Methods	Beneficial	15	T8: I have been teaching online with my students for several weeks and I had the chance to share the practical details you taught with them, so I can practice in the field.
	Rich in content	8	T9: I had never thought that I could learn so many apps in such a short time.
	Providing permanent learning based on practice	9	T6: Hands-on digital applications have provided us with a more permanent learning. These digital applications will play a significant role in boosting the classroom interaction
	Knowledge and skill development	5	T7: I got to learn applications that I had already used but had no idea about their contents and features. I can say that I have learned a number of teaching methods and technologies that I have heard of but have not used in any way before. I learned new applications that I was not aware of in any way, and thanks to your completely innovative lessons, we can enrich our content.
Changes in Participants' Acquisitions in Technological Competencies	Use of different training programs	6	T8: Google Drive has impressed me a lot; everyone will see my comments regarding the projects while working with my students on a project basis. I firmly believe that project-based education will provide me with new opportunities. I know how to use Power Point, however, using Google Slide is much more practical and time-saving, especially when accessing resources, and when adding images and videos. Google Forms has been a miracle for me; it has made my work much easier and faster. Google classroom is so wonderful.
	Preparing lessons on online platform	5	T17: I realized that I could help my students be a part of the whole learning process, not only by watching, but also by taking part in online activities.
Perceptions regarding the training	Beneficial	17	T4: It was very difficult to attract the attention of the students with the traditional education approach. I had been thinking about what we could do. Thanks to this, we have learned a lot. I am very excited to put it into practice as soon as possible
	Opportunity to use and apply instructional technologies	4	T8: In line with the feedback from my students and parents, I can also put forward the view that this training is beneficial in terms of what it has contributed to me.
	Turning the crisis period into an opportunity	9	T16: It was very nice to learn and experience something new together with my colleagues by turning the crisis period into an opportunity.

4. Discussion

The COVID-19 pandemic has initiated a rapid digital transformation in societies and students and teachers as well as their families, school administration and society as a whole had to make significant adjustments to this period. Teachers and schools had to undergo this unexpected digital transformation without being fully prepared to meet the requirements of children's basic education. While some teachers showed great flexibility, creativity and perseverance in responding to the challenging situation of COVID-19, others encountered significant difficulties. More thought and emphasis need to be given on the empowerment of teachers, enhancements of the schools' technology and teacher education to enable teachers to take on a leader role in the digital transformation of education [18].

Another study examines the professional competencies of teachers working in secondary education institutions. They stated that teachers had difficulties in providing student motivation and preparing suitable digital materials for the lesson during the Covid-19 process, and they participated in the technological skill development trainings provided during this process [19].

During the Covid-19 pandemic, Uganda has taken teacher education as an education policy and made recommendations on teacher candidates who have the competencies required for 21st century skills and how these skills will be improved [20].

The "Next Generation Teacher Training" held during the COVID-19 pandemic was designed to assist the teachers at a time in which schools were closed and starting emergency distance education was starting. The online training provided the teachers with the opportunity to learn about online applications or improve their competencies in using these apps in delivering synchronous or asynchronous courses to their students through online learning modalities. The "New Generation Teacher" training gathered teachers with different technological competencies under a single online-learning platform. Within the scope of the study, the positive and negative aspects of the online in-service teacher training and its effectiveness have been investigated and unveiled. Ensuring the digital transformation of the teachers during the COVID-19 pandemic process is among the goals of this study. Since the ADDIE model has been adopted as the instructional design model of the training, teachers, who participated in the training online, experienced almost no problems in developing or improving their technological competencies. The online training provided a platform where all participants with different competencies worked collaboratively and contributed to each other. Izquierdo & Carabajo (2021), Acting as the teachers of their colleagues, summarized their experience of collaborative education. The educational program consisted of three phases. The aim of the program was to train teachers in designing, implementing and using technological tools online. They also conducted a workshop which also aimed at motivating and improving the emotional state of teachers, who were adversely affected by being subject to isolation due to the COVID-19 quarantine. This process has gradually empowered teachers' ability to use ICT tools, while also engaging in real-life communication, helping teachers start the new online academic term in a positive manner [21]. Due to the pandemic conditions, people and institutions now spend more time online more than ever before. Due to the extensive use of the internet, internet access or poor internet connection have become significant problems in our country. This is an important issue that has negative effects on many activities. Poor internet

connection has inevitably affected us negatively in terms of delivering the teacher training program online. This issue was underlined by some participant teachers while answering the questionnaire. Since the online teacher training was carried out with practical implementations, teachers with low technological competencies experienced difficulties using online applications at the beginning of the program. Consequently, those who did not know anything or knew a little about how to use online learning apps and modalities have had the opportunity to improve themselves and use these apps and modalities effectively. In general, the fact that they use what they have learned in this training in their online classrooms shows that this training has initiated a digital transformation in teachers. Considering the uncertainties experienced by teachers all over the world in this process, the teachers who participated in this training stated that they were very happy to turn this crisis period in education into an opportunity. Montoya and González (2019) emphasized that teaching models should be developed for in-service training. It also concluded that digital resources, events and courses should increase [22]. Teachers need more knowledge of tools to design digital resources to support their virtual environments [23] and they are also in need of more pedagogical and technological assistance from experts in the design of the virtual environment [24].

5. Conclusions and Future Studies

The urgent need for distance education due to the pandemic has led teachers to develop their technological competencies, which may otherwise last a span of many years. This study was carried out in order to reveal the positive and negative aspects of online in-service training and to ensure the necessary changes in the digital competencies of teachers during the COVID-19 pandemic. The research carried out during the pandemic has revealed the areas required in order for teachers to improve their professional development. This study will shed light on future studies.

The negative aspects of online teacher training revealed in this study can be taken into account and solutions can be generated or proposed in the studies to be carried out on this issue. Furthermore, taking into account the desire of teachers to work collaboratively, future studies can be tailored to address the need to produce digital content based on the collaborative project-based work of teachers. In this regard, this study constitutes a solid ground for future studies on this issue.

References

1. Agrawal, R., Srikant, R.: Fast Algorithms for Mining Association Rules. In Proceedings of Bozkurt, A.; Sharma, R. Emergency remote teaching in a time of global crisis. *Asian Journal of Distance*, 2020, i-vi. <https://doi.org/10.5281/zenodo.3778083>
2. Egelı, S.; Özdemir, M. An Overview of the Reflections Coronavirus (Covid-19). *Education and Society in The 21st Century*, 2020, Vol.9 Issue 27, 779-804.
3. UNESCO (United Nations Educational, Scientific and Cultural Organization). Startling digital divides in distance learning emerge. Available online: <https://en.unesco.org/news/startling-digital-divides-distance-learning-emerge> (accessed on 15 February 2022)

4. UNESCO (United Nations Educational, Scientific and Cultural Organization). Guidances for online education by IITE and its partners. Available online: <https://iite.unesco.org/news/guidances-for-online-education-during-covid-19-pandemic-by-iite-and-its-partners/> (accessed on 01 February 2022)
5. ilo.org. (2020, May 16). Available online: https://www.ilo.org/wcmsp5/groups/public/---ed_emp/---ifp_skills/documents/publication/wcms_743485.pdf (accessed on 28 January 2022)
6. Telli, E. Distance Learning Experiences of Teachers During Covid-19 Process: European Commision. Available online: <https://epale.ec.europa.eu/en/blog/distance-learning-experiences-teachers-during-covid-19-process> (accessed on 22 December 2021)
7. Can, N.; Koroğlu, Y. Evaluation of Distance Education Escalating During Covid-19 And Its Investigation From The Perspective Of Education Laborers. Vol.3 Issue 4, 370. Available online: <http://bilimveaydinlanma.org/content/images/pdf/mdt/mdtc3s4/covid-19-doneminde-yayginlasan-uzaktan-egitimin-degerlendirilmesi-ve-egitim-emekcileri-acisindan-incelenmesi.pdf>
8. The World Bank. Available online: <https://www.worldbank.org/en/topic/edutech/brief/how-countries-are-using-edtech-to-support-remote-learning-during-the-covid-19-pandemic> (access on 15 January 2022)
9. CEDEFOP. Available online: <https://www.cedefop.europa.eu/en/news-and-press/news/cyprus-responses-covid-19-outbreak>
10. Sönmez, E. D.; Cemaloğlu, N. Okullaşma Sürecinde Uzaktan Evde Eğitime Geçiş. *Insan & İnsan* 2021 Vol.8 Issue 27, 63-82. <https://doi.org/10.29224/insanveinsan.799402>
11. Fauzi, I.; Khusuma, I. Teachers' Elementary School in Online Learning of COVID-19 Pandemic Condition. *Jurnal Iqra': Kajian Ilmu Pendidikan*, 2020, 58-70. <https://doi.org/10.25217/ji.v5i1.914>
12. König, J. J.; Jäger-Biela, D.; Glutsch, N. Adapting to online teaching during COVID-19 school closure: teacher education and teacher competence effects among early career teachers in Germany. *European Journal of Teacher Education*, 2020 43:4, 608-622. <https://doi.org/10.1080/02619768.2020.1809650>
13. Carrillo, C.; Flores, M. COVID-19 and teacher education: a literature review of online teaching and learning practices. *European Journal of Teacher Education*, 2020 43:4, 466-487. <https://doi.org/10.1080/02619768.2020.1821184>
14. Hebebcı, M. T.; Bertiz, Y.; Alan, S. Investigation of views of students and teachers on distance education practices during the Coronavirus (COVID-19) Pandemic. *International Journal of Technology in Education and Science (IJTES)*, 2020, 267-282. <https://doi.org/10.46328/ijtes.v4i4.113>
15. Donitsa-Schmid, S.; Ramot, R. Opportunities and challenges: teacher education in Israel in the Covid-19 pandemic. *Journal of Education for Teaching*, 2020, 46:4, 586-595. <https://doi.org/10.1080/02607476.2020.1799708>
16. Morrison, K.; Ross, S. M.; Kalman, H. K.; Kemp, J. E. *Designing effective instruction* (6th ed.). Hoboken, NJ: Wiley, 2011
17. Büyüköztürk, Ş. *Sosyal bilimler için veri analizi el kitabı*. Ankara: PegemA Yayıncılık, 2005, 17
18. İlviri, N., & Sumita Sharma, L. V.-O. Digital transformation of everyday life – How COVID-19 pandemic transformed the basic education of the young generation and why information management research should care? *International Journal of Information Management*, 2020, Vol 55. <https://doi.org/10.1016/j.ijinfomgt.2020.102183>
19. Yikici, B.; Altınay, F.; Altınay, Z.; Sharma, R.C.; Daglı, G. Adoption of Online Education and Pedagogy as New Codes of Life for New Future in Rural Regions. *Sustainability*, 2022, 14, 5528. <https://doi.org/10.3390/su14095528>
20. Ssemपाल, C.; Mpišo, P.S.; and Mary, J.; Mitana, V.; “Initial Teacher Training in the Wake of Uganda’s National Teacher Policy and Covid-19 Lockdown: A Technical or Ethical

- Challenge?." American Journal of Educational Research, vol. 10, no. 1, 2022, 46-53. doi: 10.12691/education-10-1-5
21. Llerena-Izquierdo, J.; Ayala-Carabajo, R., University Teacher Training During the COVID-19 Emergency: The Role of Online Teaching-Learning Tools. Information Technology and Systems. ICITS 2021. Advances in Intelligent Systems and Computing, 2021, vol 1331. https://doi.org/10.1007/978-3-030-68418-1_10
 22. Montoya, N. E.; González, E. V. Competencias TIC en docentes de nivel técnico y tecnológico. Un estudio de caso en un centro de formación del SENA. Revista Virtual Universidad Católica del Norte, 2019, 58, 74–95. <https://doi.org/10.35575/rvucln.n58a3>
 23. Laro, E. Innovar enseñando: la educación del futuro. Las TICs como factor motivador en la enseñanza. REJIE Nueva Época. Revista Jurídica de Investigación e Innovación Educativa, 2020, 21(1), 11–2 <https://doi.org/10.24310/rejie.2020.v0i21.7530>
 24. Michos, K.; Hernández, D. CIDA: A collective inquiry framework to study and support teachers as designers in technological environments. Computers and Education, 2020, 143(1), 1–26. <https://doi.org/https://doi.org/10.1016/j.compedu.2019.103679>

Ayden Kahraman has been working as an ICT Teacher since 1999. In 2009 Ayden obtained her master's degree in Computer Education with her thesis on "How E-Learning effects students success and passion in Mathematics" in Near East University. From 2006 to 2010 she worked as a part time lecturer in the fields web design and the usage of Photoshop in Near East University. She took place in the Information and Communication Technology book commission created by the Ministry of Education. She is still continuing her studies and currently working on her PhD in Computer and Instructional Technology under the subject "Distance Learning". In 2020 she has been a massive part of improving online education in Cyprus by lecturing the teachers and has been assigned to The Ministry of Education under the unit "Professional Development".

Huseyin Bicen has been the Head of the Human Resources Development in the Education Department since 2013. He founded the Near East University Distance Learning Centre and served as the Director between the years of 2013-2019, the Vice Dean of the Faculty of Open and Distance Education in 2015-2019, and Acting Dean of the same faculty since 2019. Hüseyin Bicen has contributed to various social responsibility projects as well as Environmental Education and Technology Integration works, and is a member of the Administrative Board of the Accessible Informatics Platform. Hüseyin Bicen has published high-impact articles titled Distance Education, Social Media in Education, Technology Addiction, Massive Open Online Courses, Artificial Intelligence, Gamification, and Flipped Classroom in both national and international conference papers and book chapters with academic content indexed in the Social Sciences Index. In addition, he is an active participant in the science committees of international conferences as well as the referee boards of projects and journals.

Received: October 17, 2021; Accepted: June 25, 2022.

Data Mining Technology in the Analysis of College Students' Psychological Problems

Jia Yu* and JingJing Lin

NanChang JiaoTong Institute, Nanchang, Jiangxi, China,330022
fishpanda2011@163.com

Abstract. This paper expounds on the research status of data mining and the status quo of college students' psychological health problems, deeply analyzing the feasibility of introducing data mining technology into the analysis of college students' psychological health. After studying and analyzing the decision tree technology of data mining, and taking the psychological health problem data of the students in a university in 2021 as the research object, this paper uses the decision tree to analyze the psychological health problem data. The main work includes the following: determining the mining object and mining target; preprocessing the original data; and according to the characteristics of the data used, choosing the C4.5 algorithm of the decision tree to construct the decision tree of the students. Finally, based on the analysis and comparison of the decision tree model before and after pruning, classification rules are extracted from the optimal decision tree model, thus providing a scientific decision-making basis for mental health education in colleges and universities. After comparing the classification results with the known categories in the test set, the accuracy rate was found to be 75%. Using the alternative error pruning method and test data set, the classification accuracy was 79%, and after PEP pruning was 84%.

Keywords: data mining, data preprocessing, decision tree, college students' psychological problems.

1. Introduction

Data mining, also known as knowledge discovery, is conducted to discover hidden information deposits from massive data. It is a comprehensive application of statistics, artificial intelligence, databases, and other technologies [1]. Knowledge discovery in databases (KDD) was first proposed by Dr. Gregory Piatetsky Shapiro at the 11th International Joint Conference on Artificial Intelligence (IJCAI) in August 1989. However, at that time, multimedia outlets tended to use the term "data mining" rather than KD [2]. From 1989 to 1994, the American Association for Artificial Intelligence held four international symposiums on KDD. In 1995, the American Association for Artificial Intelligence renamed the symposium as an international conference. The first conference was held in the same year, and it became an annual event. ACM established the data mining and knowledge discovery committee (SIGKDD) in 1998 [3], and began to organize the ACM SIGKDD International Conference (KDD) in 1999. The

* Corresponding author

conference is the top annual conference in the field of data mining research. In addition, there are many conferences, organizations, and academic groups focusing on data mining and knowledge discovery, among which the more famous ones are IEEE ICDM, PAKDD, SDM, FSKD, MLDM, and VLDB. Data mining technology has been successfully applied in many fields, such as finance, insurance, securities, telecommunications, transportation, and retail. Many foreign computer companies also attach great importance to the development and application of data mining systems. Typical data mining systems include Enterprise Miner, intelligent miner, set miner, Clementine, warehouse studio, see5, cover story, knowledge discovery workbench, Extra, quest, and dB miner. Data mining has become a standard term and encompasses text mining, image mining, web mining, prediction analysis, and processing massive data (big data). Google vigorously promotes web mining and text mining, and launched the powerful Google Analytics in 2006 [4], which can carry out access data statistics and analysis on a target website and provide a variety of parameters for website owners to use.

A survey showed that the situation of college students' weariness, dropping out of school, suicide, and hurting others is mostly caused by mental health problems, and the number of students with poor mental health has been on the rise. A survey of 126,000 college students across in China shows that 20.3% of them have psychological health problems, mainly manifesting as fear, anxiety, obsessive-compulsive disorder, depression, and neurasthenia [5].

College students face different psychological health problems in different stages of development and deal with them in different environments. Freshmen have some psychological problems in some university. Freshmen are in the second "weaning period" of their life. If they leave home and do not adapt to the new college environment, they are prone to the contradiction between independence and dependence. Facing the gap between the ideal and reality, they are then prone to frustration, anxiety, depression, and even neurasthenia. However, their mental health will obviously improve as they adapt to university life. Before graduation, students' psychology will also change a lot. Junior college students are ready to take the entrance examination of postgraduate, and undergraduates are ready to take the entrance examination of postgraduate. Most of the remaining people are under pressure due to difficulties in finding employment after graduation. Confusion, fear, and irritability disrupt their psychological balance and make some people depressed, lose confidence, unable to find a goal in life, and even feel that life is meaningless [6].

The present paper attempts to apply data mining technology to the analysis of college students' psychological problems, in order to excavate the useful knowledge hidden in the data. We then hope to use this knowledge to predict the mental health status of college students more accurately, so as to provide a scientific basis for the planning and decision-making of mental health education. An additional goal is to make mental health education more effective [7].

2. Related Work

This section mainly introduces the purpose of classification, classification steps, and evaluation criteria of classification methods. We then introduce several commonly used classification algorithms, and analyze and compare the application scope, advantages, and disadvantages of these algorithms.

Data classification is a very important function in data mining and is employed to construct a classification pattern based on the given data. A model is then used to classify unknown data records in a database. At present, [x] is mainly used in commercial applications, but is gradually being applied to other fields[8].

First, the data in the training sample set is analyzed. According to the characteristics of the training samples with known class labels, a description (model) of the known class is constructed. Then, the class description (model) is used to classify the unknown data, and the class label of the unknown data is predicted.

Data classification can be summarized into two steps: building a model, and then using that model for classification.

2.1. Construction model (classifier)

We build a model (classifier) that describes a predefined data set or concept set. Each tuple in the scheduled data set has a class label; that is, it belongs to a scheduled class. The model is constructed by analyzing the attribute description of the database tuple. The data tuple set is regarded as a sample set, and the sample subset randomly selected from the sample set to construct the classification model constitutes a training sample set, in which each tuple is a training sample. Because the class label of each training sample is known in advance, the learning process in constructing the model is a guided learning process. Usually, the first step learning is used to construct the model in the form of decision tree, formula, or rule. Then, this model can be used to classify other samples and provide a deeper understanding of the database.[9]

2.2. Using a model (classifier) for classification

First, the prediction accuracy of the model should be evaluated. In classification method, samples independent of training samples and with known class labels are randomly selected as test samples; that is, the test data set is completely different from the training data set. Each test sample in the test data set is used to learn the prediction class according to the classification model and compare it with the known class label. If the labels are the same, the classification is successful. The accuracy of the model refers to the percentage of the number of test samples correctly classified by the model. If the accuracy is acceptable, the model can be used to classify the data tuples with unknown class labels.

3. Application of data mining in the analysis of college students' psychological problems

In Section 2, a decision tree algorithm was selected as the data mining algorithm to be employed for the analysis of college students' psychological problems. In this chapter, we use the decision tree algorithm to create a decision tree model of college students' psychological problems, prune it, and then extract rules from it. These rules are then used to find those attributes that have a greater impact on psychological problems. We then analyze their impact degree so as to provide a decision-making basis for psychological counselors of college students. The final decision tree model is also used to predict new data so as to aid in the prevention of psychological problems in college students.

3.1. Determining the research object and mining target

In this study, the mining data are the basic personal information of students provided by the psychological counseling center of a university and 2012 survey results of university students (the survey employs the symptom checklist 90). This paper randomly selects a psychological symptom and establishes a decision tree model through classification mining to determine whether college students have the psychological symptom.[10]

3.2. Data acquisition

In order to obtain the required data, this paper uses symptom checklist 90 (SCL-90) (see Appendix I) to evaluate 1700 students in 42 classes, 29 majors, and 5 departments in a university. The age of the students tested ranged from 17 to 24 years. A total of 1700 self-rating scales were distributed, and 1640 were actually recovered. According to SCL-90 instructions (see Appendix II), the scores of each student's psychological factors were calculated.

The data needed for the analysis of psychological problems are managed by the database management system SQL Server 2008. We create a "psychology" database under the graphical interface of SQL Server Management Studio of SQL Server 2008. In this "psychology" database, we then create a "personal psychological problems" table, and a "personal basic information" table. Table 1 presents the definition of the table structure of the "personal psychological problems" table. Table 2 defines the table structure of the "personal basic information" table. The "personal psychological problems" table stores the scores of all the psychological factors of all the tested students. The scores of each psychological factor can be used to judge whether the students have symptoms related to nine kinds of psychological problems. The data of the "personal basic information" table is directly obtained from the psychological counseling center of the University.

Table 1. Definition of table structure of "personal psychological problems"

Field name	type	maximum length	meaning
XH	C	10	student number
QT	n	4	somatization
QP	n	4	forced
Mg	n	4	interpersonal sensitivity
YY	n	4	depression
JL	n	4	Anxiety
Hostile	n	4	DD
KB	n	4	terror
PZ	n	4	paranoia
JS	n	4	psychotic

Table 2. Table structure definition of "personal basic information" table

Field name	type	maximum length	meaning
XH	C	10	student number
XM	C	10	name
XB	C	2	gender
CS	d	8	date of birth
MZ	C	10	ethnic groups
ZY	C	10	major
Source of SY	C	20	students
Is DS	C	2	the only child
Is DQ	C	2	a single parent family
JJ	C	10	family economic status

3.3. Data preprocessing

Data cleaning. Since some students do not actively participate in responding to SCL-90, or are distracted or hindered by some external factors, not all surveys were returned, and some were irregularly filled in. Thus, this data can only be used after cleaning. The specific method employed in this paper is as follows. First, the non-standard filling items in the self-assessment scale are regarded as vacant items, and then the number of vacant items is counted. If there are no more than three vacant items, the most selected item in the item will be used as the items option. Otherwise, the students' self-assessment scale will be voided directly by the method of ignoring samples. After data cleaning, 200 self-rating scales were removed, and 1440 were retained for later use. Since the calculation result of each self-rating scale corresponds to one record in the "personal psychological problems" table, the number of records in the "personal psychological problems" table is 1440.

Data integration. Data integration is the task of integrating the related data according to the mining goal and build a new and closely coupled data set. The data used in this paper is related to students' "personal basic situation" table and "personal psychological problems" table. Through XH (student number), the tables of "personal basic information" and "personal psychological problems" are naturally linked to form a table of "basic information and psychological problems".

Data selection. In order to improve the efficiency of mining, it is necessary to further simplify the data set, that is, delete those useless data according to the mining task, and determine the data set to actually be used in mining. The data selection work in this paper proceeds as follows:

1. Directly delete XH (student number), XM (name), and CS (date of birth) in the "basic information and psychological problems" table because these attributes are meaningless for mining.

2. The MZ (nationality) in the table of "basic information and psychological problems" is deleted because the minority students account for only 1.9% of all students (most of them are of Han nationality, and this attribute has little effect on the mining results).

3. Since the specific task of this paper is to analyze which basic attributes of students are related to interpersonal sensitivity symptoms and how relevant they are, the Mg (interpersonal sensitivity) attribute in the "basic situation and psychological problems" table is selected for data mining.

Finally, in the table of "basic situation and psychological problems", there are seven attributes describing the basic situation of students, namely, XB (gender), ZY (major), sy (place of origin), DS (only child or not), DQ (single parent family or not), JJ (family economic status), and Mg (interpersonal sensitivity). The table of "psychological problem analysis" is used as the data set to construct the decision tree model of whether students have interpersonal sensitivity symptoms.

Data conversion. Since some of the data in the data set is not conducive to data mining, before mining, we should carry out data conversion according to the principle of continuous data discretization. In this paper, the data conversion of the "psychological problem analysis" table includes the following steps:

1. The attribute value of Mg (interpersonal sensitivity) is continuous, and it is discretized. According to the symptom Checklist-90 instructions for use, "when using the 0-4 level scoring method, which psychological factor score is greater than or equal to 1, indicating that the person has symptoms in which psychological problems." Because this paper adopts the 0-4 grade scoring method, the value of the psychological factor is either 1 (symptomatic) or 0 (asymptomatic). In this way, the continuous data is discretized, which is convenient for mining.

2. Although the values of some attributes are discrete, there are too many categories, and thus the data must be converted before mining. For example, there are 29 kinds of attribute values of majors. We classify majors into three categories, namely, literature and history, science and engineering, and art. Student places of origin are scattered across the country, and there are bound to be many categories of attribute values. We

classify the attribute value of students' place of origin into three categories: urban, county, and rural. In this way, the number of data categories is reduced, which is conducive for mining.

3. In order to improve the efficiency of data mining, we try not to use Chinese characters. Chinese attribute values are replaced by English characters, numbers, or a combination of English characters and numbers. For example, xb0 represents male, zy1 represents literature and history, etc.

3.4. Constructing the decision tree

Basic strategy for decision tree construction. Most decision tree algorithms are greedy algorithms. Based on the training sample set and the associated class labels, the decision tree is constructed by top-down recursive partitioning [5]. First, an attribute is found as the splitting attribute of the training sample set using the splitting criterion of the algorithm. This attribute is taken as the root node of the tree, and the branches are established according to the different values of the attribute. The training sample set is divided into subsets. This whole method is called recursively for each branch to establish new nodes and branches. As the tree grows, the training sample set is recursively divided into smaller and smaller subsets until all subsets contain only samples of the same class; that is, they are all leaf nodes. Finally, a decision tree classification model similar to a flow chart is generated, and it can be used to classify new samples. A path from the top root node to the bottom leaf node is called a classification rule.

Attribute selection metrics. Attribute selection metrics, also known as split rules, determine how to split samples on a given node. Here are two popular attribute selection metrics: information gain and gain rate.

Let d be the training sample set with class label, and the class label attribute has m different values. There are m different classes, which is the sample set of classes in D . $|D|$ is the number of samples in D , and $|$ is the number of samples in D .

1. **Information gain.** Let node n store all samples of data partition D . The expected information for the sample classification in D is given by the following formula:

$$Info(D) = -\sum_{i=1}^m p_i \log_2(p_i) \quad (1)$$

where $[x]$ is the probability that any sample in D belongs to and is calculated by $| / |D|$. In fact, the above formula only uses the proportion of the number of samples of each class to the total number of samples. This is also known as the entropy of D . Entropy is a statistic used to measure the degree of chaos in a system.

Suppose the samples in D are divided by attribute a , which has different values. If the value of attribute a is a discrete value, then the attribute a can divide d into subsets, where the value of the sample in attribute a is a discrete value. These subsets correspond to the branches growing from node n . The expected information for sample

classification based on attribute a to attribute d can be obtained by the following formula:

$$Info_A(D) = \sum_{j=1}^y \frac{|D_j|}{|D|} Info(D_j) \tag{2}$$

where [x] is the weight of the subset with a value based on attribute a. The expected information needed for the classification of samples divided into D.

Knowing the value of attribute a results in a decrease in entropy, which can be obtained from the following formula:

$$Gain(A) = Info(D) - Info_A(D) \tag{3}$$

Classification allows us to extract information from a system and reduce the degree of confusion in that system; this results in the system becoming more regular, orderly, and organized. The more chaotic the system, the greater the entropy. It is obvious that the optimal splitting scheme is the one with the largest entropy reduction. Therefore, the ID3 algorithm selects the attribute with the maximum information gain (a) as the splitting attribute on the node. GainAN.

2. Gain rate. The ID3 algorithm uses information gain as the attribute selection metric and is biased to select attributes with more attribute values. Consider the following example. Based on the split of attribute XH (student number), i.e., because everyone’s student number is different, there will be as many partitions as the number of student number attribute values; these partitions are pure, and each partition has only one data record. According to formula (2), we can obtain the expected information of the sample according to XH (student number): (d) = 0 XH. According to formula (3), the information gain of this attribute is the largest, and it will be used as the first splitting attribute. However, for classification, it is meaningless to divide based on student number. C4.5 and ID3 have the same basic principle, but the difference is that in order to compensate for ID3’s disadvantage of using information gain to select attributes with more values, C4.5 uses gain rate instead as the attribute selection metric (splitting rule). The definition of information gain rate is as follows [48]:

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo(A)} \tag{4}$$

In the above formula, split information is used to normalize the information gain:

$$SplitInfo_A(D) = - \sum_{j=1}^r \frac{|D_j|}{|D|} \log_2 \left(\frac{|D_j|}{|D|} \right) \tag{5}$$

SplitInfo(D) represents the information generated by dividing the training sample set into plans corresponding to the outputs of the attribute test. ADAvv

Constructing the decision tree of college students' psychological problems. The data set used in this paper has three attribute values of specialty, student origin, and family economic status; meanwhile, the other attributes have only two attribute values. If the ID3 algorithm is used, it will tend to select attributes with more attribute values when selecting split attributes, which will affect the quality of mining. Conversely, the C4.5 algorithms can solve this problem well while still inheriting all the advantages of the ID3 algorithm. In order to make the mining results more accurate, we choose to use the C4.5 algorithms to construct our decision tree.

In this paper, we randomly selected 2/3 of the data set (960 records) to comprise the training sample set for decision tree mining, and the other 1/3 (480 records) as the test sample set for decision tree testing. Below we describe the whole process of building the decision tree model using the C4.5 algorithms.

1. Using formula (1), formula (2), formula (3), formula (4), and formula (5), the information gain rate of each split attribute in the training sample set is calculated.
2. By comparing the information gain rate of each split attribute, the split attribute with the maximum information gain rate is set as the root node of the decision tree. If the attribute has several values, the data set is split into several sub data sets. If the attribute has only one value, the split ends.
3. Step 1 and step 2 are performed recursively on each of the split sub data sets.

The training sample set is randomly selected from the data, we can know that the class label attribute Mg (interpersonal sensitivity) has two different values: 1 (symptomatic) and 0 (asymptomatic). Therefore, the training sample set has two different categories. There are 270 samples corresponding to class value 1, and 690 samples corresponding to class value 0.

According to formula (1), we first calculate the expected information of training sample set classification:

$$Info(D) = -\frac{270}{960} \log_2 \frac{270}{960} - \frac{690}{960} \log_2 \frac{690}{960} = 0.85714844$$

Next, we need to calculate the expected information of each split attribute. For example, XB (gender) has two different values: xb0 (male) and xb1 (female). Therefore, xb0 and xb1 can be divided into two categories. There are 370 samples in xb0, of which 40 samples have 1330 values in attribute mg, and 40 samples have 0 values in attribute mg. There are 590 samples in xb1, among which 230 samples have 1360 values in attribute mg, and 230 samples have 0 values in attribute mg.

Meanwhile, it can be seen that DS (only child or not) has the largest information gain rate of the attributes. Taking DS (only child or not) as the root node leads to two branches: only child and non-only child. The samples are thus divided according to these branches. We repeat the above steps to classify the sub data set of each branch, and then create new branches. With the increase in the number of branches, the sample data set is recursively divided into smaller sub data sets. Finally, the decision tree model is generated.

3.5. Tree pruning

Pruning is an effective method to simplify a decision tree. During the construction of the decision tree, the algorithm does not consider the existence of noise and outliers in data. In this way, many branches are constructed based on anomalies in the sample set, which leads to overfitting between the generated decision tree and the training samples. Furthermore, the generated decision tree will be overgrown, which reduces its practicability and readability and increases its dependence on the sample data. Moreover, while the accuracy of the decision tree may be very high, when the decision tree is used for the classification of new data, the accuracy can become very low. In order to deal with the problem of overfitting data and to obtain more general classification rules, it is necessary to prune the decision tree. In this paper, the decision tree has 50 leaf nodes, from which 50 classification rules can be extracted. This is a large number of rules, and some of them are too long to understand.

Pruning is carried out to improve the readability and classification accuracy of the decision tree model.

Pre-pruning. First pruning is to prune a tree by stopping its construction in advance. Because this pruning is done before node splitting, it is called pruning first. If the tree splitting ends at a node, the node will be replaced by a leaf, and there may be several samples belonging to other categories in the training sample set covered by this node [6]. Therefore, information gain, statistical significance, Gini index, and substitution error rate can be used to evaluate the pros and cons of splitting. If the partition of a node results in a split lower than a predetermined threshold, the partitioning is stopped. However, it is often difficult to determine the appropriate threshold. If the threshold value is too high, the decision tree will be too simple and lose its application value; if the threshold value is too low, the tree will be pruned too little, and thus some redundant branches cannot be cut off. The specific method of pruning first also includes presetting the minimum number of training samples contained in the training sample subset of each node, stopping the node splitting when it is less than the predetermined value, and presetting the highest level of the decision tree to limit its growth.

1. The principle of alternative error pruning.

In the process of constructing the decision tree, when the samples contained in the branches of a given node belong to the same class or there are no remaining attributes that can be used to continue to split the sample set, the division of the decision tree is stopped. If the subset of a branch of the decision tree is further split during the calculation, and if the number of samples in different classes varies greatly, the formula for error substitution rate can be introduced:

$$\text{pure} \frac{n - n'}{m} \quad (6)$$

where $[x]$ denotes the number of samples in the branch, $[x]$ denotes the number of samples of most categories on the branch, and $[x]$ represents the total number of samples in the training sample set. If the substitution error rate calculated by the formula is lower than the predetermined threshold, the branch will be transformed into a leaf node and

identified by most categories on the branch so as to prevent overfitting to a certain extent.

2. Pruning decision tree based on substitution error rate

In this paper, before constructing the decision tree model, we set the threshold at 0.3%. While building the tree, the substitution error rate of each branch is calculated according to formula (6). If the substitution error rate is lower than the predetermined threshold, the branch is cut and replaced with a leaf node, and then the category corresponding to most of the samples in the cut branch is used as the class label of the leaf node. For example, when node 4T in Figure 4.3 splits based on XB (gender), the value of XB (gender) is 70 on the branch with xb0 (male), and is 7068, the total number of known samples is 960. According to formula (6), the substitution error rate of the branch can be calculated as 0.2%, which is lower than the predetermined threshold of 0.3%. Then the division is stopped and the branch is replaced by a leaf node. The class label of the leaf node is considered to be category 0 (asymptomatic). The substitution error pruning method is employed when using C4.5 to generate the decision tree model.

Post-pruning. The post-pruning algorithm first conducts fitting and then simplification. First, the algorithm constructs a decision tree model that completely fits the training sample set, and then deletes several subtrees and replaces them with leaf nodes. Next, the class of most of the training samples in the subtree is used to identify the class of the leaf node. Since post-pruning is based on all the information of the decision tree and conducted according to a certain standard, the effect of post-pruning is generally better than that of first pruning. The four most commonly used post-pruning algorithms are reduced error pruning (Rep), probabilistic error pruning (PEP), minimum error pruning (MEP), and cost complexity pruning (CCP). When choosing the pruning algorithm, we should balance the accuracy and complexity of the decision tree, combine the characteristics of different algorithms with the actual sample set, and choose one algorithm or a combination of several algorithms according to the specific situation. Generally speaking, the rep algorithm is the simplest, but it is not suitable for a sample set with small amount of data; PEP is one of the most accurate algorithms; the CCP algorithm gets a smaller tree size than the rep algorithm; and the MEP algorithm gets a larger tree size but has a lower precision than the PEP algorithm.

1. Selection basis of the PEP pruning algorithm

We choose the PEP algorithm for pruning due to the following two considerations:

(1) The PEP algorithm does not need a pruning set, and this is useful when pruning a decision tree model based on a small data set. Since the data set used in this paper is small, the PEP algorithm is a good choice.

(2) The PEP algorithm has high pruning efficiency. Because this paper is to study the practical application of data mining technology in the analysis of college students' psychological problems, rather than the research of data mining algorithm, so the work efficiency is higher. Furthermore, the PEP algorithm adopts a top-down strategy, which makes it faster and more accurate than other pruning algorithms.

2. Principle of the PEP pruning algorithm

The PEP pruning algorithm was proposed by Quinlan in order to overcome the shortcoming that the rep algorithm needs to prune a data set separately. It does not need

pruning set, but prunes based on the error estimation of training data set. However, this will lead to a large error in the estimation error rate. Therefore, Quinlan introduced a continuity correction. In other words, it is assumed that each leaf node automatically misclassifies 1/2 of the instances it covers, and a constant is added to the training error of the subtree. When calculating the standard error rate, the continuous correction follows a binomial distribution.

We use [x] to represent the original tree; [x] represents the subtree with the node as the root node; (T) represents the number of misclassified instances at the node; and (T) represents the number of all instances covered at the node.

The classification error rate of node T is

$$r(t) = e(t) / n(t) \tag{7}$$

The PEP algorithm modifies this to

$$r'(t) = [e(t) + 1/2] / n(t) \tag{8}$$

Let s be the subtree TT of T, and l(s) denote the number of leaf nodes of TT. Then, the classification error rate of TT is

$$r'(T_i) = \frac{\sum_s [e(S) + 1/2]}{\sum_s n(S)} = \frac{\sum e(S) + \frac{L(S)}{2}}{\sum n(S)} \tag{9}$$

For the sake of simplicity, the error rate is replaced by the number of errors:

$$e'(t) = e(t) + 1/2 \tag{10}$$

The PEP algorithm adopts a top-down method. If the calculation result of a non-leaf node is

$$e'(T_i) \leq e'(T_i) + SE[e'(T_i)] \tag{12}$$

then t is cut off and replaced by its corresponding leaf. The standard error is defined as

$$SE[e'(T_i)] = \sqrt{\frac{e'(T_i)[n(t) - e'(T_i)]}{n(t)}} \tag{13}$$

Because each subtree is visited at most once during pruning, the PEP algorithm is faster and better than other algorithms and is recognized as one of the most accurate pruning algorithms. However, the top-down pruning strategy of PEP algorithm will bring the same "field of view effect" as the first pruning method, that is, when a node is pruned according to the pruning conditions, the node may have subordinate nodes that should not be pruned according to the pruning conditions.

1. Pruning the decision tree with PEP

The PEP algorithm is used to calculate each non-leaf node in a top-down way. This is done according to formulae (10), (11), and (13). The pruning results of each non-leaf node are shown in Table 4.5. The calculation results of non-leaf nodes, T24, T27, T29, T29, T32, t33, can make equation (12) hold. Therefore, the subtree with the six non-leaf nodes as root nodes is cut and replaced with leaf nodes. Then, the class label of the leaf node is identified by the class of most of the samples in the subtree. Finally, we obtain a pruned decision tree model.

Obviously, the size of the pruned decision tree model is smaller than that of the directly generated decision tree model, and this improves the readability of the decision tree and the classification speed of new data.

Result analysis. From the generated rules, it can be seen that the most relevant attribute of students with interpersonal sensitivity symptoms is whether they are the only child. After analysis, we can draw the following conclusions: among the students with interpersonal sensitivity symptoms, the proportion of those who are an only child is significantly higher than those who are a non-only child; the proportion of female students is significantly higher than that of male students; the proportion of art majors is higher than that of other majors; and the proportion of students from single parent families is higher than that of students from non-single parent families. Among the only child attribute, female students from rural areas or majoring in art generally have symptoms. In addition, it is more common for students with poor family economic status to have symptoms, and most of them come from rural areas. Therefore, when carrying out psychological counseling and counseling related to interpersonal sensitivity, we should pay special attention to high-risk groups, such as students who are an only child, females majoring in art, students from poor families, and students from single parent families, and give them timely and appropriate guidance and help. These rules obtained through data mining technology can help university management departments firmly grasp the initiative of work, practice the guiding ideology of prevention first and treatment second, and carry out targeted psychological counseling.

4. Model accuracy evaluation

This study randomly selected 2/3 of 1440 records (960 records) as the training sample set; the remaining 1/3 (480 records) was used to form the test data set. It was known whether each student had interpersonal sensitivity symptoms in the test data set. After comparing the classification results with the known categories in the test set, the accuracy rate was found to be 75%. Using the alternative error pruning method and test data set, the classification accuracy was 79%, and after PEP pruning was 84%. Obviously, the classification accuracy of the decision tree model generated after pruning by PEP was the highest. This was followed by the decision tree model generated by introducing the substitution error first pruning method, and lastly the decision tree model generated directly. The pruned decision tree model obviously has more accurate classification than the directly generated decision tree model. Overall, the decision tree model constructed in this paper achieved the expected classification effect.

5. Conclusion

According to the requirements of decision-making analysis of mental health education for college students, the classification and mining of college students' psychological problems is fully realized. First, we determine the mining objects and data mining objectives. Based on SCL-90 results and basic information of students, a decision tree model of whether students have interpersonal sensitivity symptoms is constructed. Then the training sample set is obtained by preprocessing the data. According to the characteristics of the training sample set, the C4.5 algorithm is selected to construct the

decision tree model and prune it. Then, the classification rules are extracted from the decision tree model and analyzed. Finally, the accuracy of the model is evaluated. This paper also compares the original tree with the pruned tree in terms of scale, extracted classification rules, and classification accuracy. The conclusion is that the pruned decision tree model is simpler, easier to understand, and more efficient than the directly generated decision tree.

Acknowledgment. This work was supported by Humanities and Social Science Research Projects in Colleges and Universities in Jiangxi Province in 2020 : " A Study on the Relationship between Psychological Capital, Academic Burnout and Subjective Well-being of Independent College Students" SZZX20138.

References

1. Zhao Ding. Research on emotional data analysis and psychological early warning system based on data mining and intelligent computing [J]. Electronic production, 2021 (02): 88-90
2. Xu Meijuan, Xiao Xiong, xiong Ye, Dong Liangliang. Analysis of the "heart" way out of data mining technology in college mental health education [J]. Science and technology and innovation, 2020 (22): 43-44 + 47
3. Dai Qun, Wang Yibo. Research on academic early warning system of college students based on data mining [J]. Rural staff, 2020 (20): 230-231
4. Liang Yueying, Yuan Fang, Zhang Jian. Analysis of research hotspots of health education for college students at home and abroad based on data mining [J]. Chinese Journal of multimedia and network teaching (first ten issues), 2020 (10): 46-48
5. Yang Xun. Research on early warning of College Students' Psychological Crisis Based on big data technology [J]. Digital world, 2020 (10): 88-89
6. Chang Ni, Yu Yanjun. Dilemma and countermeasure analysis of psychological crisis early warning in Higher Vocational Colleges from the perspective of big data [J]. Intelligence, 2020 (26): 92-94
7. Gu Yongcheng. Analysis and Research on students' mental health based on Internet behavior data [D]. Guangdong University of technology, 2020
8. Fu ronghua. Application of association algorithm based on decision tree in rural college students' information system [J]. Hubei Agricultural Sciences, 2020,59 (10): 150-153 + 158
9. Chen haoquan, Hu Ruiyu, Zhao Zheng, Wang Shi. Design of mental health problem prevention platform based on data mining and database [J]. Science and technology horizon, 2020 (14): 49-51
10. Lin Jingyi, Li Dakun, Wu Pingxin, Wang Xu, Zhou Yan. Modeling and analysis of mental health early warning based on social data mining [J]. Electronic technology and software engineering, 2020 (08): 172-173

Jia Yu has Master of Applied Psychology, and works as Lecturer. She graduated from the Jiangxi Normal University in 2012 and worked in NanChang JiaoTong Institute. Her research interests include mental health of college students and counseling.

JingJing Lin has Master of Management. She graduated from the Jiangxi University of Finance and Economics in 2011. She worked in NanChang JiaoTong Institute. Her research interests include government behavior and public policy.

Received: April 04, 2021; Accepted: May 20, 2022.

CIP – Каталогизacija y publikaciji
Народна библиотека Србије, Београд

004

COMPUTER Science and Information
Systems : the International journal /
Editor-in-Chief Mirjana Ivanović. – Vol. 19,
No 3 (2022) - . – Novi Sad (Trg D. Obradovića 3):
ComSIS Consortium, 2022 - (Belgrade
: Sibra star). –30 cm

Polugodišnje. – Tekst na engleskom jeziku

ISSN 1820-0214 (Print) 2406-1018 (Online) = Computer
Science and Information Systems
COBISS.SR-ID 112261644

Cover design: V. Štavljanin
Printed by: Sibra star, Belgrade