Contents

Editorial

Papers

Computer Science and Information Systems

Special Issue on Emerging Services in the Next-Generation Web: Human
Meets Artificial Intelligence

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 2019 two-year impact factor 0.927,
· 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 (MS Word or 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 Word or 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 18, Number 2, Special Issue, April 2021

CONTENTS

Editorial

Papers

Guest Editorial
Emerging Services in the Next-Generation Web: Human Meets Artificial Intelligence

Neil Y. Yen[1], Hwa-Young Jeong[2], Kurosh Madani[3], Francisco Isidro Massetto[4]

[1] School of Computer Science and Engineering, University of Aizu,
Tsuruga, Aizuwakamatsu, Fukushima 9658580, Japan
neilyyen@u-aizu.ac.jp
[2] Humanitas College, Kyung Hee University,
1 Hoegi-dong, Dongdaemun-gu, Seoul, 130701, Korea
hyjeong@khu.ac.kr
[3] Laboratoire LISSI, University Paris-Est Créteil,
61 Avenue du Général de Gaulle, 94000 Créteil, France
madani@u-pec.fr
[4] Campus Santo André, Federal University of ABC,
Av. dos Estados, 5001 - Bangú, Santo André - SP, 09210580, Brazil
sung@mme.dongguk.edu

Recent prompt development in ICT (Information and Communications Technology) has caused a great impact to our daily lives. We can use a browser on laptop to access our works, an app on phone to chat with our friends, and a pre-deployed sensor at home to execute our commands remotely. Almost all of above mentioned instances imply that an impartible relation between human beings and ICT services exists.

Service provision is always a key factor to ensure the success of in ICT as well as the web development. In the past, developers, and researchers as well, often provide services by predicting what, and how, target users would be expecting. Empirical study, e.g., questionnaire, field study, etc., of course, is conducted to achieve the purpose. But however, reaction time of services on the web to be updated is way less than expectations from users (i.e., human beings). This issue can be formulated as the more we can understand the human, the more precise services we can provide to our users. Prediction, and/or anticipation, of human beings through the support of artificial intelligence techniques thus becomes an emerging topic in order to better develop the next-generation web.

What is the difference between prediction and anticipation in technosocial systems? Is there a common anticipatory feature in biological structures, cultural structures, and technological ones? Humans remain, either individually or collectively, very poorly skilled when it comes to foresee the outcomes of their actions and take inspired decisions. The practice of prediction has made effective progress in the last decades in

certain disciplines and thanks to intelligent systems, but mostly as a mechanistic and probabilistic protocol, based on reactive causation and often keeping the human factor out of the loop because of its complexity. Anticipatory system is an implementation to compensate disadvantage of the system with the factors from human. In a view of computing and engineering, anticipatory system is the one that can effectively make the forecasting, where the outcome of the forecasting affects the forecaster, and the one whose current states can be influenced by the future states. Effectual anticipation is a desired model of the future that acts in the present, a way of acting which does not obey the instinct of immediate gratification but uses final causation and deeper aspirations. It is a sensibility for destiny ramifications, the capacity to imagine and project into the future the consequences of our intentions.

In order to achieve the goal of anticipation, intelligence, in which we shall look at it from both human intelligence and machine intelligence parts, plays an important role. The intelligence here defines an integration of different techniques, approaches and thoughts to overcome individual limitations and achieve synergistic effects. It has been applied to provide human-like expertise embodied by domain knowledge, uncertain reasoning and adaptation to complex environments for specific purposes (e.g., human supports). It covers a wide areas ranging from science, engineering and business and is considered as an essential approach to tackle real-world research problems. Considering mentioned phenomenon, a promising shift from system centric to human centric is revealed by the growth of technology and related commercial products, such as Apple i-devices, Google Glass, Tesla electric vehicle, and etc., which dramatically change our perspective upon information technology, in recent years. Researchers (and company as well) tend to provide tailored and precise solutions wherever and whenever human beings are active according to individuals' needs. Making technology usable by and useful to, with interdisciplinary concerns and hybridization of intelligent technologies, human beings in ways that were previously unimaginable has becoming an emerging issue to explore potential supports in the next era.

This special issue aims at revealing the fundamental technologies and potential research focuses on the web and its potential applications. We especially expect to find out the relation between human beings and everyday-growth intelligence. Understanding the context is the key to make achievement. The notion of context, in fact, is not new but require further explanation. The new scenario of intelligence under different types of context is more complicated. The need of multimodality (or interdisciplinary) should be defined between different context and different type of intelligence. On the other hand, intelligence among a group of people relies on typical media, or simply known as the approach that implements the intelligence. Although intelligent techniques became a common way to implement smart systems, a wide spectrum of issues (e.g., different computing paradigms and their applications) need to be taken into deep consideration. How to efficiently use different approaches to design an efficient, friendly-accessed, and high-performance hybrid intelligent systems to ease users (and their existing environment) remains a challenge. The development of a new

style of hybrid intelligence will need sophisticated adaptation techniques for different smart devices under different context.

This special issue received 95 submissions from 16 countries where the corresponding authors were majorly counted by the deadline for manuscript submission with an open call-for-paper period of 4 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, 12 papers were accepted for publication. These accepted papers mainly look at our issue from the union of human-centric design, machine learning, blockchain techniques, statistical methods, ICT-enabled service provision, smart living, social, privacy and security related issues for next-generation web which have brought lively discussions to the publics.

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

The first paper titled "A Novel Distant Target Region Detection Method Using Hybrid Saliency-Based Attention Model Under Complex Textures", by Jaepil Ko et al., proposes a hybrid visual attention model to effectively detect a distant target. The proposed model employs the human visual attention mechanism and consists of two models, i.e., the training model, and the detection model. When the image containing the target is input into the detection model, a task of selectively promoting only features of the target using pre-trained data is performed. The authors found that the desired target is detected through the saliency map created as a result of the feature combination. In this work, the model has been tested on various images, and the experimental results demonstrate that the proposed model detected the target more accurately and faster than other previous models.

The second paper titled "Exploring the Effectiveness of Deep Neural Networks with Technical Analysis Applied to Stock Market Prediction", by Ming-Che Lee et al., presents a work that explores the feasibility and efficiency of deep network and technical analysis indicators to estimate short-term price movements of stocks. A four-layer Long Short-Term Memory (LSTM) model was constructed. This work uses well-known technical indicators such as the KD, RSI, BIAS, Williams% R, and MACD, combined with the opening price, closing price, daily high and low prices, etc., to predict the trend of stock prices. It shows that the combination of technical indicators and the LSTM deep network model can achieve 83.6% accuracy in the three categories of rising, fall, and flatness.

The third paper titled "Text recommendation based on time series and multi-label information", by Yi Yin et al., proposes a novel method to ameliorate the correlation analysis issue in the recommendation method using the time series. The authors specify

a certain text collection according to the interests of users and integrate the varied label values of the text and build the correlation coefficient between text and its related text with the differential analysis. Finally, the similarity degree of the text is calculated out using the improved cosine similarity correlation matrix to promote a recommendation of similar text. The experiments indicate that the proposed method can ensure the quality of the text, with an improvement of accuracy by 8.63% as well as an improvement of recall rate by 5.25%.

The fourth paper titled "Message Propagation in DTN Based on Virtual Contact of Behavior Model", by Ho-Hsiang Chan et al., presents a work that simulates message propagation in a delay tolerant network (DTN), which is a kind of network structured to deliver messages intermittently, using virtual contact of a behavior model. The paper considers a scenario in which nodes make virtual contact in cyberspace and incur message delivery based on their behavior patterns. The verifying experiment is conducted using both survey and simulation that analyzes how messages propagated in different behavior pattern groups. It is derived from the simulation that to quicken message propagation, directing messages to one of the behavior groups yields the maximum benefits. It provides the basis for further research on collecting data of desired scenarios to establish respective propagation models.

The fifth paper titled "Enhanced image preprocessing method for an autonomous vehicle agent system", by Kaisi Huang et al., proposes a deep time-economical Q network (DQN) input image preprocessing method to train an autonomous vehicle agent in a virtual environment to solve the training cost issue of neural networks. The current frame top-view image is combined with the images from the previous two training iterations. The DQN model uses this combined image as input. The experimental results indicate higher performance and shorter training time for the DQN model trained with the preprocessed images compared with that trained without preprocessing.

The sixth paper titled "A Study of Universal Zero-Knowledge Proof Circuit-based Virtual Machines that validate general operations & reduce transaction validation", by Soon Hyeong Jeong et al., studies the zero-knowledge proof algorithm for general operation verification in the blockchain network. In this system, the design of a zero-knowledge circuit generator capable of general operation verification and optimization of verifier and prover was also conducted. This work develops an algorithm for optimizing key generation. Based on all of these, the zero-knowledge proof algorithm was applied to and tested on the virtual machine so that it can be used universally on all blockchains.

The seventh paper titled "Image Target Detection Algorithm Compression and Pruning Based on Neural Network", by Yan Sun et al., optimizes and compresses some algorithms by using early image detection algorithms and image detection algorithms based on convolutional neural networks to handle the issues of a large number of parameters and high storage and computational costs in detected models. This work discusses the Faster-RCNN algorithm and the YOLO algorithm. A target detection

model based on the Significant area recommendation network is proposed to solve the problem that the candidate frame is not significant which is extracted in the Faster-RCNN algorithm. Experiments show that the image detection algorithm based on compressed neural network images has certain feasibility.

The eighth paper titled "Collaborative Filtering Recommendation Algorithm in cloud computing environment", by Pei Tian, studies the collaborative filtering detection algorithm in the cloud computing environment for personalized recommendation technology. The algorithm migrates the collaborative filtering detection technology and applies it to the cloud computing environment. It shortens the recommendation time by using the advantages of clustering. A new recommendation algorithm can improve the accuracy of recommendation and proposes a parallel collaborative filtering recommendation algorithm based on the project. The algorithm is designed with a programming model. The experimental results show that the proposed algorithm has a shorter running time and better scalability than the existing parallel algorithm.

The ninth paper titled "The Application of Virtual Reality Technology in the Digital Preservation of Cultural Heritage", by Hong Zhong et al., mainly studies the application research of virtual reality technology in the digital preservation of cultural heritage. First, the system creates an immersive environment for users, displays the objects realistically in the virtual reality system, thereby digitizing the technical protection of cultural heritage. Secondly, it uses the virtual environment model of material cultural protection to build and use the terrain to generate and edit. Finally, the radial basis function is used to calculate the value in the virtual environment, so that the digital preservation of cultural heritage is more accurate. Experimental data shows that 35.54% and 64.46% of users are more likely to use the handle to interact with three-dimensional objects. This study indicates that the virtual environment reality technology specification is more efficient than the original technology in the process of digitizing cultural heritage.

The tenth paper titled "Extraction of Mosaic Regions through Projection and Filtering of Features from Image Big Data", by Seok-Woo Jang, proposes an algorithm that detects mosaic regions blurring out certain blocks using the edge projection. The proposed algorithm initially detects the edge and uses the horizontal and vertical line edge projections to detect the mosaic candidate blocks. Subsequently, geometrical features such as size, aspect ratio, and compactness are used to filter the candidate mosaic blocks, and the actual mosaic blocks are finally detected. The experiment results show that the proposed algorithm detected mosaic blocks more accurately than other methods.

The eleventh paper titled "Network Analysis of Social Awareness of Media Education for Primary School Students Studied through Big Data", by Su-Jeong Jeong et al., aims to examine the social debate about media education in Korean society, how media education is being conducted in this important primary school period. The data was collected in the last 5 years (2014.08.07-2019.08.07) from internet portal sites with

keywords of "primary school media education" and "primary school media literacy". Semantic network analysis, CONCOR analysis, and content analysis were used as data analysis methodology. This study reveals that it is the right time to provide future education that can have a sound digital identity so that media education can be achieved in a media-friendly local community and educational environment.

The twelfth paper titled "Machine Learning Based Distributed Big Data Analysis Framework for Next Generation Web in IoT", by Sushil Kumar Singh et al., introduces a machine learning-based distributed big data analysis framework for the next-generation web in the internet of things (IoT) to solve the issue of latency, accuracy, load balancing, centralization, and others in the cloud layer when transferring the IoT data. This study utilizes feature extraction and data scaling at the edge layer paradigm for processing the data. Extreme learning machine is adopting in the cloud layer for classification and big data analysis in IoT. The experimental evaluation demonstrates that the proposed distributed framework has a more reliable performance than the traditional framework.

A Novel Distant Target Region Detection Method Using Hybrid Saliency-Based Attention Model under Complex Textures

Jaepil Ko[1] and Kyung Joo Cheoi[2, *]

[1]Department of Computer Engineering, Kumoh National Institute of Technoligy,
Daehak-ro 61, Gumi-si, Gyeongbuk, 39177 Korea
nonzero@kumoh.ac.kr
[2]Department of Computer Science, Chungbuk National University,
Chungdae-ro 1, Seowon-gu, Cheongju-si, Chungbuk, 28644 Korea
kjcheoi@chungbuk.ac.kr

Abstract. In this paper, a hybrid visual attention model to effectively detect a distant target is proposed. The model employs the human visual attention mechanism and consists of two models, the training model, and the detection model. In the training model, some of the features are selected to train in the process of extracting and combining the early visual features from the training image of the target by bottom-up manner, and these features are trained and accumulated as trained data. When the image containing the target is input into the detection model, a task of selectively promoting only features of the target using pre-trained data is performed. As a result, the desired target is detected through the saliency map created as a result of the feature combination. The model has been tested on various images, and the experimental results demonstrate that the proposed model detected the target more accurately and faster than other previous models.

Keywords: target, hybrid, saliency, attention

1. Introduction

So far, lots of various studies on target detection have been conducted. Many of the studies on target detection have shown attempts to use human visual attention mechanisms for target detection [1, 2].In particular, in many intelligent robotic systems developed to assist humans [3], selecting only information useful for the current task from a large number of image information is very important in terms of efficient use of information as well as computational efficiency. To select such useful information, a various method has been proposed to increase the efficiency of computation by adaptively selecting only the high-priority information related to the current work from the numerous information of the input image using the selective visual attention mechanism of a human being. The mechanism of selective visual attention has long been first studied by cognitive scientists and has attracted a lot of interest in the computer vision area because this capability helps find conspicuous objects or regions.

---

* Corresponding author

Treisman suggested that the attention region is determined by integrating various features in the input image. Based on this feature integration theory [4], visual attention studies have been done largely into two approaches, a bottom-up approach, and a top-down approach. The first computational model for human visual attention was the model presented in [4] and was inspired by the feature integration theory [4]. Numerous successful bottom-up computational models have been developed based on this model [6-10]. This visual attention model has been expanded to be used for videos using motion information which is one of the temporal features, as a bottom-up clue [19-22]. In addition, a model using depth information was proposed to apply to the 3D image [23-26]. The main differences between the most successful computational models proposed so far are in methods of extracting features and generating the saliency map [27]. The word 'Saliency' means the intensity of a feature of an input image and can be described as a difference between a pixel and its surrounding neighborhood [1]. An area with high saliency is an area attracting attention.

Many successful bottom-up visual attention models have been proposed, but these models have a limitation that the attention region in the image does not always match the desired target. Detecting conspicuous objects or regions still remains a difficult issue. Therefore, a number of studies based on the top-down approach have been conducted to solve these kinds of issues. The top-down approach is based on the fact that a human pays attention to an object or part of an image that he or she already knows before another object [28]. This approach extracts and combines the features in the same process as the bottom-up approach, but makes the objects that want to be more noticeable by applying high-level knowledge to the extracted basic features [29]. High-level knowledge used here is feature information such as color, shape, and intensity learned about the object and can be obtained by human learning function. This learning knowledge simplifies information about objects according to certain rules. For example, consider the case that we are learning about a beverage can. If the cans of beverages are red and blue, and red has a relatively higher visual attention than blue, we do not learn both blue and red, but rather simplify them to red. Actually, Coca-Cola cans are mixed with white and red, but we just cognize that "Coca-Cola" cans are red because of these reasons.

Although a number of models have been proposed, the fast and accurate detection of salient regions remains a challenge in target detection, particularly in cases of complex textures. In this paper, a hybrid visual attention model for distant target detection that is robust to the color environment is proposed. The model proposed here is designed to extend the capabilities of the previous model that uses high-level information on the target. It overcomes the limitations of the bottom-up visual attention approach, which does not accurately detect the desired object, and the limitation of the top-down visual attention approach, which has only the feature values related to the object to be detected.

This paper is organized as follows. In Section 2, previous studies of target detection using top-down information that shares the basic frame with the proposed model are presented, and the proposed model is presented in Section 3. In Section 4, experimental results and discussions were described to evaluate the performance of the model. Finally, conclusions were drawn with some general observations and recommendations for ongoing work in Section 5.

2.   Related Works

Saliency detection technique using a visual attention mechanism is widely used in the fields of target detection [30, 31]. However, as mentioned in the previous section, the bottom-up model often does not find the desired target. Most conventional detecting models are based on training to detect a specific target by the difference of saliency in the local context of image [1].

The model in [8], a top-down attention model that searches and learns the optimal set of linear map weights for a given object in an image was proposed. The model in [32] used mixture information of bottom-up and top-down information, but this model has a limitation from the fact that these two kinds of information were combined with a fixed weight. In addition, if a shape feature (eg, a circle shape) is selected as top-down information, other objects of the same shape are difficult to find. In the model of [33], features were extracted by the method in [6] to search for the object, and the values of the most prominent part of each of the 42 feature maps of the training image were learned with the naive Bayesian network. When detecting a specific object, multiple features maps were filtered by applying the pre-learned subbands, and then center-surround operator was applied to feature maps to enhance the region which is much different to surroundings. And these multiple maps were simply multiplied to make one single saliency map. This model has a limitation that all values other than the learned feature values are lost and that the relationship between features cannot be maintained because all feature maps are multiplied when generating the saliency map. In [34], a top-down attention model for robot navigation was proposed. This model calculates the robot's position by learning Gist features and landmarks. The biggest limitation of this model is that it is practically impossible to apply because it has to actually visit the place and learn. In [35], a top-down attention model that can be applied to finding a person's face was proposed, but it needs to select the color of the clothes manually whenever a top-down saliency map is generated. The model in [36] extracted the attention area by inputting top-down information to the model in [6]. In this model, color, direction, and shape features were extracted, the candidate areas were compared to the similarity of the target, and the candidate areas were weighted according to the results of the similarity comparison. This model does not use learning processing, and only the comparison of one specified object by specifying the search object area in the image. Xiao [37] stored basic features such as lines, points, and circles that make up a target and extracted a target from the input image using them, but it has limitations in that the standard of features for representing objects is not accurately presented.

Recently, deep learning has been successfully utilized in target detection. Automatic feature extraction methods with convolutional neural networks combined with transfer learning achieved top level performance on saliency estimation [38-47] have been proposed. CNN (Convolutional Neural Networks) capture typical high-level features to detect salient objects that are prominent in a particular size and category and achieved an advanced performance for saliency detection issues. Though these deep learning models have shown preferable results, they extract features on special levels, and all levels of information are significant. It also needs a supervised learning process. The existing problem is how to choose the network layers, although each network layer is significant, nevertheless, full convolution network layers increase workload. The other problem is what kind of low-level features to integrate and how to integrate all level features into multiple resolutions. Moreover, these models do not produce a temporal

sequence of eye movements, which can be very important not only in developing a system that deals with video streams, but also in understanding human vision.

Although lots of models have been proposed, the fast and accurate detection of salient regions remains a challenge in target detection, particularly in cases of complex textures. Summarizing previous studies that explored objects using the top-down visual attention approach discussed so far, there are some limitations in the method of saliency map generation and performance evaluation. First, in most models, weight, color, intensity, and shape feature maps are integrated with the same weighted without considering the relative differences between features. In this case, if a specific feature value of the search object is changed in the search image, the search becomes difficult. Let's suppose that a red can has been trained, but when the model needs to detect a darker red can. When the top-down information is input, the intensity and shape features become more prominent than the color feature of the red can. Therefore, if there are cans of different colors of the same size in the search image if the intensity and shape features are the same, the color features are pushed out to other colors, so that the red cans cannot be easily found. Second, in the previous model, a saliency map was made by selecting only some feature maps from the extracted feature map such as intensity and color. Since humansHumans take all of the various features into account when paying attention, the method of considering only some features does not fully mimic human visual processing. Third, the kind of experimental images used in the performance evaluation experiments has been very monotonous. If the feature values of the training objects are similar when learning, other objects other than the target to be searched for may be found. Performance evaluation experiments were performed in many models using only experimental images contained only one trained object. Therefore, the performance of the interference between the trained objectscannot be evaluated.

In this paper, a model with the following characteristics is proposed in order to overcome the above-mentioned limitations of the existing model. To overcome the first described limitations, trained data is generated by training the relative differences between features. This trained data can be used in a detection model to detect targets more efficiently. To overcome the second described limitations, a saliency map that takes into account all the features is made. This eliminates the loss of features, allowing the proposed model to handle a wider variety of information, and also allow the model to have a structure similar to the human visual process. To overcome the third described limitation, experiments were performed on images containing a number of trained objects. Through these experiments, the proposed model can be shown to detect the desired target well in the image containing other trained objects.

3.   The Methodology

The proposed model was developed to effectively detect a distant target using the top-down information by expanding the bottom-up visual attention model proposed in [27]. To detect targets in the proposed model, the training process is required before searching for targets. Pre-trained data causes bias in the feature extraction and combining phases to find the desired target.

The proposed model consists of the training model and detection model as shown in Fig. 1. In the training model, training images for the target are input and processed to generate trained data. The training model has two processes, feature extraction, and training. Such early visual features such as color, intensity, and form-orientation are extracted from the input image, and information of each feature is selectively selected and trained, and then trained data is generated. For training, the naive Bayesian network was used. In the detection model, a target is detected using the trained data from an image containing the target to be detected. In the detection model, the early visual feature extraction and saliency map generation process goes through the same process as that of the training model. In the training model, the feature values for training are newly calculated and selected during this process, but in the detection model, weights that are calculated using pre-trained data are given to feature maps.
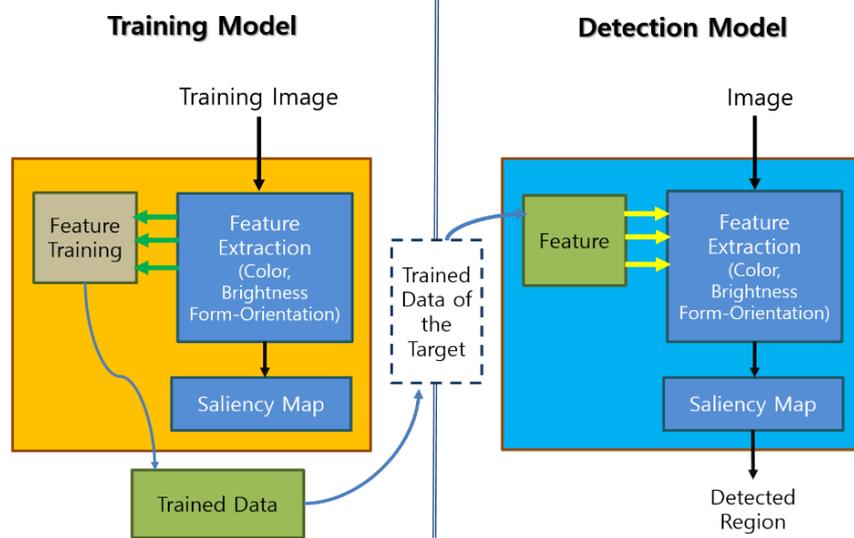
Fig. 1. The overall process of the proposed model

3.1.   Training Model

In the training model, the general features of the training image of the target are extracted, and the unique attributes of the target are trained.

Feature extraction and saliency map generation

The basic bottom-up process of extracting the early visual features and weighting them together to produce the saliency map proceeds in the same way as in the model of [27]. In the proposed training model, some of the features extracted in this process are selected and used for training. The overall process of extracting early visual features and integrating them into a saliency map by weighting them is shown in Fig. 2.Feature extraction and saliency map generation process in "Training" component can be described detail in 5 steps as follows.
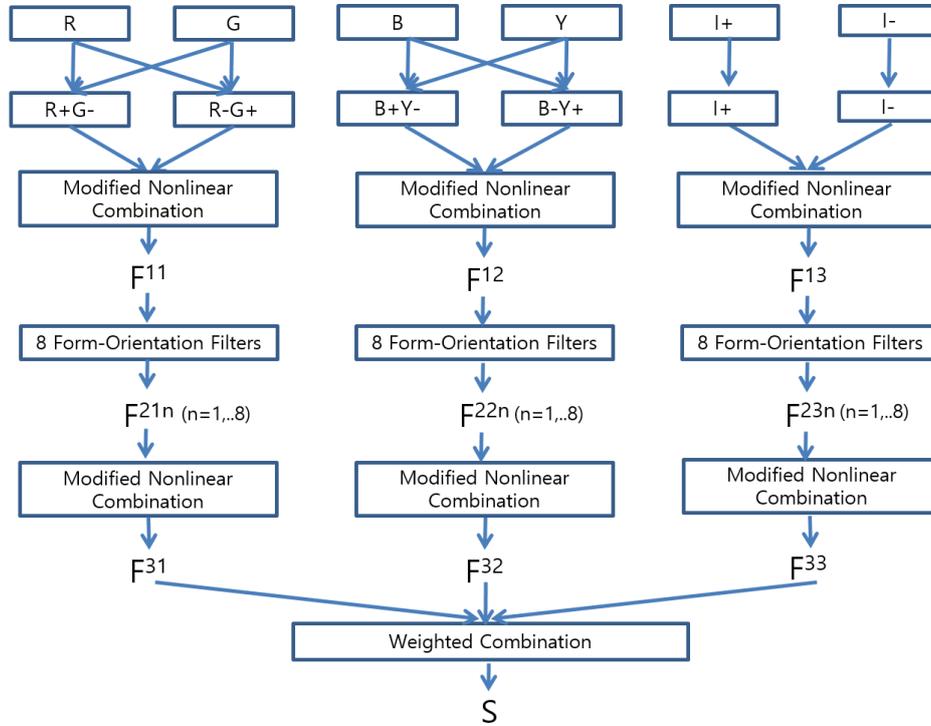
Fig. 2. Features extracted from the proposed training model

First, early visual features, R(red), G(green), B(blue), Y(yellow), I+(ON-intensity) and I-(OFF-intensity) are extracted by equation (1) from input image.

$$R = r - \frac{g+b}{2}, \quad G = g - \frac{r+g}{2}, \quad B = b - \frac{r+g}{2}$$
$$Y = r + g - (2(|r-g|+b)), \quad I+ = \frac{r+g+b}{3}, \quad I- = 1 - \frac{r+g+b}{3} \quad (1)$$

In equation (1), R is the red channel of the image, G is the green channel, and B is the blue channel. Since the intensity features seen by the human eye may have a high saliency either on the bright part of the image or on the dark part on the contrary, ON intensity features (I+) with high feature values for bright parts and OFF intensity features (I-) with high feature values for dark parts are generated.

Second, early visual features are then reorganized into R+G-, R-G+, B+Y-, B-Y+, which are relative color pairs for R/G, B/Y colors extracted based on an opponent-process theory of color vision [48]. These opposite features are generated by equation (2).

$$R + G- = R - G, R - G+ = G - R, \quad B + Y- = B - Y, B - Y+ = Y - B \quad (2)$$

Through an improved nonlinear combining method shown in equation (3), R+G- and R-G+ are combined into F^{11}, and B+Y- and B-Y+ are combined into F^{12}, and I+ and I- features are combined into F^{13}.

$$F_{x,y}^k = \frac{F_{x,y}^k - MinF}{MaxF - MinF},$$

$$F_{x,y}^k = F_{x,y}^k \times Diff(F^k), \qquad Diff(F^k) = (MaxF^k - AveLF^k)^2$$

$$MaxF^k = \max(F_{x,y}^k), AveLF^k = average[\text{local } \max(F_{x,y}^k)],$$

$$MaxF = \max(F_{x,y}^1, \dots, F_{x,y}^k), MinF = \min(F_{x,y}^1, \dots, F_{x,y}^k)$$

(3)

In equation (3), k is the number of imput map, and $Diff(F^k)$ is the relatvive activity value of F^k. At this point, 'activity' is a unit that indicates the intensity of attention of the input image, which means that the higher the amount of activity, the more noticeable features are included than the surroundings.

Third, three form-orientation feature maps (F^{21} for R/G color, F^{22} for B/Y color, F^{23} for intensity) are generated by extracting form-orientation features (F^{21n}, F^{22n}, F^{23n}, n=1~8) that have eight orientations from the extracted two color features (F^{11}, F^{12}) and intensity feature (F^{13}) respectively. Center-surround computations with 8 orientations ($0\pi/8$, $1\pi/8$, … $7\pi/8$) [27] are used as form-orientation filter.and combining them using an improved nonlinear combining method.

Fourth, achieved form-orientations are combined by nonlinear combination method into of F^{31}, F^{32}, and F^{33}.

Finally, saliency map (S) is generated by weighted combination method by equation (4). In each process, all input images are normalized between 0 and 1.In equation (5) k is the number of feature maps, W^k is the weight of feature map, and $Diff(F^k)$ is the relative activity value of F^k.

$$S = W_1 \times F^{31} + W_2 \times F^{32} + W_3 \times F^{33},$$

$$W_k = \frac{\sum_{i=1}^n Diff(F^i)}{\sum_{i=1}^n Diff(F^i) - Diff(F^k)}$$

(4)

Selected feature values for training

In the proposed training model, various feature values were involved in training to learn the unique features of the object. The feature values in the various feature maps generated from the training image and the relative feature values between the feature maps were all considered, and the following four types of feature values shown in Table 1 were calculated and participated in the training.

Table 1. Selected feature values for training

Selected feature value	Reason
R, G, B, Y, I +, I- feature map values corresponding to maximum saliency location of F^{11}, F^{12}, F^{13} feature map	To find out the intensities of the unique colors of the target.
Relative activity amount of each pair of R+ G-, R-G+, B+Y-, B-Y+, I+, and I- feature maps	To find out which features of the input map were reflected in creating F^{11}, F^{12}, and F^{13} feature maps.
Relative activity amount of each of F^{21n}, F^{22n}, F^{23n} feature maps.	To find out which features of the input map were reflected in creating F^{21}, F^{22}, F^{23} feature maps.
Weights given to the F^{21}, F^{22}, F^{23} feature maps used when generating the saliency map (S)	To find out which features of the input map were reflected in creating a saliency map.

Fig. 3 shows the example of a training feature selection mechanism considering the features in the various feature maps and Fig. 4. shows the example of a training feature selection mechanism considering the features between the various feature maps. In Fig. 3, R, G feature map values corresponding to maximum saliency location of F^{11} feature map were selected, and in Fig. 4, the relative activity amount of orientation feature maps of F^{21n} was selected.
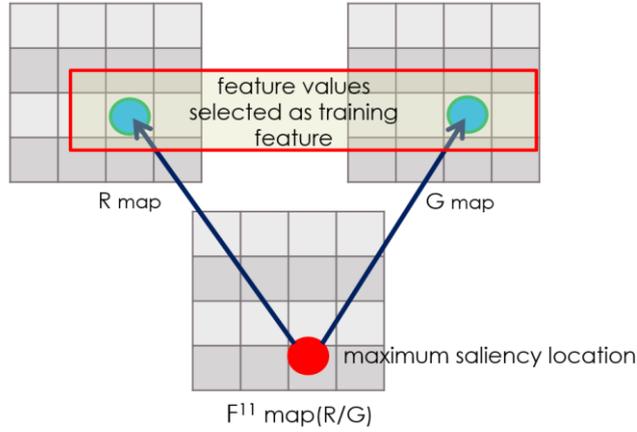
Fig. 3. The example of a training feature selection mechanism considering the features in the various feature maps
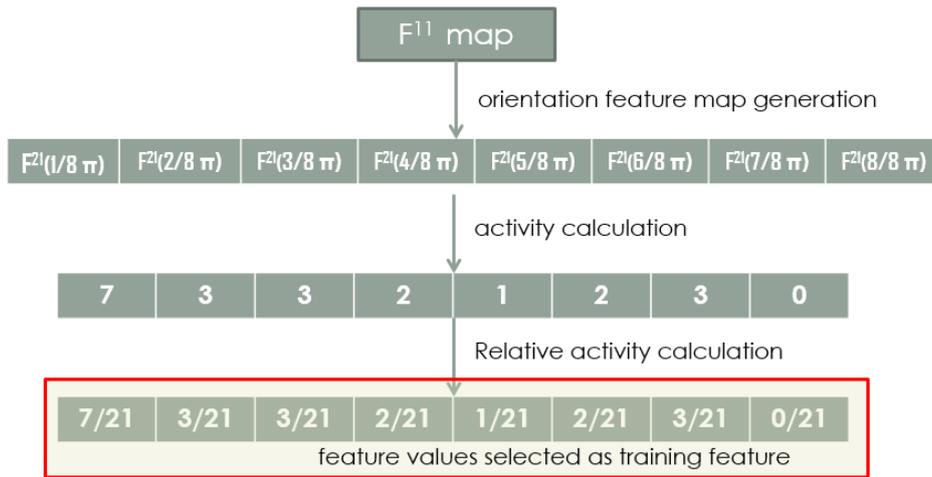
Fig. 4. The example of a training feature selection mechanism considering the features between the various feature maps

Trained Data

The features extracted and selected from each category's training images were stored as mean(μ) and standard deviation(σ^2) as shown in equation (5) through naive bayesian. In equation (1), is the value of input data and is the number of data input. In this way, a plurality of images is trained to construct trained data. The trained data is represented by a probability equation as shown in equation (6), where $N()$ is a normal distribution curve, $p()$ is a probability, and n is the total number of features to be trained. That is, the probability that the value 'θ' emerges from the feature 'F' is the result value when

'theta' is input into a normal distribution curve composed of the mean(μ) and standard deviation(σ^2) of the trained 'F'.

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i, \qquad \sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^2. \tag{5}$$

$$\prod_{j=1}^{n} p(F_j \mid \theta_j) \propto N(F_j; \mu_j; \sigma_j). \tag{6}$$

3.2.   Detection Model

In the detection model, trained data is used to detect the target object. The early visual feature extraction and saliency map generation process goes through the same process as that of the training model. In the training model, the feature values for training are newly calculated and selected during this process, but in the detection model, weights that are calculated using pre-trained data are given to feature maps.

Biasing the feature values of R, G, B, Y, I+, I-

In order to make the region corresponding to target in the extracted R, G, B, Y, I +, I- feature maps to have high feature values, high weight is assigned to each feature map that is similar to the maximum feature values of R, G, B, Y, I+, and I- of the trained data. The weighting method is as follows. First, a Gaussian distribution curve is drawn using the mean (μ) and standard deviation (σ^2) of the trained data. And then the values of the feature map are increased by the result value of passing each value of the feature map through a gaussian distribution curve sd shown in equation (7). In equation (7), x means the feature values in the feature map, and y means the value of the results when the x was inputted to the gaussian distribution curve. Feature value x is modified by adding the result value y with x itself. This adjusts the value of the feature similar to the target to a higher value.

$$y = gaussian(x), \qquad x = x + y \tag{7}$$

At this time, the value of the curve is adjusted to match the maximum value of the y-axis of the gaussian distribution curve with the maximum value of the feature map so that the feature values similar to the trained data in the output feature map should be the maximum feature values.

Biasing in nonlinear combination process for generation of F^{11}, F^{12}, and F^{13}

When top-down information is input to the R, G, B, Y, I+, and I- feature maps, F^{11}, F^{12}, and F^{13} are generated with an improved nonlinear combination. In the nonlinear combination process, additional activity amount of trained data is input to select the

salient feature map. Fig, 5 shows the basic mechanism of biasing activity values of feature maps. In Fig, 5, the biggest of the calculated activity values of the input maps is 15 of the first map, but in trained data, the third of the input maps has the largest value of 9. So, it is biased by the value of the third map of the input maps.
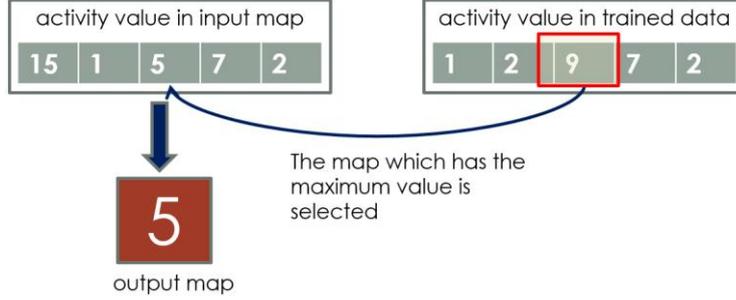
Fig. 5. The basic mechanism of biasing activity values of feature maps

The nonlinear combination reflecting the top-down information can be expressed as equation (8). In equation (8), C means input map, CN means a map with assigned weight, and T_a means activity amount of C map in the trained data.

$$CN_{x,y}^k = C_{x,y}^k \times T_a^k. \tag{8}$$

Biasing in nonlinear combination process for generation of $\mathbf{F^{31}}$, $\mathbf{F^{32}}$, and $\mathbf{F^{33}}$

The eight form-orientation feature maps F^{21n}, F^{22n}, F^{23n} are generated for each color and intensity feature by performing a center-surround operation that mimics the cellular reactivity seen in the "ON-centered, OFF-surround" receptive field of humans [5]. The eight form-orientation feature maps are integrated into the orientation with the largest response because one feature was divided into eight orientations. At this time, in the nonlinear combining process, the top-down information is input to the form-orientation activity of each direction in the trained data as the top-down information to generate an output map having the prominent direction of the target.

Biasing in weighted combination process for generation of the saliency map

When weighted combining is performed, the weight of trained data is added as top-down information to generate a saliency map that has the relationship between the color and intensity of the target. The saliency map is generated by equation (9) and L_1, L_2, L_3 is the weight corresponding to each map in the trained data. The maximum salient location of the saliency map generated by the weighted combination is the portion that matches the target.

$$S = L_1 \times F^{31} + L_2 \times F^{32} + L_3 \times F^{33}. \tag{9}$$

4.   Experiments and the results

To evaluate the performance of the model, the model was applied to the problem of detecting targets, such as a pen, triangle-shaped safety sign, and beverage can. In addition, to carry out a quantitative evaluation, the proposed model was compared with the model in [5], [7], and [26].

Most models that use top-down information are difficult to compare because of the lack of training and test images. However, because the model of [8] provides both training image, test image, and general evaluation criteria, it is possible to compare the performance with the proposed model. The proposed model was also compared with previous bottom-up models in [6] and [27]. The model of [6] is the most frequently referred to in the bottom-up model studies, and it was selected for the performance comparison with other models in the future. The model in [27] was selected because it has the same feature extraction method as the proposed model, but not use top-down information.

4.1.   Training images

For 3 types of detection experiments and robustness experiments of color and brightness, images of red, blue, and black pens with strong color and orientation features commonly found in the real-world were created as training images. In addition, 32 triangle-shaped safety signs and 45 red beverage cans taken at random campus locations and times were used as training images. The training image was made by 8 angular changes for one training image for one object, and 12 brightness changes and 10 step size changes for each angular change. The brightness change was changed in 5 steps from -30 to +30, and the change in size was changed in 0.2 steps from 1.2 to 3.0 times. In summary, as shown in Table 2, a total of 176 training images was created for the training image for one object.

Table 2. Number of newly Selected feature values for training

composition of training images for each object	variation1		variation2		Total No.
	rotation	8	brightness	12	96
			size	10	80
					176

4.2.   Test images

Test images used in specific pen detection experiments. Test images used in the experiment for detecting a specific pen is a scene containing the one, two, and three red, blue, and black pens that are trained as shown in Table 3. A total of 147 test images were used. Test images used in triangular-shaped safety signs and beverage can detection experiments. 32 triangular-shaped safety signs and 59 red cans were used as test images in triangular-shaped safety signs and beverage can detection experiments. An example of the test image used in the detection experiment is shown in Fig. 6.

Table 3. Test image used in specific pen detection experiments

	1 trained object	2 trained object	3 trained object	Total
Total	63	63	21	147

Test images used in color robustness experiments

The robustness test for color is to see how the color of distractors and the background of the scene with the object effects the target detection. The images with a red, blue, gray, sky blue, and brown background and the images with red, blue, and green distractors among images used in specific pen detection experiments were used as the Test images used in color robustness experiments. A total of 100 images were used.

Test images used in brightness robustness experiments

The robustness test for brightness is to check the effect of brightness change on target detection. 294 images with brightness variation from +30 to –30 applied to the original test images used in specific pen detection experiments were used as test images in brightness robustness experiments.
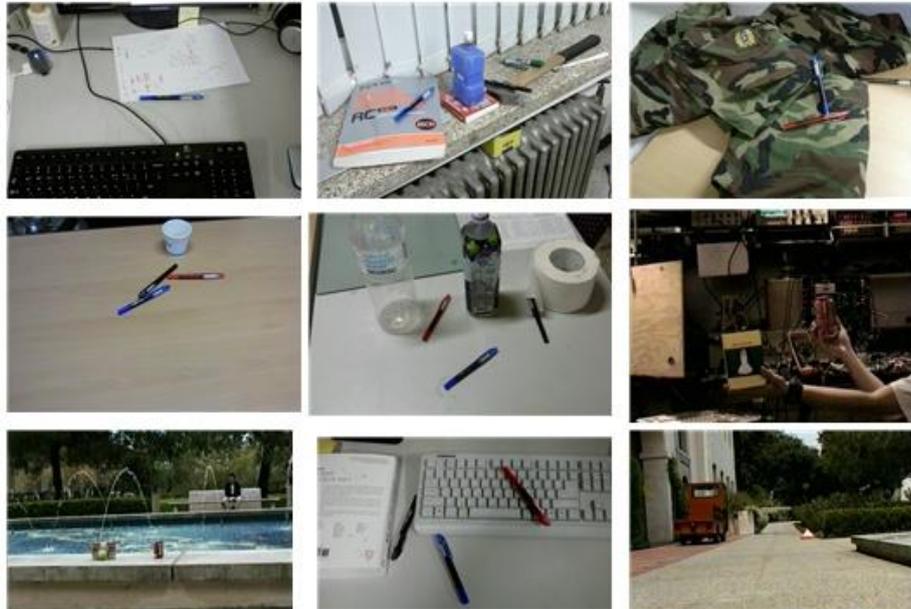
Fig. 6. An example of the test image used in the detection experiment

4.3.   Performance Evaluation Criteria

Two performance evaluation criteria, i) The successful target detection rate (p) using equation (10) for the images detected within the specified number of times in the test image, ii) the average viewpoint variation number until the target found, were used.

$$p = \frac{r}{r + q} \tag{10}$$

In equation (10), r is the number of images that successfully detected the target, q is the number of images that failed at detecting the target. If a model detects a target in 3 images out of 5 test images, the successful target detection rate (p) is 0.6. Among successfully detected images, if this model detected the target in the third search from the first image, the third search from the second image, the second-search in the third image, then the average viewpoint variation number until target found is 2.6. The successful target detection rate (p) can be used to evaluate the detection efficiency in the entire test image, and the average viewpoint variation number until target found can be used to evaluate how fast the model detected the target. The higher the successful target detection rate, the lower the viewpoint variation number until the target found, the better the performance.

4.4.   Results

Fig. 7 shows an example of the result of a saliency map of the proposed model using top-down information and the result that is not. The target to be detected in the input image is a blue pen. Without using trained data, with only bottom-up features, the proposed model detects the red pen at first-search. The saliency map on the left side of Fig. 7 shows that the red pen is the most salient object. But what if we want to detect a blue pen, not a red pen? The proposed model allows you to find blue pens at once. The saliency map on the right side of Fig. 7 was made using trained data, showing that the blue pen, not the red one, is the most salient object.
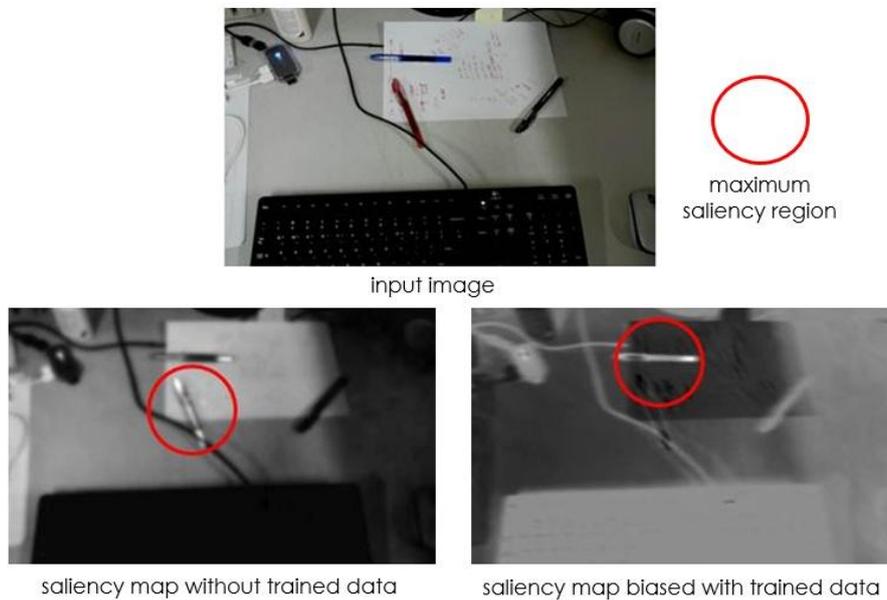
Fig. 7. Example of the result of the saliency map without top-down information (left) and the result with top-down information (right)

Fig. 8 demonstrates the example of detection results for 3 types of target detection experiments in the proposed method. The three images on the left side of Fig. 8 are examples of results from specific pen detection experiments. The image on the first line is an image where two trained objects (red and blue pens) are placed on a bookshelf with several books. The target is a blue pen. The proposed model detected the target after the second-search but the other models failed at detecting the target within the specified number of search-times. The image in the second row on the left of Fig. 8 shows a yellow alarm clock on a light brown table, two trained objects (a red pen and a blue pen), and a gray digital clock. The target is a blue pen. The proposed model and the models of [6] and [8] detected the target at once, but the model of [27] failed at detecting the target within the specified number of search-times. The image in the third row on the left of Fig. 8 is a test image with a military uniform on a brown table and three trained objects (red pen, blue pen, and black pen) placed on the uniform. The

target is a red pen. The proposed model detected the target at once, but the other models failed at detecting the target within the specified number of search-times.

The three images in the middle of Fig. 8 are examples of the results from triangular-shaped safety sign detection experiments, and the three images on the right are examples of the results from beverage can detection experiments. We can see that the proposed model detects the target in various environments in the first-search.
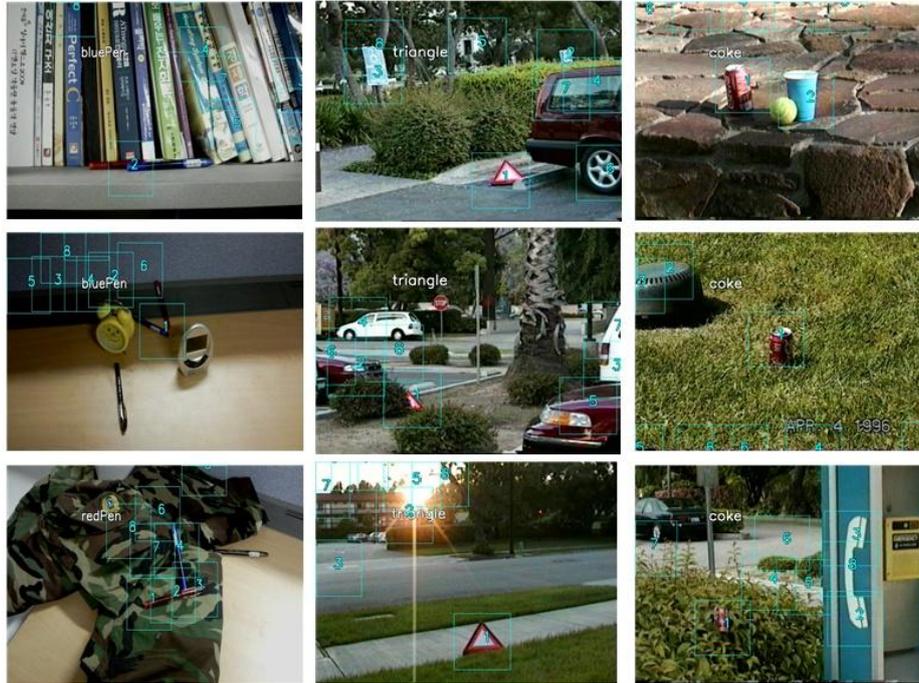
Fig. 8. Example results on 3 types of detection experiments: pen detection (left), triangular-shaped safety sign detection (middle), beverage can detection (right)

Table 4 and Table 5 summarizes the performance evaluation result of the proposed model with the comparison result of 3 models. Overall, the proposed model achieved the successful target detection rate with an average of 94.33% on 3 types of target detection experiments, and 97% on color and brightness robustness experiments. An average viewpoint variation number until the target found from the proposed model is 1.37 on 3 types of target detection experiments, and 1.33 on color and brightness robustness experiments. And these results are the best performance among other comparison models.

The proposed model detected the target almost at first-search, but other models didn't. The model in [8] has the best performance except for the proposed model, and it detected the target on average 1.67th search on 3 types of target detection experiments and 1.87th search on color and brightness robustness experiments. However, the proposed model has better performance with 1.37 on 3 types of target detection and 1.33 on color and brightness robustness experiments. Even the successful target detection rate is superior to that of [8].

These results demonstrate that the proposed model outperforms other previous models. The color robustness test and the brightness robustness test were experiments to confirm how robust the proposed model was for the brightness variation and the disturbing color. Previous models were difficult to secure versatility only by experimenting on the images of the monotonous environment. The color robustness test results and the brightness robustness test results demonstrate that the proposed model is very robust to color and brightness. In particular, even if the background color of the test image is similar to the target and there are some distractors, the performance of the model did not deteriorate much.

Table 4. Overall results on 3 types of target detection experiments

	The successful target detection rate				Viewpoint variation number until the target found			
	This Model	[8]	[27]	[6]	This Model	[8]	[27]	[6]
Pen	89	75	67	60	1.21	1.8	3.57	4.31
Safty sign	99	90	79	75	1.59	1.8	4.02	5.6
Beverage can	95	89	82	67	1.31	1.35	5.3	6.5
average	94.33	84.67	76	67.33	1.37	1.67	4.3	5.47

Table 5. Results on color and brightness robustness experiments

		The successful target detection rate				viewpoint variation number until the target found			
		This Model	[8]	[27]	[6]	This Model	[8]	[27]	[6]
Color	Red background	100	100	92	75	1	1.25	2.5	2.13
	Bluebackground	100	100	83	50	1	1.40	2.2	2.80
	Gray background	100	100	75	67	1	1.14	3.0	2.28
	Sky blue Background	100	92	83	83	1	2.25	2.38	2.25
	Brown background	100	92	75	75	1	1.43	2.5	2.29
	Red distractor	100	100	83	75	1.14	2.14	3.71	3.29
	Blue distractor	100	100	100	100	1.67	1.83	4.08	3.42
	Green distractor	100	100	92	83	1.11	1.67	2.5	3.11
Bright-ness	+30	89	75	65	63	2.06	2.9	4.15	4.12
	-30	91	73	60	61	2.27	2.73	4.95	4.19
average		97	93.2	80.8	73.2	1.33	1.87	3.20	2.99

5. Conclusions

In this paper, a robust distant target detection model employing the human visual attention mechanism was proposed. The proposed model consisted of a training model for training top-down information and a searching model for detecting targets using top-down information. In the training model, some primitive features of training objects extracted in a bottom-up manner were selected and trained. The detection model biases feature maps and adjust the saliency map to detect desired targets. In the process of detecting, the bottom-up saliency and top-down saliency were both considered. The proposed model trained the relationship information between color, intensity, form-orientation features in order to overcome the limitations of the previous training model that could not accurately express the features of an object because of the relationship between color and brightness was not considered. Also, all training images with changes under certain criteria were used for training in order to overcome the limitations of the previous model which did not produce reliable results by selectively using the training images. The proposed model detected the target using all features in order to overcome the limitations of previous detection models which used only a few features to detect targets which could cause detection failures due to loss of information.

The entire process of the proposed detection model is similar to that of the training model, but instead of training selected features, top-down information was input and biased. Top-down information is input in three phases: first, the color and brightness values of the trained data were input in the process of generating early visual feature maps of color and intensity. Among the early feature values of the color and brightness feature maps, the values similar to the color and brightness values in the trained data were modified to be more salient. In addition, the activity information of the trained data is used in a nonlinear combination process, and it leads to making features of the target objects more salient. Second, relative activity values of feature maps in the trained data were input in the nonlinear combination process of form-orientation feature maps. The detailed process of the nonlinear combination method was modified by trained data. Finally, the weights of the form-orientation feature maps in the trained data were input in the weighted combination process.

To evaluate the performance of the proposed model, experiments were conducted to apply the proposed model to detection problems such as detecting a specific pen, triangular safety objects, and can. Also, color robustness and brightness robustness experiments were additionally performed. To evaluate the performance of the proposed model, experiments were conducted to apply the proposed model to detection problems such as detecting a specific pen, triangular safety objects, and can. Also, color robustness and brightness robustness experiments were additionally performed. In addition, to carry out a quantitative evaluation, the proposed model was compared with the previous model. The previous model only used the images contains only one trained object so that they failed to evaluate performance considering the effects between different trained objects. In this experiment, to overcome these limitations of the previous model, an image containing several trained objects was used as the experimental image. The color robustness test and the brightness robustness test are experiments to confirm how robust the proposed model is for the brightness variation and the disturbing color. Existing models were difficult to secure versatility only by experimenting on the images of the monotonous environment. These experiments were conducted to confirm whether the proposed model overcomes the limitations of these

existing models. Quantitative and qualitative analyses of the experimental results showed that the proposed model outperformed the traditional model in target detection and was comparable to the state-of-the-art models.

References

1. Borji, A. Itti, L.: State-of-the-art in visual attention modeling. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 35, No. 1, 185–207. (2013)
2. Bi, Z., Dou, S., liu, Z., Li, Y.: A Recommendations Model with Multiaspect Awareness and Hierarchical User-Product Attention Mechanisms. Computer Science and Information Systems, Vol. 17, No. 3, 849–865. (2020)
3. Paliwal, N., Vanjani, P., Liu, J., Saini, S., Sharma, A.: Image processing-based intelligent robotic system for assistance of agricultural crops. International Journal of Social and Humanistic Computing. Vol. 3, No. 2, 191-204. (2019)
4. Treisman A.M., Gelad, G.: A Feature-integration Theory of Attention. Cognitive Psychology, Vol. 12, No. 1, 97-136. (1980)
5. Koch, C., Ullman, S.: Shifts in selective visual attention: towards the underlying neural circuitry. Human Neurobiology, Vol. 4, 219-227. (1985)
6. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, 1254–1259. (1998)
7. Itti, L., Koch, C.: Computational modeling of visual attention. Nature Reviews, Neuroscience, Vol. 2, 194–203. (2001)
8 Itti, L., Koch, C.: Feature combination strategies for saliency-based visual attention systems. Journal of Electronic Imaging, Vol. 10, No. 1, 161-169. (2001)
9. Dhavale, N., Itti, L.: Saliency-based multifoveated MPEG Compression, In Proceedings of IEEE International Symposium on Signal Processing and its Applications, Paris, France, France,229-232. (2003)
10. Yun, Z., Shah. M.: Visual Attention Detection in Video Sequences Using Spatiotemporal Cues. In Proceedings of 14th annual ACM international conference on Multimedia, Santa Barbara, CA, USA,815-824. (2006)
11. Torralba, A., Oliva, A., Castelhano, M. S., Henderson, J. M.: Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. Psychological Review, 113, 766–786. (2006)
12. Li, S., Lee, M.: An Efficient Spatiotemporal Attention Model and Its Application to Shot Matching. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 17, No. 10, 1383-1387. (2007)
13. Li, H., Su, X., Wang, J., Kan, H., Han, T., Zeng, Y., Chai, X.: Image processing strategies based on saliency segmentation for object recognition under simulated prosthetic vision. Artificial Intelligence in Medicine, Vol. 84, 64-78. (2018)
14. Li, H., Han, T., Wang, J., Lu, Z., Cao, X., Chen, Y., Li, L., Zhou, C., Chai, X.: A real-time image optimization strategy based on global saliency detection for artificial retinal prostheses, Information Sciences 415–416, 1-18. (2017)
15. Lei, J., Wang, B., Fang, Y., Lin, W., Callet, P.L., Ling, M., Hou, C.: A universal framework for salient object detection. IEEE Transactions on Multimedia, Vol. 18, No. 9, 1783–1795 (2016).
16. Wang, Z., Xiang, D., Hou, S., Wu, F., Background-driven salient object detection. IEEE Transactions on Multimedia, Vol. 19, No. 4, 750–762. (2017).
17. Park, M., Cheoi, K.: Selective Visual Attention System Based on Spatiotemporal Features, Lecture Notes in Computer Science, Vol. 5068, 203-212. (2008)

18. Cheoi, K., Park, M.: Visual Information Selection Mechanism Based on Human Visual Attention. Journal of Korea Multimedia Society, Vol. 14, No. 3, 378-391. (2011)
19. Chen, C., Li, S., Wang, Y., Qin, H., Hao, A. Video saliency detection via spatial-temporal fusion and low-rank coherency diffusion. IEEE Transactions on Image Processing, Vol. 26, No. 7, 3156–3170. (2017).
20. Xi, T., Zhao, W., Wang, H., Lin, W.: Salient object detection with spatiotemporal background priors for video. IEEE Transactions on Image Processing, Vol. 26, No. 7, 3425–3436. (2017).
21. Liu, Z., Li, J., Ye, L., Sun, G., Shen, L. Saliency detection for unconstrained videos using superpixel-level graph and spatiotemporal propagation. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 27, No. 12, 2527-2542. (2017).
22. Wang, W., Shen, J., Yang, R., Porikli, F.: A unified spatiotemporal prior based on geodesic distance for video object segmentation. IEEE Transactions on Pattern Analasis and Machine Intelligence, Vol. 40, No. 1, 20–33. (2018).
23. Wang J., Silva M., Callet P., Ricordel V.: Computational model of stereoscopic 3D visual saliency. IEEE Transactions on Image Processing, Vol. 22, No. 6, 2151–2165. (2013)
24. Feng, D., Barnes, N., You, S., McCarthy, C.: Local background enclosure for RGB-D salient object detection. In Proceedings of International Conference on Computer Vision and Pattern Recognition, 2343–2350. (2016).
25. Song, H., Liu, Z., Du, H., Sun, G.: Depth-aware saliency detection using discriminative saliency fusion. In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 1626–1630. (2016).
26. Wang, A., Wang, M.: RGB-D salient object detection via minimum barrier distance transform and saliency fusion. IEEE Signal Processing Letters, Vol. 24, No. 5, 663–667. (2017).
27. Cheoi, K., Kim, M.: Adaptive Spatiotemporal Feature Extraction and Dynamic Combining Methods for Selective Visual Attention System. Wireless Pers. Commun. Vol. 98, 3227–3243. (2018).
28. Oliva, A., Torralba, A., Castelhano, M., Henderson, J.: Top-down control of visual attention in object detection. Proceedings of International Conference on Image Processing (pp. 253–256).Barcelona, Catalonia: IEEE Press. (2003).
29. Torralba, A., Oliva, A., Castelhano, M., Henderson, J.: Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. Psychological Review, 113, 766-786. (2006)
30. Borji, A., Cheng, M., Hou, Q., Jiang H., Li, J.: Salient object detection: A survey. Computational Visual Media. Vol. 5, No. 7, 117-250. (2019)
31. Ren, Z., Gao, S., Chia, L., Tsan,g U.: Region-based saliency detection and its application in object recognition. IEEE Transactions on Circuits & Systems for Video Technology, Vol. 24, No. 5, pp. 769-779. (2013)
32. Yu, Y., Mann, G., Gosine, R.: An Object-Based Visual Attention Model for Robotic Applications. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), Vol. 40, No. 5, 1398-1412. (2010)
33. Elazary, L., Itti, L.: A Bayesian model for efficient visual search and recognition. Vision Research, Vol. 50, No. 14, 1338-1352. (2010)
34. Siagian, C., Itti, L.: Biologically Inspired Mobile Robot Vision Localization. IEEE Transactions on Robotics, Vol. 25, No. 4, 861-873. (2009)
35. Lee, K., Buxton, H., Feng, J.: Cue-guided search: a computational model of selective attention. IEEE Transactions on Neural Networks, Vol. 16, No. 4. (2005)
36. Zhang, J. Li, Z., Jingjing, G., Zhixing, L.: A Study of Top-down Visual Attention Model Based on Similarity Distance. In Proceedings of 2nd International Congress on Image and Signal Processing, Tianjin, China, 1-5. (2009)
37. Xiao, J., Cai, C., Ding, M., Zhou, C.: The Application of Novel Target Region Extraction Model Based on Object-accumulated Visual Attention Mechanism. In Proceedings of Fourth International Conference on Natural Computation, Jinan, China, 116-120. (2008)

38. Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N., Tang, X., Shum, H.: Learning to Detect a Salient Object. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33, No. 2, 353–367. (2011)
39. Li, G., Yu, Y.: Visual saliency based on multiscale deep features. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 5455-5463. (2015)
40. Wang, L., Lu, H., Ruan, X., Yang, M.: Deep networks for saliency detection via local estimation and global search. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 3183-3192. (2015)
41. Li, X., Zhao, L., Wei, L., Yang, M., Wu, F., Zhuang, Y., Ling, H., Wang, J: DeepSaliency: Multi-Task deep neural network model for salient object detection. IEEE Transactions on Image Processing. Vol, 25, No. 8, 3919-3930. (2016)
42. Wang, L., Lu, H., Zhang, P., Ruan, X.: Saliency Detection with Recurrent Fully Convolutional Networks. In European Conference on Computer Vision, 825-841. (2016)
43. Zhang, P., Zhuo, T., Huang, W., Chen, K., Kankanhalli, M.: Online object tracking based on CNN with spatial-temporal saliency guided sampling. Neurocomputing. Vol. 257, 115-127. (2017)
44. Zhang, J., Li, B., Y. Dai, F. Porikli, He, M.: Integrated deep and shallow networks for salient object detection. In Proceedings of IEEE International Conference on Image Processing, 271–276. (2017)
45. Zhang, P., Wang, D., Lu, H., Wang, H., Yin, B: Learning uncertain convolutional features for accurate saliency detection. In Proceedings of Interational Conference on Computer Vision, 212–221. (2017)
46. Wang, L., Lu, H., Wang, Y., Feng, M., Wang, D., Yin, B., Ruan,X.: Learning to detect salient objects with image-level supervision. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 3796–3806. (2017)
47. Zhang, J., Zhang, T., Dai, Y., Harandi, M., Hartley, R.: Deep unsupervised saliency detection: A multiple noisy labeling perspective. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition,9029–9038. (2018)
48. Hurvich, L., Jameson, D.: An opponent-process theory of color vision. Psychological Review, Vol. 64, No. 6, 384-404. (1957)

Kyung Joo Cheoi received Ph.D. degrees in Computer Science from Yonsei University, Korea in 2002. During 2002-2005, she worked as a research engineer in TI-specialist-Tech. of LG CNS, Korea. She is currently an associate professor of the Department of Computer Science in Chungbuk National University. Her research interests include computer vision, image processing, and machine learning.

Jaepil Ko received Ph.D. degree in Computer Science from Yonsei University, Korea in 2004. He is currently a faculty member of computer engineering department at Kumoh National Institute of Technology. He has served as a director of the Artificial Intelligence Society of KIISE. His research interests include pattern recognition, computer vision and image processing.

Exploring the Effectiveness of Deep Neural Networks with Technical Analysis Applied to Stock Market Prediction

Ming-Che Lee[1], Jia-Wei Chang[2,*], Jason C. Hung[2], and
Bae-Ling Chen[3]

[1] Department of Computer and Communication Engineering,
Ming Chuan University, Taoyuan City, Taiwan
leemc@mail.mcu.edu.tw
[2] Department of Computer Science and Information Engineering,
National Taichung University of Science and Technology, Taichung City, Taiwan
jiaweichang.gary@gmail.com, jhung@gm.nutc.edu.tw
[3] College of Intelligence,
National Taichung University of Science and Technology, Taichung City, Taiwan
chenbl@nutc.edu.tw

Abstract. The sustainable development of the national economy depends on the continuous growth and growth of the capital market, and the stock market is an important factor of the capital market. The growth of the stock market can generate a huge positive force for the country's economic strength, and the steady growth of the stock market also plays a pivotal role in the overall economic pulsation and is very helpful to the country's high economic development. There are different views on whether the technical analysis of the stock market is efficient. This study aims to explore the feasibility and efficiency of using deep network and technical analysis indicators to estimate short-term price movements of stocks. The subject of this study is TWSE 0050, which is the most traded ETF in Taiwan's stock exchange, and the experimental transaction range is 2017/01 ~ 2019 Q3. A four layer Long Short-Term Memory (LSTM) model was constructed. This research uses well-known technical indicators such as the KD, RSI, BIAS, Williams% R, and MACD, combined with the opening price, closing price, daily high and low prices, etc., to predict the trend of stock prices. The results show that the combination of technical indicators and the LSTM deep network model can achieve 83.6% accuracy in the three categories of rise, fall, and flatness.

Keywords: deep neural network, long short-term memory, technical analysis, fintech.

1. Introduction

How to make good use of the funds at hand for investment and financial management and financial planning is often the most concerned topic for modern people. Common investment and financial management methods include the purchase and sale of derivative financial commodities such as stock trading, funds, futures, and options, foreign currency investment, fund investment, and insurance planning. The variety of

---

* Corresponding author

financial commodities in the market has different characteristics, and investment returns and risks also vary. In recent years, the popularity of financial digitization and the vigorous development of artificial intelligence have also driven the future trend of mobile finance and new types of Financial Technology (FinTech). The development of the stock market has been under the influence of liberalization and internationalization. With the advancement of Internet technology and the growing popularity of financial knowledge, stock investment has become a part of investment and management in life [1], [2]. Investors' funds will not be confined to this class of stocks, and the spread of buying and selling stocks to earn changes in stock prices is also one of the favorite operations of retail investors [3]. Investment and financial management are closely related to the pulse of social development, and it is also the topic of most concern for young people entering the society. However, investors who do not have professional financial background or knowledge may be vulnerable to losses due to opaque market information [4], [5]. At the same time, due to the huge market information, the variety of financial products and the diversity of technical analysis indicators, novice investment and financial management faced with a wealth of information and could not absorb and judge [6], [7].

In recent years, Deep Neural Network [8] and various Deep Learning algorithms have shined in the major competitions of Pattern Recognition and Machine Learning. The rapid development of deep neural networks has not only opened up new areas of machine learning research, but also various applications have gradually appeared around people's lives, such as speech recognition, emotion recognition, natural language processing and image recognition [9], [10], [11]. In the field of deep learning, Recurrent Neural Network (RNN) [12], Long Short Term Memory (LSTM) [13] are particularly suitable for processing time-series data, such as natural language processing, machine translation, speech recognition, and financial index prediction. At present, the mainstream financial commodity analysis and decision tools are mainly based on fundamental analysis, chip analysis and technical index analysis. The main purpose of this research is to explore the feasibility and effectiveness of the application of technical analysis indicators in deep networks. The remainder of this paper is as follows: section 2 is the literature and techniques review; section 3 is the methodology of this research; section 4 is the experimental design, results and discussion, and the last section is conclusion and future research.

2.   Literature Review

Stock analysis tools can be divided into 'Fundamental Analysis' and 'Technical Indicator Analysis' in essence. Fundamental analysis is a method of valuation of securities or stocks, which uses financial analysis and economics research to evaluate corporate value or predict securities (such as stocks or bonds) [14], [15].

The fundamental analysis is to study the reasons for price changes, including economic factors, non-economic factors, internal market factors, current industrial conditions, domestic and foreign economic conditions, etc [16], [17], [18]. The data to be collected for the fundamental analysis is huge, and not every relationship with the stock price is equally important. There are many parts that need to be judged by individuals. In addition, some information may be hidden by the company. If investors do not have strong financial and economic analysis capabilities and internal

information, they may suffer losses [19]. Technical analysis, also known as trend analysis or market analysis, is to analyze the data of past prices and trading volumes and convert them into graphs or indicators. It uses statistical methods to analyze historical data to predict future prices [18], [20]. From the perspective of technical analysis, changes in stock prices and trading volume will affect the behavioral decision-making patterns of investors [21]. As long as the changes in stock prices and trading volumes are used to predict trends, excess returns will be obtained. Technical analysis mostly shows the behavior of investors in the past with graphics, and analyzes the past behavior of investors to predict the future trend of the market [22], [23], [24]. The basic theory of technical analysis is that stock price fluctuations and trends change mainly from market supply and demand, and are determined by the transaction behavior of all investors after integrating all relevant economic factors and information, resulting in stocks rising or falling to form a trend [25]. Based on past experience, the stock market or stock price usually leads the economic fundamentals by half a year to nine months, that is, the stock market or stock will rise or fall first, and then economic data will appear [26]. Based on this, researchers believe that technical analysis not only has the function of leading indicators, but also reflects the future trend of stock prices. Investment stocks can select individual stocks that have the potential to rise as long as the technical indicators are used properly, and sell at a high point of the swing to increase capital returns [27].

Technical analysis originated from "Dow Theory" published by Dow in 1930 [28]. Dow Theory assumes that all information will be reflected in stock prices. The main method of Dow theory is to divide market fluctuations into three trends according to the time period-long-term trends that last for 1 to several years, secondary movements that last for weeks to months, and short-term fluctuations at the intraday level. Based on the characteristics of "high point refresh, low point rise", it is judged as a bull market (or a bear market on the contrary) as a basis for buying and selling to capture the long-term trend of the market [29]. Then Elliott 1871 [30] proposed the "Elliott wave principle" to further improve the entire technical analysis system. The theory believes that the stock market behaves like the waves of the sea will rise and fall. There should be five upward waves in a complete cycle and three downward waves.

Subsequently Malkiel [31] proposed the efficient market hypothesis, assuming that the stock market is an efficient market, all the stock market information is immediately and completely reflected on the stock price, and it is concluded that the technical analysis of the stock market is invalid. Malkiel's [32] empirical research indicates that the return on the use of technical analysis is inferior to the buy and hold trading strategy. Hudson et al. [33] used empirical models to study the UK stock price from 1935 to 1994, and the results showed that there was no excess return. However, in subsequent research, Bohan used the S&P 500 weekly data to apply RSI stock market technical indicators to prove the validity of technical analysis [34]. Hinich and Peterson pointing out that the stock price series exhibits a non-linear change, and empirical evidence shows that the Moving Average (MA) technical indicator has a significant performance in predicting the US Dow Jones Industrial Average Index (DJIA) [35]. Pruitt [36] combined three technical indicators: cumulative trading volume, relative strength and weakness, and moving averages to develop an investment strategy called CRISMA (Cumulative Volume Relative Strength Moving 10 Average). The research period and objects were 1976 to 1985. For 204 stocks of the years, empirical results found that its trading strategy is better than the buy and hold strategy. Lo, Mamaysky,

and Wang [37] automatically identify patterns in technical analysis using non-major kernel regression methods. During the study period and objects were the stock prices of NYSE / AMEX and NASDAQ in the United States from 1962 to 1996. Empirical evidence shows that according to the technical analysis, the proposed pattern-based approach can get excess returns and beat buy-and-hold strategies. In addition, recent studies have shown that technical analysis indicators can obtain excess returns compared to the broader market in investment trading strategies [38], [39], [40], [41], [42].

In recent years, due to improvements in deep learning algorithms, computing efficiency, and excellent fault tolerance, deep neural networks have been widely used in the fields of stock price prediction and many derivative financial commodities such as futures, options and even house prices. Yu et al. [43] used a hybrid AI neural model combined with WTM text exploration technology to predict monthly WTI crude oil prices from 2000 to 2002 by collecting specific vocabulary and monthly WTI crude oil closing prices in monthly news from 1970 to 1999. The result confirms that the prediction accuracy in the case of ANN alone is 61%, while the accuracy of the hybrid AI neural model can reach 80%. Zhuge et al. [44] used semantic analysis to substitute the data of the community's opinions and comments on the Internet into the LSTM model for prediction. Compared with the traditional time series model and ANN model, the prediction of the Shanghai Stock Exchange's comprehensive stock price index has better accuracy. Nelson et al. [45] used the LSTM model to predict stock prices on the Brazilian Stock Exchange. The study period was from 2008 to 2015, and compared with traditional models and strategies, and found that the LSTM model has higher returns and lower risks. Soon [46] compared the performance of Feed-forward Neural Network (FFNN) and RNN when predicting the closing price of Commerce International Merchant Bank (CIMB) Kuala Lumpur Stock Exchange (KLSE), and its model used opening price, closing price and exchange rate information. The results show that both FFNN and RNN models can reach 90% prediction accuracy. Chen and Wei [47] proposed a Convolutional Neural Network (CNN) prediction model based on a company relationship graph. The experimental data uses the CSI 300 Index (CSI 300), which is close to 3,000 companies, traded between April 29, 2017 and December 31, 2017, with an average accuracy of approximately 60%. Many subsequent studies have also shown that machine learning and deep neural networks have superior performance when applied to the trend of time series data such as stock market trends [48], [49], [50], [51], [52], [53], [54], [55].

3.   Methodology

3.1.   Recurrent Neural Network, Long Short-Term Memory and System Model

General neural networks, such as Deep Convolutional Neural Network (DCNN) [56], processed samples are Independent and Identically Distributed (IID), and the problem solved is also a classification problem, a regression problem or a feature expression problem. However, more real problems are not satisfying IID, such as language translation, automatic text generation. They are a sequence of problems, including time series and spatial sequences. Compared to DCNN, sequence data or time series data is

more suitable for processing with Recurrent Neural Network (RNN) [57]. The reason that RNN is recurrent is that it performs the same operation on every element of a sequence, and the subsequent output depends on the previous calculation.
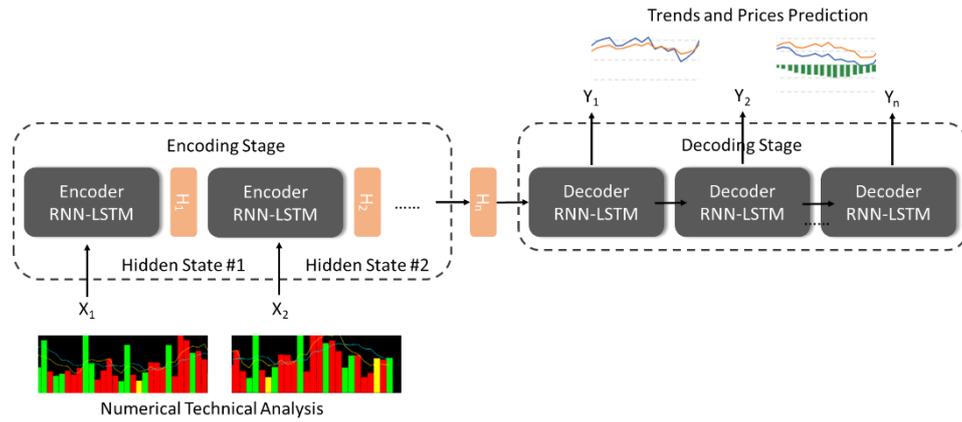
Fig. 1. Technical analysis and the proposed RNN-LSTM model

Another way to look at RNN is to think that it has some "Memory" that captures some of the previously calculated information. Long Short-Term Memory (LSTM) [58] is a special RNN, mainly to solve the gradient disappearance and gradient explosion problems in long sequence training. Because LSTM has the characteristics of remembering long-term trends and forgetting short-term fluctuations and errors, and can handle non-linear function problems, it is quite suitable for forecasting non-linear events that are easily affected by investors' mentality. This research uses LSTM neural network as the target stock price prediction model, the proposed model is shown as Fig. 1.

In our architecture, the LSTM is composed of two sets of RNNs, Encoder and Decoder. The input of the network is the filtered technical indicator features and basic information of individual stocks, and the output is the trend category, future trends and regression values. The LSTM model uses the gradient descent method to continuously transfer the training error to the neuron training and minimize the error. During the training process, the weights of each time will be continuously modified according to the errors found during the training. The weight update and error minimization are as follows:

$$E = \sum_j \frac{1}{2}(t_j - y_j)^2 \tag{1}$$

In equation (1), E is the error function of the network, t_j is the target output, and y_j is the prediction. Then find the partial derivatives for each weight ω_{ji}:

$$\frac{\partial E}{\partial \omega_{ji}} = \frac{\partial \left(\frac{1}{2}(t_j - y_j)^2\right)}{\partial y_j} \frac{\partial y_j}{\partial \omega_{ji}} \tag{2}$$

$$\frac{\partial E}{\partial \omega_{ji}} = -(t_j - y_j)\frac{\partial y_j}{\partial \omega_{ji}} \qquad (3)$$

Then use Chain Rule to expand the activation function of each neuron:

$$\frac{\partial E}{\partial \omega_{ji}} = -(t_j - y_j)g'(h_j)\frac{\partial h_j}{\partial \omega_{ji}} \qquad (4)$$

$$\Delta\omega_{ji} = \alpha(t_j - y_j)\frac{\partial g}{\partial h_j} \cdot x_i \qquad (5)$$

α is the learning rate, g is the activation function of each neuron, and g' is the corresponding first derivative.

RNN-LSTM is a network in which nodes are connected along a sequence to form a directed graph, showing temporal dynamic behavior of time series. The feedforward calculation is as follows:

$$h(t) = Vx(t) + Uh(t-1)$$
$$O(t) = W\big(Vx(t) + Uh(t-1)\big), \forall t \qquad (6)$$

where $x(t)$ is the input of time t, $h(t)$ is the output of the hidden layer at time t, and $O(t)$ is the output at time t. U, V, W are the weight matrices for input layer to hidden layer, hidden layer to output layer, and hidden layer to the next time point hidden layer, respectively. In a time series recursive network, the gradient error will be passed back and forth along the time series layer by layer (back-propagation through time):

$$\Delta W = \frac{\partial E(t)}{\partial W}, \Delta V = \frac{\partial E(t)}{\partial V}, \Delta U = \frac{\partial E(t)}{\partial U} \qquad (7)$$

3.2.   Technical Analysis and Strategy

Technical analysis is a quantitative analysis of the price of a commodity based on statistics, and the signals of buying and selling are obtained through changes in technical indicators. There are many related discussions on the validity of technical analysis. From the past literature, the technical indicators, calculation methods, sample data or sample period used will affect the research results. This paper aims to explore the effectiveness of the Simple Moving Average (SMA), Stochastic Oscillator (KD), relative strength index (RSI), and Moving Average Convergence/Divergence (MACD) indicators as the effectiveness of deep learning training attributes, and test whether the technical analysis method is feasible in deep network architecture. According to the characteristics of individual financial indicators, this study converts them into appropriate normalized input values. The meaning of each indicator is as follows:

Simple Moving Average (SAM)

The concept of a moving average (MA) can be said to be the earliest and most basic method in technical analysis tools. Its theoretical basis is to average prices over a period

of time according to the concept of "Average Cost" by Dow Jones. When the bulls are moving, the moving average is showing an upward trend due to higher and higher prices. Conversely, when the market is showing a short pattern, the moving average is showing a downward trend due to lower and lower prices. Most people in the market are profitable and easy to get out of the "long" trend. In turn, the long-term moving average is driven upwards, which makes the trend of the trend upward. The equation of SAM is shown below:

$$SAM = \frac{p_1 + p_2 + p_3 + \cdots + p_n}{n} \qquad (8)$$

when calculating continuous values, one can directly use the original SAM increment:

$$SAM_{t1,n} = SAM_{t0,n} - \frac{p_1}{n} + \frac{p_{n+1}}{n} \qquad (9)$$

KD Line

The stochastic index, also known as the KD line [59], measures the position of the closing price at the highest and lowest price ranges to determine trends and entry and exit points. The random exponential coordinates are in the range of 0-100. The K line represents the closing price and the highest price within a certain period of time, the percentage of the lowest price. When the K line is higher than the D line, but the D line is broken in the overbought area, it is a sell signal (dead cross). When the K line is lower than the D line, but breaks through the D line in the oversold area, it is a buy signal (golden cross). To calculate the KD value, one must first obtain the RSV indicator:

$$RSV = \frac{C_n - L_n}{H_n - L_n} \times 100\% \qquad (10)$$

where n is the ttransaction date interval, C_n is the closing price of the n-th day, H_n is the highest price in the past n days and L_n is the lowest price in the past n days, and

$$K_n = \alpha \times RSV_n + (1 - \alpha) \times K_{n-1} \qquad (11)$$

$$D_n = \alpha \times RSV_n + (1 - \alpha) \times D_{n-1} \qquad (12)$$

If there is no K value or D value of the previous day, 50% can be substituted. In total, The K value is the 3-day smoothing moving average of the RSV value, and the D-value is the 3-day smoothing moving average of the K value.

Relative Strength Index (RSI).

The Relative Strength Index (RSI) [60] is a relative strength index, an indicator of the strength of the market's rise and fall. RSI is a technical indicator based on the strength of market fluctuations. This indicator can indicate the change in the strength of bullish

and bearish forces. When the RSI value rises, it means that the bullishness of the market is greater than the bearish force. Conversely when RSI falling, it means that the market is more bearish than bullish. Then, the average value of the uptrend during the period is used as a percentage of the sum of the rise and the average of the downtrend to represent the RSI of the relative strength of the buyer and the seller. The equation is shown below:

$$RSI = \frac{EMA_{(U,n)}}{EMA_{(U,n)} + EMA_{(D,n)}} \qquad (13)$$

suppose the price changes upwards is U and downward is D. In the day when prices rise, U = Today's closing price minus yesterday's closing price and D = 0. In the days when prices fell, U = 0 and D = yesterday's closing price minus today's closing price. $EMA_{(U,n)}$ is the U average in n days, and $EMA_{(D,n)}$ is the D average in n days.

Bias Ratio (BIAS)

BIAS [61] represents the gap between the stock closing price and the moving average of the day to analyze the degree of stock price deviation. Its function is mainly to measure the degree of deviation of the stock price from the moving average during the fluctuation process. When the stock price fluctuates sharply, it deviates from the moving average trend. As a result, possible retracements or rebounds, as well as the movement of stock prices within the normal fluctuation range, form the credibility of continuing the original trend.

$$BIAS_n = \frac{Close - MA_n}{MA_n} \times 100\% \qquad (14)$$

$BIAS_n$ indicates the deviation rate of the past n days, *Close* is the closing price of the nth day, and MA_n is the moving average of the past n days. Positive BIAS implied price may fall, otherwise it may rise.

William's Oscillator (Williams%R)

The William indicator [62] is an oscillator that measures the ratio of the peak (highest price) created by both long and short sides to the daily closing price and the ratio of stock price fluctuations within a certain period of time, providing a signal that the stock market trend is reversed. The William%R uses the pendulum principle to discern overbought or oversold stocks, discriminates highs and lows, and proposes effective investment signals.

$$Williams\%R = \frac{H_n - C_n}{H_n - L_n} \times 100\% \qquad (15)$$

where C_n is the closing price on the n*th* day, and H_n and L_n are the highest and lowest prices for the past *n* days.

Moving Average Convergence / Divergence (MACD)

The MACD [63] was proposed by Gerald Appel in the 1970s to study the strength, direction, energy, and trend cycle of stock price changes in order to capture the timing of stock buy and sell. The MACD uses the difference between the 12-days EMA and the 26-days EMA as a signal to determine the operation.

$$DIF_t = EMA_{t(close,12)} - EMA_{t(close,26)}$$

$$9MACD_t = MACD_{t-1} \times \frac{8}{10} + DIF_t \times \frac{2}{10}$$

(16)

where $EMA_{t(close,12/26)}$ is the exponential moving average of the closing price on day t (12/26), and $9MACD_t$ is a 9-day exponentially smooth moving average

4. Experiments

4.1.   Experimental Setup

This study uses the FTSE TWSE Taiwan 50 (TWSE 0050) Index as the research sample [64]. Most mutual funds are stock funds. These funds use stocks as the main investment target and can also be traded according to the professional judgment of the fund manager. ETFs are mutual funds that passively track the performance of an index and are listed in a centralized market. ETFs combine the trading characteristics of closed-end funds and open-end funds in trading methods. They can be traded on exchanges, and can also be purchased and redeemed. The main difference between ETFs and stocks is that buying an ETF is equivalent to buying securities that are tracked during the purchase period, effectively dispersing the risk of a single target and avoiding the price of a single target. The ETF operation aims to replicate the performance of the index, the investment portfolio is also consistent with the index constituent stocks, and the shareholding is also quite transparent. The TWSE0050 selected in this study has both the characteristics of stocks and ETFs. Investors can buy this stock at the market value in the stock exchange market, or they can invest in a fixed amount through banks. TWSE 0050 constituents include the top 50 listed companies on the Taiwan Stock Exchange, accounting for more than 70% of the total market capitalization of the Taiwan Stock Exchange. The data required for this study was obtained through the Taiwan Stock Exchange, and the transaction date used was 2017/1/4 ~ 2019/10/15 Q3, a total of 684 records. The first 2/3 of the data is set as training data, and the last 1/3 is test data. Each experiment features include daily opening price, closing price, highest price, lowest price, ups and downs, volume and turnover, and the normalized indicators.
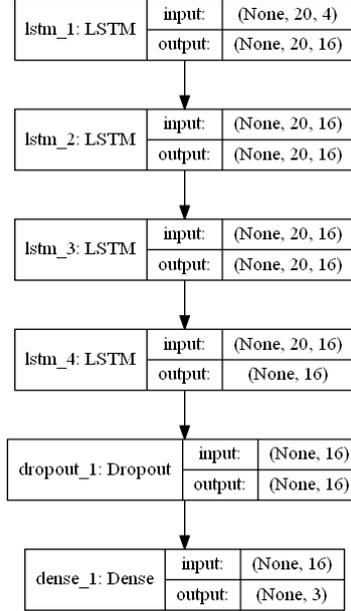
Fig. 2. The proposed LSTM layered architecture

The proposed LSTM layered architecture is set as Fig. 2. The model contains 4 LSTM layers and a full-connected layer. The experiments in this study were trained by ASUS ESC8000 G3, with Intel Xeon processor E5-2600 v3, NVIDIA GeForce GTX 1080ti GPU×8, and 64GB RAM. In the classification case, the data label is adjusted as follows:

1. Trend up: $Close_{t+1}\text{-}Close_t > Close_t \times \alpha$

2. Flat trend: $Close_t \times \alpha > Close_{t+1}\text{-}Close_t > -Close_t \times \alpha$

3. Trend down: $Close_{t+1}\text{-}Close_t < -Close_t \times \alpha$

where $Close_t$ is the closing price of day t, $Close_{t+1}$ is the closing price of day $t+1$, and α is an adjustable parameter. In subsequent experiments, α is set to 1.0%. Deep network model uses four layers of LSTM with cross-entropy loss function, and the LSTM input timestamp is 20 days. The dropout rate was set to 0.5. The learning model is trained with a batch size 16 for 30~100 epochs using the inverse decay learning rate policy, which de-fined as follows:

$$\ell = \frac{\ell_0}{1 + (r \times t)} \tag{17}$$

where ℓ_0 is the initial learning rate at epoch 0, t is the number of current epochs, r is a hyperparameters to be tuned. The learning rate started at 0.01 and reduced every epoch with $r = 0.1$.

4.2.   Technical Index Attribute Setting

As mentioned in the previous section, the technical analysis indicators used in this research are KD, RSI, BIAS, William%R and MACD. The financial range given by each indicator has different financial significance. In this study, each value of technical indicators is classified into three categories A, B, and C according to their financial significance, indicating that the future stock price may rise, stay flat, and fall.

KD Normalization

The cross breakthrough of the K line and the D line is more accurate when it is above 80 or below 20. When the cross is around 50, it means that the market trend is caught in the market. At this time, the trading signals provided by the cross breakthrough are invalid. When the K line is higher than the D line but the next day K falls below the D value, it is a sell signal (death cross); when the K line is lower than the D line but the next day K breaks the D value upward, it is a buy signal (gold cross). In this experiment, the KD indicator is converted into the following attributes and input into the model:

$$\begin{cases} A: K_{t-1} < D_{t-1}, D_t < K_t, \text{and } (20 < K_{t-2} < 80 \text{ or } 20 < K_{t-1} < 80 \text{ or } 20 < K_t < 80) \\ B: K_{t-1} > D_{t-1}, D_t > K_t \text{ and } (20 < K_{t-2} < 80 \text{ or } 20 < K_{t-1} < 80 \text{ or } 20 < K_t < 80) \\ C: \text{Otherwise} \end{cases} \quad \textbf{(18)}$$

RSI Normalization

RSI defines the value of relative strength between 0 and 100. When the indicator rises to 80, it indicates that the stock market has been overbought. If it continues to rise, if it exceeds 90, it means that it has reached the warning zone of severe overbought. The stock price has formed a head. Conversely, if it is lower than 20, it indicates that the market is oversold, and the stock price may enter the bottom. Therefore, the RSI numerical threshold for this study is set to 20 and 80:

$$\begin{cases} A: RSI_t \leq 20 \\ B: 20 < RSI_t < 80 \\ C: RSI_t \geq 80 \end{cases} \quad \textbf{(19)}$$

BIAS Normalization

A positive BIAS is called a positive deviation, and a value of more than 3.5% may have a price drop correction, and the stock price may fall; a negative BIAS is called a negative deviation, and a value of -3% or less may have a price increase correction. In order to make the value of the attribute more concise and easy to understand, we convert the BIAS value into the following value, where t is a trading day:

$$\begin{cases} A: \ BIAS_t \leq -0.03 \\ B: \ -0.03 < BIAS_t < 0.035 \\ C: BIAS_t \geq 0.035 \end{cases} \tag{20}$$

W%R Normalization

When the value of the W%R is greater than 80, it is oversold, and the stock price trend will bottom out; when the value of the W%R is less than 20, it is overbought and will recommend selling. This study converts W%R into the following categories, where t is a trading day:

$$\begin{cases} A: W\%R_t \geq 80 \\ B: 20 < W\%R_t < 80 \\ C: W\%R_t \leq 20 \end{cases} \tag{21}$$

MACD Normalization

When DIF is higher than MACD but the next day DIF falls below the MACD value, it is a sell signal and indicating that the stock price may fall in the future; While DIF is lower than MACD but the next day DIF breaks the MACD value, a buy signal indicating that the future stock price may rise. We convert the MACD value into the following categories, where t is a trading day:

$$\begin{cases} A: DIF_{t-1} < 9MACD_{t-1} \ and \ 9MACD_t < DIF_t \\ B: DIF_{t-1} > 9MACD_{t-1} \ and \ 9MACD_t > DIF_t \\ C: Otherwise \end{cases} \tag{22}$$

4.3.   Experimental Results

In this experiment, the number of epoch ranges from 30 to 100 until convergence. The test uses basic features plus technical analysis indicators, including normalized KD, RSI, BIAS, Williams%R and MACD. The experimental results are shown in Figs. 3-5.
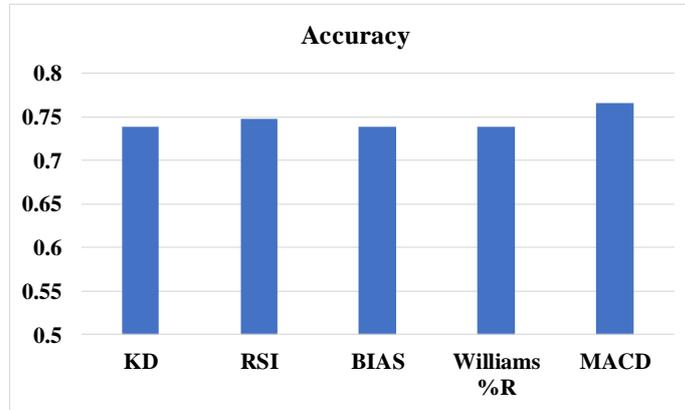
Fig. 3. Technical analysis indicator accuracy

(a)   (b)
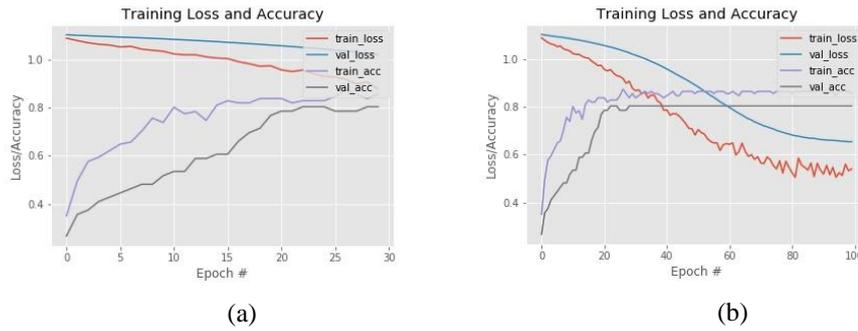
Fig. 4. Training Loss and Accuracy (a) with epochs 30 and (b) with epochs 100
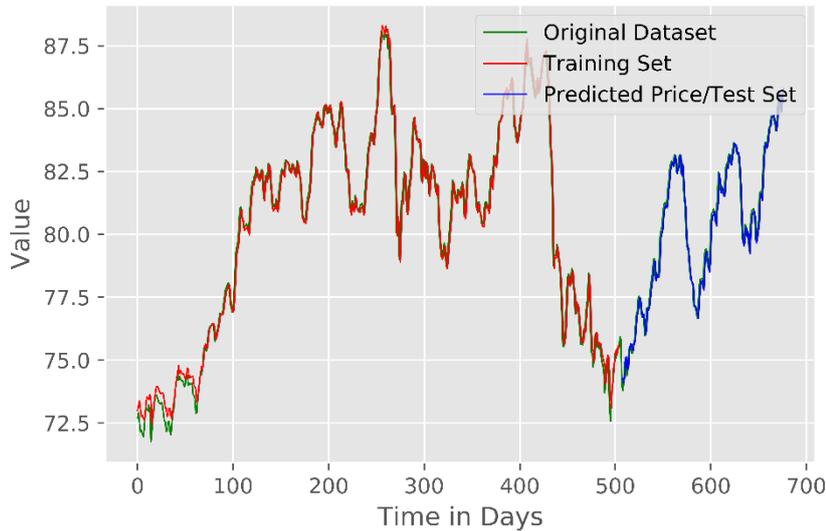
Fig. 5. Regression Results

The experimental results show that the accuracy of individual KD, RSI, BIAS, Williams%R, and MACD are 0.74, 0.75, 0.74, 0.74, and 0.76, respectively. The average accuracy of all indicators is 0.75. Among all the indicators, MACD obtained the highest estimation accuracy. Observation shows that the MACD indicator smoothes the closing price of the stock price according to the moving average and calculates the arithmetic average before integrating. The indicator uses the signs of the short-term and long-term moving average trends and performs double smoothing. Compared to other basic moving average-based indicators, MACD can provide more learnable information on the time series deep neural network. Further observations show that the current mainstream financial technology indicators are highly correlated with the moving average, therefore the network may automatically smooth out the differences in its deep structure. In addition, if all technical indicators are combined, the accuracy can reach 83.6%, and reach convergence in about 50 epochs. Fig. 5 is the regression results. The green line is the original data, the red line is the regression prediction result of the training data, and the blue line is the prediction result of the test data. The results show that our proposed model can accurately predict the trend and turn of the stock price in most cases.

5.   Conclusions

Differs from previous studies [65], [66], this research constructs a four-layer LSTM deep neural network and explores the effectiveness of technical indicators in the deep network. The experimental data uses TWSE0050 transaction data from 2017 to 2019 Q3. The research results show the prediction accuracy combined with comprehensive technical indicators can up to 83%. The use of individual indicators can also achieve an

average accuracy of 75%. The results of this study demonstrate that the use of technical analysis to the prediction of stock prices in the deep network is indeed feasible and effective. The indicators used in this research are general-purpose technical analysis indicators. The nature and calculation methods are applicable to all daily volatile financial commodities, such as exchange rate indexes, crude oil prices, futures and other derivative financial commodities. Therefore, the effectiveness of this study can be applied to different financial commodity transactions.

The focus of future research is to improve the accuracy of model estimates. In addition to technical analysis indicators, the fundamentals of individual stocks, chip analysis, financing/securities lending, etc., are also important reference indicators and this study will further examine how these important business factors can be added to the deep model.

References

1. Pagano, M.: Financial Markets and Growth: An Overview. European Economic Review, Vol. 37, No. 2-3, 613-622. (1993)
2. Mantegna, R. N.: Hierarchical Structure in Financial Markets. The European Physical Journal B-Condensed Matter and Complex Systems, Vol. 11, No. 1, 193-197. (1999)
3. Campbell, J. Y., Champbell, J. J., Campbell, J. W., Lo, A. W., Lo, A. W., & MacKinlay, A. C.: The Econometrics of Financial Markets. Princeton University Press. (1997)
4. De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J.: Noise Trader Risk in Financial Markets. Journal of Political Economy, Vol. 98, No. 4, 703-738. (1990)
5. Saint-Paul, G.: Technological Choice, Financial Markets and Economic Development. European Economic Review, Vol. 36, No. 4, 763-781. (1992)
6. Fouque, J. P., Papanicolaou, G., & Sircar, K. R.: Derivatives in Financial Markets with Stochastic Volatility. Cambridge University Press. (2000)
7. Stiglitz, J. E.: Financial Markets and Development. Oxford Review of Economic Policy, Vol. 5, No. 4, 55-68. (1989)
8. LeCun, Y., Bengio, Y., & Hinton, G.: Deep Learning. Nature, Vol. 521, No. 7553, 436-444. (2015)
9. Cao, C., Liu, F., Tan, H., Song, D., Shu, W., Li, W., ... & Xie, Z.: Deep Learning and its Applications in Biomedicine. Genomics, Proteomics & Bioinformatics, Vol. 16, No. 1, 17-32. (2018)
10. Miikkulainen, R., Liang, J., Meyerson, E., Rawal, A., Fink, D., Francon, O., ... & Hodjat, B.: Evolving Deep Neural Networks. In Artificial Intelligence in the Age of Neural Networks and Brain Computing, 293-312. Academic Press. (2019)
11. Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X.: Deep Learning and its Applications to Machine Health Monitoring. Mechanical Systems and Signal Processing, Vol. 115, 213-237. (2019)
12. Mikolov, T., Karafiát, M., Burget, L., Cernocký, J., & Khudanpur, S.: Recurrent Neural Network based Language Model. In Interspeech, Vol. 2, 3. (2010)
13. Hochreiter, S., Bengio, Y., Frasconi, P., & Schmidhuber, J.: Gradient Flow in Recurrent Nets: The Difficulty of Learning Long-term Dependencies. (2001)
14. Taylor, M. P., & Allen, H.: The Use of Technical Analysis in the Foreign Exchange Market. Journal of international Money and Finance, Vol. 11, No. 3, 304-314. (1992)

15. AS, S.: A Study on Fundamental and Technical Analysis. International Journal of Marketing, Financial Services & Management Research, Vol. 2, No. 5, 44-59. (2013)
16. Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y., & Ngo, D. C. L.: Text Mining for Market Prediction: A Systematic Review. Expert Systems with Applications, Vol. 41, No. 16, 7653-7670. (2014)
17. Sankar, C. P., Vidyaraj, R., & Kumar, K. S.: Trust based Stock Recommendation System–A Social Network Analysis Approach. Procedia Computer Science, Vol. 46, 299-305. (2015)
18. Ahmadi, E., Jasemi, M., Monplaisir, L., Nabavi, M. A., Mahmoodi, A., & Jam, P. A.: New Efficient Hybrid Candlestick Technical Analysis Model for Stock Market Timing on the Basis of the Support Vector Machine and Heuristic Algorithms of Imperialist Competition and Genetic. Expert Systems with Applications, Vol. 94, 21-31. (2018)
19. Bettman, J. L., Sault, S. J., & Schultz, E. L.: Fundamental and Technical Analysis: Substitutes or Complements?. Accounting & Finance, Vol. 49, No. 1, 21-36. (2009)
20. Anbalagan, T., & Maheswari, S. U.: Classification and Prediction of Stock Market Index based on Fuzzy Metagraph. Procedia Computer Science, Vol. 47, No. 5, 214-221. (2015)
21. Brown, D. P., & Jennings, R. H.: On Technical Analysis. The Review of Financial Studies, Vol. 2, No. 4, 527-551. (1989)
22. Edwards, R. D., Bassetti, W. H. C., & Magee, J.: Technical Analysis of Stock Trends. CRC Press. (2007)
23. Antoniou, A., Ergul, N., Holmes, P., & Priestley, R.: Technical Analysis, Trading Volume and Market Efficiency: Evidence from an Emerging Market. Applied Financial Economics, Vol. 7, No. 4, 361-365. (1997)
24. Kirkpatrick II, C. D., & Dahlquist, J. A.: Technical Analysis: the Complete Resource for Financial Market Technicians. FT Press. (2010)
25. Rockefeller, B.: Technical Analysis for Dummies. John Wiley & Sons. (2019)
26. Schwager, J. D.: A Complete Guide to the Futures Market: Technical Analysis, Trading Systems, Fundamental Analysis, Options, Spreads, and Trading Principles. John Wiley & Sons. (2017)
27. Scott, G., Carr, M., & Cremonie, M.: Technical Analysis: Modern Perspectives. CFA Institute Research Foundation. (2016)
28. Schannep, J.: Dow Theory for the 21st Century. Technical Indicators for Improving Your Investment Results, Vol. 4. (2008)
29. Brown, S. J., Goetzmann, W. N., & Kumar, A.: The Dow theory: William Peter Hamilton's Track Record Reconsidered. The Journal of finance, Vol. 53, No. 4, 1311-1333. (1998)
30. Elliott R. N.: The Wave Principle. In: Prechter RR Jr (ed) 1994. R. N, Elliott's Masterworks. (1938)
31. Malkiel, B. G., & Fama, E. F.: Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance, Vol. 25, No. 2, 383-417. (1970)
32. Malkiel, B. G.: A Random Walk Down Wall Street: Including a Life-cycle Guide to Personal Investing. WW Norton & Company. (1999)
33. Hudson, R., Keasey, K., & Dempsey, M.: Share prices under Tory and Labour Governments in the UK since 1945. Applied Financial Economics, Vol. 8, No. 4, 389-400. (1998)
34. Bohan, J.: Relative Strength: Further Positive Evidence. The Journal of Portfolio Management, Vol. 8, No. 1, 36-39. (1981)
35. Hinich, M. J., & Patterson, D. M.: Evidence of Nonlinearity in Daily Stock Returns. Journal of Business & Economic Statistics, Vol. 3, No. 1, 69-77. (1985)
36. Pruitt, S. W., & White, R. E.: The CRISMA Trading System: Who Says Technical Analysis Can'. Journal of Portfolio Management, Vol. 14, No. 3, 55. (1988)
37. Lo, A. W., Mamaysky, H., & Wang, J.: Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation. The Journal of Finance, Vol. 55, No. 4, 1705-1765. (2000)

38. Nazário, R. T. F., e Silva, J. L., Sobreiro, V. A., & Kimura, H.: A Literature Review of Technical Analysis on Stock Markets. The Quarterly Review of Economics and Finance, Vol. 66, 115-126. (2017)

39. Detzel, A. L., Liu, H., Strauss, J., Zhou, G., & Zhu, Y.: Bitcoin: Predictability and Profitability via Technical Analysis. SSRN Electronic Journal. (2018)

40. Ye, F., Zhang, L., Zhang, D., Fujita, H., & Gong, Z.: A Novel Forecasting Method based on Multi-order Fuzzy Time Series and Technical Analysis. Information Sciences, Vol. 367, 41-57. (2016)

41. Jiang, F., Tong, G., & Song, G.: Technical Analysis Profitability Without Data Snooping Bias: Evidence from Chinese Stock Market. International Review of Finance, Vol. 19, No. 1, 191-206. (2019)

42. Urquhart, A., & Zhang, H.: The Performance of Technical Trading Rules in Socially Responsible Investments. International Review of Economics & Finance, Vol. 63, 397-411. (2019)

43. Yu, L., Wang, S., & Lai, K. K.: Forecasting Crude Oil price with an EMD-based Neural Network Ensemble Learning Paradigm. Energy Economics, Vol. 30, No. 5, 2623-2635. (2008)

44. Zhuge, Q., Xu, L., & Zhang, G.: LSTM Neural Network with Emotional Analysis for Prediction of Stock Price. Engineering Letters, Vol. 25, No. 2. (2017)

45. Nelson, D. M., Pereira, A. C., & de Oliveira, R. A.: Stock Market's Price Movement Prediction with LSTM Neural Networks. In 2017 International Joint Conference on Neural Networks (IJCNN), 1419-1426. IEEE. (2017)

46. Soon, G. K., On, C. K., Rayner, A., Patricia, A., & Teo, J.: A CIMB Stock Price Prediction Case Study with Feedforward Neural Network and Recurrent Neural Network. Journal of Telecommunication, Electronic and Computer Engineering (JTEC), Vol. 10, No. 3-2, 89-94. (2018)

47. Chen, Y., Wei, Z., & Huang, X.: Incorporating Corporation Relationship via Graph Convolutional Neural Networks for Stock Price Prediction. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management, 1655-1658. ACM. (2018)

48. Chou, J. S., & Nguyen, T. K.: Forward Forecast of Stock Price using Sliding-window Metaheuristic-optimized Machine-learning Regression. IEEE Transactions on Industrial Informatics, Vol. 14, No. 7, 3132-3142. (2018)

49. Rasekhschaffe, K. C., & Jones, R. C.: Machine Learning for Stock Selection. Financial Analysts Journal, Vol. 75, No. 3, 70-88. (2019)

50. Kim, H. Y., & Won, C. H.: Forecasting the Volatility of Stock Price Index: A Hybrid Model Integrating LSTM with Multiple GARCH-type Models. Expert Systems with Applications, Vol. 103, 25-37. (2018)

51. Li, Q., Tan, J., Wang, J., & Chen, H.: A Multimodal Event-driven LSTM Model for Stock Prediction Using Online News. IEEE Transactions on Knowledge and Data Engineering. (2020)

52. Do, Q., & Trang, T.: Forecasting Vietnamese Stock Index: A Comparison of Hierarchical ANFIS and LSTM. Decision Science Letters, Vol. 9, No. 2, 193-206. (2020)

53. Ukrit, M. F., Saranya, A., & Anurag, R.: Stock Market Prediction Using Long Short-Term Memory. In Artificial Intelligence and Evolutionary Computations in Engineering Systems, 205-212. Springer, Singapore. (2020)

54. Feng, F., He, X., Wang, X., Luo, C., Liu, Y., & Chua, T. S.: Temporal Relational Ranking for Stock Prediction. ACM Transactions on Information Systems (TOIS), Vol. 37, No. 2, 1-30. (2019)

55. Zhang, Y., Chu, G., & Shen, D.: The Role of Investor Attention in Predicting Stock Prices: The Long Short-term Memory Networks Perspective. Finance Research Letters, 101484. (2020)

56. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P.: Gradient-based Learning Applied to Document Recognition. Proceedings of the IEEE, Vol. 86, No. 11, 2278-2324. (1998)
57. Rumelhart, D. E., Hinton, G. E., & Williams, R. J.: Learning Representations by Back-propagating Errors. Cognitive modeling, Vol. 5, No. 3, 1. (1988)
58. Hochreiter, S., & Schmidhuber, J.: Long Short-term Memory. Neural Computation, Vol. 9, No. 8, 1735-1780. (1997)
59. Murphy, J. J.: Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications. Penguin. (1999)
60. Wilder, J. W.: New concepts in technical trading systems. Trend Research. (1978)
61. Abdulali, A.: The Bias Ratio™: Measuring the Shape of Fraud. Protégé Partners, New York. (2006)
62. Williams, L. R.: How I Made One Million Dollars... Last Year... Trading Commodities. Windsor Books. (1979)
63. Appel, G.: Technical analysis: power tools for active investors. FT Press. (2005)
64. TWSE 0050. [Online] Available: https://www.twse.com.tw/en/ETF/fund/0050 (current November 2020)
65. Tao, Z., Muzhou, H., & Chunhui, L.: Forecasting stock index with multi-objective optimization model based on optimized neural network architecture avoiding overfitting. Computer Science and Information Systems, Vol. 15, No. 1, 211-236. (2018)
66. Hu, H., Tang, Y., Xie, Y., Dai, Y., & Dai, W.: Cognitive computation on consumer's decision making of internet financial products based on neural activity data. Computer Science and Information Systems, Vol. 17, No. 2, 689-704. (2020)

Ming-Che Lee received the Ph.D. degrees from National Cheng Kung University, Taiwan, in 2008. Since 2008, he has been an Associate Professor in computer and communication engineering with Ming Chuan University, Taoyuan, Taiwan. His research interests include deep learning, algorithms, and artificial intelligence.

Jia-Wei Chang is currently an Assistant Professor with the Department of Computer Science and Information Engineering, National Taichung University of Science and Technology. He is also the Chair of Young Professionals with the IET Taipei Local Network. His research interests include natural language processing, the Internet of Things, artificial intelligence, data mining, and e-learning technologies.

Jason C. Hung is an Associate Professor of Department of Computer Science and Information Engineering at National Taichung University of Science and Technology, Taiwan, ROC. His research interests include Multimedia System, e-Learning, Affective Computing, Artificial Intelligence and Social Computing. Dr. Hung received his Ph.D. in Computer Science and Information Engineering from Tamkang University in 2001. Dr. Hung participated in many international academic activities, including the organization of many international conferences. He is the founder of International Conference on Frontier Computing. He served as Vice Chair of IET Taipei LN. In April of 2014, he was elected as Fellow of the Institution of Engineering and Technology (FIET). He was elected as vice chair of IET Taipei LN in Nov. 2014. From June 2015, He is Editor-in-Chief of International Journal of Cognitive Performance Support.

Bae-Ling Chen is currently an Assistant Professor with the College of Intelligence, National Taichung University of Science and Technology, Taiwan.

Text Recommendation Based on Time Series and Multi-label Information

Yi Yin[1,2*], Dan Feng[1], Zhan Shi[1] and Lin Ouyang[2]

[1] Wuhan National Laboratory for Optoelectronics,
Huazhong University of Science and Technology;
Key Laboratory of Data Storage System, Ministry of Education,
Huazhong University of Science and Technology,
430074 Wuhan, Hubei, China
[*] yinyi@wust.edu.cn, {dfeng, zshi}@hust.edu.cn
[2] School of Computer Science and Technology,
Wuhan University of Science and Technology,
430073 Wuhan, Hubei, China
ouyanglin@wust.edu.cn

Abstract. One of the key functions of the method of text recommendation is to build a correlation analysis to all the text collection. At present, most of the text recommendation methods use the citation network, but less to consider the internal relations, which has become a challenge and an opportunity for the research of text recommendation. Therefore, we propose a new method to ameliorate the above problem based on the time series in this paper. We specify a certain text collection according to the interests of users and integrate the varied label values of the text, then we build the correlation coefficient between text and its related text with the differential analysis, finally the similarity degree of the text is calculated out by using the improved cosine similarity correlation matrix to promote a recommendation of similar text. Our experiments indicate that we are able to ensure the quality of text, with an improvement of accuracy by 8.63% as well as an improvement of recall rate by 5.25%.

Keywords: time series; label value; correlation coefficient; similarity degree

1.　Introduction

Nowadays it's becoming a dogma that more and more dates have been connected together as a colossal directed network with a common case of cited node [1, 2]. In the directed network, the latest data squint to achieve their own functions in view of the original data. One of the most typical applications is the processing of text data, for instance, the new text customarily needs to cite the original ones to achieve its own research and innovation.

That may easily lead to some ineluctable similarity of these mutually referred texts, the association analysis of these similarity texts has become one of the main ideals of the text recommendation to meet the needs of different users. Many previous text

---

[*] Corresponding author

recommendations have been used to take the citation relationship as the main approach to achieve their function, which means, if the text 1 cited the text 2, they are related. However, these methods are too simple to get the useful information [3]. Therefore, more and more researches of text recommendation have been indulged in exploring the label value on this basis, a variety of recommendation methods have been put forward according to the multi-attributes of the text [4, 5], like the keyword, research field as well as citation times and so on.

Still, we should pay attention to these ubiquitous problems hidden in the application of the label value in the text recommendation. (1) The label value is quite abstract, that is to say, there are still unfathomed symptoms, the user will not get the accurate text easily through these tags and he is more likely to capture the text in vain. (2) The label value analysis of the existing methods of text recommendation usually adopts an extensive weighted average, which ignores the individual differences. That model built with this method cannot reflect the different characteristics of the global data. (3) Generally speaking, there have been two accesses to the label value for the user, one way is based on the explicit formulation of the user himself, the other way is to utilize the helpful information of text to capture the label value covertly. However, both methods get the problem of how to extract information and give a standard measurement, which results in a waste of the user's time and a lower expectation of the recommendation system [6]. At the meantime, these two problems should be the main reason of people being not satisfied with many existing methods of text recommendation. (4) Most of the existing text recommendation methods only use a single scoring model extensively to mark the label value one by one and they would default to be independent of each other, which might be ignoring of these multitudinous label values.

In order to solve the problems mentioned above, we propose a new text recommendation method based on the time series, which is combined with the multi-label value. This method imports the time series into the label value. What the time series denotes was a set of statistical data which can be arranged in order of time. In this paper, the statistical index of time series is exactly the varied of label value of the text, the value of label value is directly related to the length of time, and the label value can usually be obtained through continuous summary in the time series. When the time series is introduced into the label value, the label value can be dynamically extracted and created. Thus, we can capture a set of ideal recommendable texts with time-validity.

So far, there have been some researches on the applying of the time series in text analysis. ZR Jiang et al. had applied the time series to subject analysis. S Bjork et al. [7] analysed the pattern of innovation cycle time of the economy knowledge that the French Nobel Prize winner from 1930 to 2005 had kept to base on the time series. A Lercher et al. [8] gave a research of correlation rule of text with the time series analysis, etc. However, there has been always a pity that scarcely any researcher had taken the combination of time series and label value into consideration in the area of text recommendation.

There are several benefits for the combined utilization of time series and multi-label value in text recommendation method: (1) the text data is one kind of fairly detailed data, and the thoughts attached on text data always have limitations. In order to give a response to the specific needs of users, it is always necessary to carry out a full range of multi-angle analysis of the text, so here may be a large number of label values that can

meet the needs of different users [9], as label value is one of the attributes of the text itself. (2) For the text recommendation, the time of a text is a key factor that cannot be ignored by the user. This is because the information content of the text is timeliness. For example, the timeliness of a news text is very strong; its time interval needs to be accurate to day or even less, as for academic texts, although their timeliness is not strong, it is still necessary to meet the need of users according to the time series. The reason of a text being arranged in the order of time is that the latter text is based on the development of society and science, and especially the previous text. (3) There is citation relationship between a text and its reference, which takes time of each text as the premise, that is to say, the recent text can cite the earlier text well but not on the contrary as data are time ordered. If we ignore the text sequence, the research such as text citation times and references relationship will be out of the question, the time series is the prerequisite, the citation of a text is the need to be guaranteed with very strong time series.

Based on the introduction above, the main contributions of our work in this paper include the following aspects: (1) Quantify the multi-label value of the text, which is more flexible than the previous method of obtaining information through the subject of the text and citation times monolithically. (2) Establish a combination model of a time series with multi-label information to capture the correlation coefficient between texts by the method of differential cryptanalysis, and carry out the similarities between texts [10]. (3) Consider the evaluation value of a text to the user, the practicability of the proposed method is verified on a large-scale real data set.

Our ground work in this paper is elaborated as follow. We introduce the related work in the second chapter and describe the concrete form of a time series and process of the label value in the third chapter. The basic definition and the assignment of label value, along with the computer realization method are mentioned to obtain the correlation coefficient of a text by a difference equation in the chapter four, while the fifth chapter demonstrates our method of text recommendation, then we conclude our work in section 6.

2.   Current Practice and Research

At present, some text recommendation algorithms based on time series usually extract some label information in the text shared by the users, and then generate the algorithm by making the label information connected through statistical learning method. When extracting label information, the algorithm generally pays attention to the importance of a part of label information in common knowledge. For example, focusing on keyword label indicates that the algorithm wants to recommend more relevant text to users, or focusing on the authoritative label indicates that the algorithm wants to recommend more classic text to users, etc. Different label information and models reflect different emphases of the algorithm. The analysis of the cited text usually use the graph model to represent the relationship between texts, and a large amount of text recommendation researches have adopted the method of extracting label value.

Gupta S et al. came up with a recommendation method based on the theme and the core idea of the text [11], in their method, the user had to provide a full text (including

title, abstract, text and reference) for extracting the core idea. Similarly, Tellez E S et al. introduced an independent framework to recommend the useful text [12], in which users also had to enter a full academic literature text to generate some different information, and then submitted it to the existing network information resources to realize the recommendation. Mäntylä M V et al. took the cited text abstract, introduction and conclusion to obtain better recommendation results [13], however, these methods not only actually increased the user's burden, but also could not be able to provide extra information on the basis of a section of the user's interested information in the real environment. Caruccio L et al. studied the problem of how to use references to recommend based on the user's query without an additional reference list [14], they designed a non-parametric probability model, and calculated the correlation between the two texts by using the reference information. There were some other studies focusing on improving the topic similarities of the cited texts. Huang S et al. found out that in the topic clustering of citation, using references could effectively avoid the "drifting" [15]. Harman D et al. had verified the different extraction methods of references, which would have an influence on the quality of information retrieval [16]. All of these researches are effective methods for text recommendation.

At the same time, there also have been some researches on the single role of a time series in the text. For example, the generation of timing [17], timing based clustering and classification [18], information retrieval for future [19], etc.

According to the above analysis, a large number of recommendation algorithms based on shared text have the problems of independence of label information of default text or simplification of calculation method. That is to say, the same method is used in the quantitative calculation of label. Such a label extraction method is not comprehensive in the expression of text content.

3.   Time Series and Multi-Label Value

This chapter does a quantitative analysis of label value. Each access of the user is determined by a variety of label value. Therefore, the co-analysis of multi-label value denotes the relationship between the texts, which has benefit of finding the similar text to recommend to the user.

3.1.   Defined Variable

We supposed u as an arbitrary user, and text set as $\mathbf{X} = \{x_1, x_2, \cdots, x_i, \cdots, x_n\}$. The whole label value of $\mathbf{X}$ can be denoted by $\mathbf{g}(\mathbf{X})$,

$$\mathbf{g}(\mathbf{X}) = \begin{pmatrix} g_1(\mathbf{X}) \\ g_2(\mathbf{X}) \\ \vdots \\ g_m(\mathbf{X}) \end{pmatrix} = \begin{pmatrix} g_1(x_1) & g_1(x_2) & \cdots & g_1(x_n) \\ g_2(x_1) & g_2(x_2) & \cdots & g_2(x_n) \\ \vdots & \vdots & \ddots & \vdots \\ g_m(x_1) & g_m(x_2) & \cdots & g_m(x_n) \end{pmatrix}$$

Here, there were a number of m of label values for each x_i, label value of x_i $\mathbf{g}(x_i)$ should be $\mathbf{g}(x_i) = (g_1(x_i), g_2(x_i), \cdots, g_m(x_i))$. If x_i cited a set of text $\mathbf{x}_i$, we would like to

mark this text set as $\mathbf{x}_i = \{x_i^1, \quad x_i^2, \cdots, x_i^l\}$, The whole label value of $\mathbf{x}_i$ can be denoted by $\mathbf{g}(\mathbf{x}_i)$,

$$\mathbf{g}(\mathbf{x}_i) = \begin{pmatrix} g_1(\mathbf{x}_i) \\ g_2(\mathbf{x}_i) \\ \vdots \\ g_m(\mathbf{x}_i) \end{pmatrix} = \begin{pmatrix} g_1(x_i^1) & g_1(x_i^2) & \cdots & g_1(x_i^l) \\ g_2(x_i^1) & g_2(x_i^2) & \cdots & g_2(x_i^l) \\ \vdots & \vdots & \ddots & \vdots \\ g_m(x_i^1) & g_m(x_i^2) & \cdots & g_m(x_i^l) \end{pmatrix}.$$

As x_i cite this set of text $\mathbf{x}_i$, $\mathbf{x}_i = \{x_i^1, x_i^2, \cdots, x_i^l\}$, the correlation coefficient of x_i and $\mathbf{x}_i$ should be,

$$\mathbf{s}(x_i\mathbf{x}_i) = \begin{pmatrix} \mathbf{s}_1(x_i\mathbf{x}_i) \\ \mathbf{s}_2(x_i\mathbf{x}_i) \\ \vdots \\ \mathbf{s}_m(x_i\mathbf{x}_i) \end{pmatrix} = \begin{pmatrix} s_1(x_ix_i^1) & s_1(x_ix_i^2) & \cdots & s_1(x_ix_i^l) \\ s_2(x_ix_i^1) & s_2(x_ix_i^2) & \cdots & s_2(x_ix_i^l) \\ \vdots & \vdots & \ddots & \vdots \\ s_m(x_ix_i^1) & s_m(x_ix_i^2) & \cdots & s_m(x_ix_i^l) \end{pmatrix}$$

The first equation represents a set of importance relationships between x_i and $\mathbf{x}_i$, and the second equation represents the correlation coefficient which every text had referred by x_i.

Figure 1 showed the basic frame structure of a text reference and the relevant parameters.
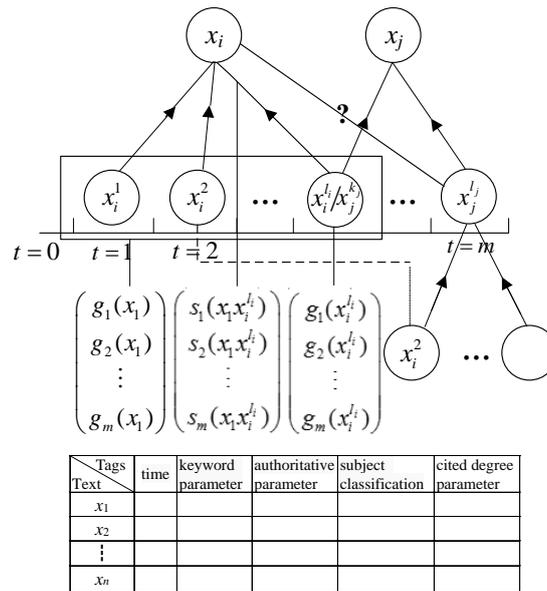
Fig. 1. Citation structure of text.

Figure 1 includes the citation correlation, label value, a time series, the correlation coefficient between the text and their correlation diagram. As we can see from this figure, for the text x_i and x_j, they have the same cited text x_i^l or x_j^k, and we can estimate

the similarity of x_i and x_j according to the comparison of two correlation coefficients

between $\begin{pmatrix} s_1(x_i x_i^l) \\ s_2(x_i x_i^l) \\ \vdots \\ s_m(x_i x_i^l) \end{pmatrix}$ and $\begin{pmatrix} s_1(x_j x_j^k) \\ s_2(x_j x_j^k) \\ \vdots \\ s_m(x_j x_j^k) \end{pmatrix}$.

Among the variants defined in this paper, if the text $\mathbf{x}_i$ is cited by x_i, which is also the text in set $\mathbf{X}$, that means $\mathbf{x}_i \subset \mathbf{X}$. In this paper, we adopted these two different recording ways to make it easier to discuss the text, for example, the text x_i^l in the figure 2, it was not only cited by x_i, but also cited by x_j, so x_i^l could be marked as x_j^k. The text x_i^2 can be cited by x_j^k as well as x_i, $\{x_i, x_j, x_i^1, x_i^2, x_i^l, x_j^k\}$ all came from the set of $\mathbf{X}$.

3.2.   Definition and Exposition of Time Series

The role of a time series in text recommendation is very important [20]. For text recommendation, there are two kinds of time, one is the published time of the text, the other is the user' access date to the text. The text published time usually reflects the existence of the text; the access time reflects the need dateline of the user for text information. In this paper, the text denoted to be the published time. If the label information comes from text, the label is a kind of special word information. If the label information comes from other statistics information, the label is not word information. Therefore, label information cannot be a part of word.

Here, t is used to represent the time parameter. For the time parameter t, there are several properties. First, as we have mentioned in chapter 1, all the texts $\mathbf{x}_i$ are cited by x_i arranged in chronological order of time t. Second, in view of the fact that text information is of timeliness, this paper makes a restriction on the publication time, which limits its minimum number of years. In this paper, we had to restrict the publication time of the text by limiting t to the minimum year in view of the timeliness of text. There will be a difference between the publication time and the minimum number of years, the minimum year denoted the initial time, let the text of the initial time or before as 0, then for all the text after this time node could be: the time parameter $t =$ publication time $-$ the initial time. For instance, when $\mathbf{x}_i$ is cited by x_i, the time parameter of text $\mathbf{x}_i$ can be arranged from small to large. If there is time span between text x_i^{l-1} and x_i^l, the time span of these two texts is also retained. Third, we can use the limitation of the publication time to avoid the multiple citation, and to reduce the amount of calculation. For example, x_i cites x_i^l, and x_i^l cites the other texts, after many such citations, some published time of papers here are less than the minimum year, so we can discard them. In other word, there is no need to calculate the correlation coefficient between these discard texts with x_i^l. If the text is sorted according to different time series, it may have an impact on the final classification result. However, this paper assumes that they are arranged according to the chronological order of natural years, forming a unified sorting order. This paper only discusses the correlation coefficients arranged according to this chronological order. Fourth, the correlation coefficient and label values are functions of time parameters, which are defined the similarity of the texts by using the difference. As for the definition and physical significance of difference

equations, the change would reflect on timeline, and we would be able to calculate the correlation coefficient of x_i and $\mathbf{x}_i$ with the difference equations. Fifth, for the texts in $\mathbf{x}_i$, there may be a number of the same time value. To solve this problem, we can do a secondary arrangement according to the value of keywords, from large to small. Which means, when $t(x_i^l) - t(x_i^{l-1}) = 0$ and $t(x_i^l) = t(x_i^{l-1}) \neq 0$, the sequence of x_i^{l-1} and x_i^l can be arranged from large to small according to one of the element value in $\mathbf{g}(\mathbf{x}_i)$. Sixth, consideration of the timeliness of the text is necessary, text information would generally show a process of decay with the increase of time, for instance, the near-term text would be more important than the older ones.

Last but not the least, texts has a citation relationship, and the citation relationship had to take time as a prerequisite, that means, the near time text can cite the earlier text well but not on the contrary. Therefore, text citation should to be guaranteed with a strong time order.

3.3.   Establish the Correlation Coefficient

Based on above definition, we discussed the value of correlation coefficient $\mathbf{s}(x_i\mathbf{x}^i)$ between x_i and $\mathbf{x}_i$. Assumed that the time series t as all the cited text $\mathbf{x}_i$ of x_i, $t = \{0, 1, 2, \cdots, m\}$. That means, t from 1 to m. When $t = 0$, the corresponding text denoted the most distal one of $\mathbf{x}_i$, on the contrary, when $t = m$, the corresponding text denoted the most proximal one of $\mathbf{x}_i$. For the correlation coefficient, since it represents the relationship between x_i and $\mathbf{x}_i$, we can establish a connection between them. At a given time, the influence degree of x_i^l to x_i is denoted as

$$\begin{pmatrix} g_1(x_i^l) \\ g_2(x_i^l) \\ \vdots \\ g_m(x_i^l) \end{pmatrix}' \cdot \begin{pmatrix} s_1(x_i x_i^l) \\ s_2(x_i x_i^l) \\ \vdots \\ s_m(x_i x_i^l) \end{pmatrix}$$

. Since the importance of text exhibits attenuation over time, and the whole process is discrete, therefore, the influence of $\mathbf{x}_i$ on x_i is

$$\begin{pmatrix} g_1(\mathbf{x}_i) \\ g_2(\mathbf{x}_i) \\ \vdots \\ g_m(\mathbf{x}_i) \end{pmatrix}' \cdot \begin{pmatrix} \mathbf{s}_1(x_i\mathbf{x}_i) \\ \mathbf{s}_2(x_i\mathbf{x}_i) \\ \vdots \\ \mathbf{s}_m(x_i\mathbf{x}_i) \end{pmatrix}$$

. The connection between x_i and $\mathbf{x}_i$ is the superposition of the influence of $\mathbf{x}_i$ on x_i in the whole time $T = \{1, 2, \cdots, \tau\}$. Then, the correlation of x_i and $\mathbf{x}_i$ could be:

$$\begin{pmatrix} g_1(x_i) \\ g_2(x_i) \\ \vdots \\ g_m(x_i) \end{pmatrix} - \begin{pmatrix} g_1(\mathbf{x}_i) \\ g_2(\mathbf{x}_i) \\ \vdots \\ g_m(\mathbf{x}_i) \end{pmatrix}' \cdot \begin{pmatrix} \mathbf{s}_1(x_i\mathbf{x}_i) \\ \mathbf{s}_2(x_i\mathbf{x}_i) \\ \vdots \\ \mathbf{s}_m(x_i\mathbf{x}_i) \end{pmatrix} = \mathbf{0} \tag{1}$$

Which can be converted to:

$$\begin{cases} g_1(x_i) - \sum_{t=1}^{l} g_1(x_i^t) \cdot s_1(x_i x_i^t) = 0 \\[2mm] g_2(x_i) - \sum_{t=1}^{l} g_2(x_i^t) \cdot s_2(x_i x_i^t) = 0 \\[2mm] \vdots \\[2mm] g_m(x_i) - \sum_{t=1}^{l} g_m(x_i^t) \cdot s_m(x_i x_i^t) = 0 \end{cases} \tag{2}$$

Each element in the equation set (2) is arranged in chronological order. We build a frame diagram of discrete system according to the equation set (2), which is shown in figure2.
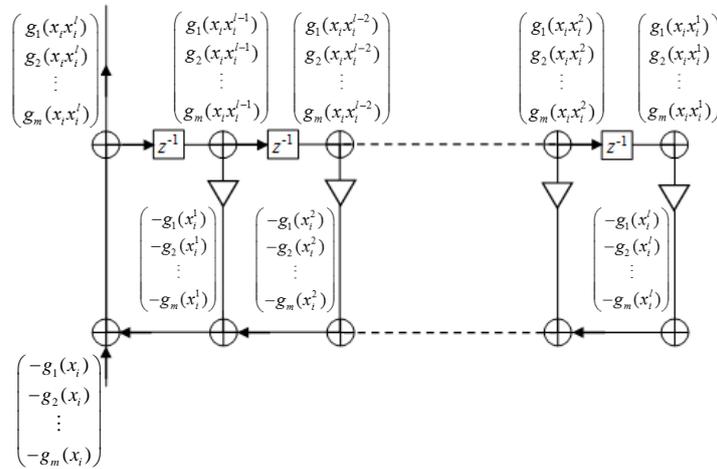
Fig. 2. System chart of the correlation function.

It shows the whole situation of text association, which is composed of n simple subsystems connected in a loop way. If the final association between texts is regarded as a steady state, figure 2 shows the state diagram of the whole association state. From this graph, the correlation equation can be established, that is the state equation, and then get the correlation coefficient between different texts by solving the state equation.

As we can see in figure2, the equation set (2) can be described as difference equation set, it was actually a linear time-invariant system and a zero state response system. By solving the difference equation, we carried out all of the correlation coefficient of x_i and $\{x_i^1, x_i^2, \cdots, x_i^l\}$. There were $m \times l$ solutions of this equation set:

$$\begin{pmatrix} s_1(x_i x_i^1) - D_1(1) \\ s_1(x_i x_i^2) - D_1(2) \\ \vdots \\ s_1(x_i x_i^l) - D_1(l) \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1l} \\ a_{11}^2 & a_{12}^2 & \cdots & a_{1l}^2 \\ \vdots & \vdots & \ddots & \vdots \\ a_{11}^l & a_{12}^l & \cdots & a_{1l}^l \end{pmatrix} \begin{pmatrix} C_{11} \\ C_{12} \\ \vdots \\ C_{1l} \end{pmatrix} \tag{3}$$

$$
\begin{pmatrix} s_2(x_i x_i^1) - D_2(1) \\ s_2(x_i x_i^2) - D_2(2) \\ \vdots \\ s_2(x_i x_i^l) - D_2(l) \end{pmatrix} = \begin{pmatrix} a_{21} & a_{22} & \cdots & a_{2l} \\ a_{21}^2 & a_{22}^2 & \cdots & a_{2l}^2 \\ \vdots & \vdots & \ddots & \vdots \\ a_{21}^l & a_{22}^l & \cdots & a_{2l}^l \end{pmatrix} \begin{pmatrix} C_{21} \\ C_{22} \\ \vdots \\ C_{2l} \end{pmatrix} \tag{4}
$$

$$
\vdots
$$

$$
\begin{pmatrix} s_m(x_i x_i^1) - D_{m1}(1) \\ s_m(x_i x_i^2) - D_{m2}(2) \\ \vdots \\ s_m(x_i x_i^l) - D_{ml}(l) \end{pmatrix} = \begin{pmatrix} a_{m1} & a_{m2} & \cdots & a_{ml} \\ a_{m1}^2 & a_{m2}^2 & \cdots & a_{ml}^2 \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}^l & a_{m2}^l & \cdots & a_{ml}^l \end{pmatrix} \begin{pmatrix} C_{m1} \\ C_{m2} \\ \vdots \\ C_{ml} \end{pmatrix} \tag{5}
$$

Then we simplified the above function as:

$$
\begin{pmatrix} \mathbf{s}_1(x_i \mathbf{x}^i) \\ \mathbf{s}_2(x_i \mathbf{x}^i) \\ \vdots \\ \mathbf{s}_m(x_i \mathbf{x}^i) \end{pmatrix} - \begin{pmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \\ \vdots \\ \mathbf{D}_m \end{pmatrix} = \begin{pmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ \vdots \\ \mathbf{V}_m \end{pmatrix}' \begin{pmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \\ \vdots \\ \mathbf{C}_m \end{pmatrix} \tag{6}
$$

Related to the equation set (2), $\mathbf{D} = (\mathbf{D}_1, \mathbf{D}_2, \cdots, \mathbf{D}_m)'$ could be able to denote as the solutions of $\mathbf{s}(x_i \mathbf{x}^i)$, and $\mathbf{V} = (\mathbf{V}_1, \mathbf{V}_2, \cdots, \mathbf{V}_m)'$ denoted the characteristic root of each equation, and $\mathbf{C} = (\mathbf{C}_1, \mathbf{C}_2, \cdots, \mathbf{C}_m)'$ denoted the coefficient of the characteristic root of each equation.

Here, the initial conditions of equation set (2) should be

$$
\begin{cases} g_1(x_i^{-t}) \cdot s_1(x_i x_i^{-t}) = 0 \\ g_2(x_i^{-t}) \cdot s_2(x_i x_i^{-t}) = 0 \\ \vdots \\ g_m(x_i^{-t}) \cdot s_m(x_i x_i^{-t}) = 0 \end{cases} \tag{7}
$$

Therefore, the values of $\mathbf{D}$, $\mathbf{V}$ and $\mathbf{C}$ can be carried out, and then the correlation coefficient $\mathbf{s}(x_i \mathbf{x}^i)$ was obtained.

In the same way, we could also get the correlation coefficient between x_j and $\mathbf{x}_j$, so that all the text $\mathbf{X}$ and their cited text can eventually be carried out as a value of correlation coefficient.

3.4.   Comparison of Correlation Coefficient

If the user u wants to obtain a series of texts related to the x_i, it is necessary to compare the correlation coefficient of the text that related to the x_i. For example, in Figure 1, $\{x_i^1, x_i^2, \cdots, x_i^l\}$ are directly related to x_i, and x_j is indirectly related to x_i. Both the similarity of x_i and $\{x_i^1, x_i^2, \cdots, x_i^l\}$ and the similarity of x_i and x_j, we had to compare their correlation coefficient. As the correlation coefficients are actually m vectors that composed with a set of label value, so we had to find out a method that was able to compare the vectors. We eventually came up with an improved cosine similarity to restrain the impact resulted from the m label value.

To obtain the similarity of x_i and x_j, we adopted an improved cosine similarity formula as below:

$$sim(\mathbf{s}(x_i x_k^i), \mathbf{s}(x_j x_j^1)) = \frac{\begin{pmatrix} s_1(x_i x_i^k) \\ s_2(x_i x_i^k) \\ \vdots \\ s_m(x_i x_i^k) \end{pmatrix}' \begin{pmatrix} \rho_1 & & & \\ & \rho_2 & & \\ & & \ddots & \\ & & & \rho_m \end{pmatrix} \begin{pmatrix} s_1(x_j x_j^1) \\ s_2(x_j x_j^1) \\ \vdots \\ s_m(x_j x_j^1) \end{pmatrix}}{\left[\begin{pmatrix} s_1(x_i x_i^k) \\ s_2(x_i x_i^k) \\ \vdots \\ s_m(x_i x_i^k) \end{pmatrix}' \begin{pmatrix} s_1(x_i x_i^k) \\ s_2(x_i x_i^k) \\ \vdots \\ s_m(x_i x_i^k) \end{pmatrix} \right]^{\frac{1}{2}} \cdot \left[\begin{pmatrix} s_1(x_j x_j^1) \\ s_2(x_j x_j^1) \\ \vdots \\ s_m(x_j x_j^1) \end{pmatrix}' \begin{pmatrix} s_1(x_j x_j^1) \\ s_2(x_j x_j^1) \\ \vdots \\ s_m(x_j x_j^1) \end{pmatrix} \right]^{\frac{1}{2}}} \quad (8)$$

Here, $\{\alpha_1, \alpha_2, \cdots, \alpha_m\}$ was applied to subtly adjust $\mathbf{s}(x_i x_i^l)$ and $\mathbf{s}(x_j x_j^l)$ into a more adaptive similarity. Parameter $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \cdots, \alpha_m\}$ is an important role, it controlled the rate of the m label values in the whole recommendation process, which denoted the restraint degree of label value.

Through this method, we can obtained a series of similar texts related to x_i, so we are able to offer the texts with high similarity to the user.

4.   Label Value

For the research of text recommendation, there have been some relatively standard text libraries at home and abroad, we can compare different recommendation methods and system performance in the common text library. And we can also adjust the function parameters, allowing users more flexible access to the text label information [21].

4.1.   Selection of Label Value

We have selected a part of texts from Web of Science[†] (WOS) to test our method in this paper. There were 31393 texts in the area of recommendation system according to the searching keywords "recommendation system" (RS). Web of Science is a product of Thomson Scientific in the United States, which is omnibus multi-disciplinary core journal citation index database and includes three famous citation databases (SCI, SSCI and A&HCI) and two major chemical information database (CCR&IC), as well as the three of SCIE, CPCI-S and CPCI-SSH Citation Database. In virtue of the recommendation method, the date from the search platform of ISI Web of Knowledge have been classified to five major kinds and 151 secondary classifications based on the subject category, and we have obtained an universal network file format (*.net) date set

---

[†] http://isiknowledge.com/

to do the citation analysis by the analytical tool of Web of Science [22]. Thus, we could not only effectively reveal the relationship between the internal connection and the texts (which provides the intrinsic link of scientific research), but also could find out the opportune ones to recommend to the user through a multi-tier analytical approach with the help of the provided analytical tool.

The main contents of label value were shown below. We could capture lots of meaningful label value by searching the keyword "text mining" on the WOS.

As can be seen in the figure3, these contents, like the same keyword, similar keywords, author name, journal name, publication time, number of citations, other text references, etc. can be selected as label value. These label value were the key factors in the composition of the text that is needed for an appropriate extraction in text recommendation.

Fig. 3. Example of label value.

4.2.   Classification of Label Value

For all of the text, we choose four kinds of label value as the valuable parameters and establish the weight for each type of label values. These four kinds of label value respectively named {keyword, the subject category, text authority, cited degree}, which could reflect the value of a text to the user as a four-dimensional vector. This paper considers the different needs of users, and the needs of users are comprehensive from the perspective of the paper data. Therefore, this article combines the type description method of statistical data to quantify the label information. The theoretical basis for the selection of the label content of the paper is that its selection method is based on statistical principles, and combined with the "statistical data type description" method to select the label.

(1) Keywords. It is the most important criteria for users to obtain the text, as well as the most important label in the method of text recommendation [23]. For the search of texts, the more detailed keywords are, the more accurate the results are. However, for

the text recommendation, less keywords may lead to a large number of irrelevant texts and an unstable accuracy at mean time, while too many keywords may result in missing some texts and a reduced recall rate. Therefore, the choice of the type and the number of keywords would have a direct impact on the text mining. In general, the role of secondary keywords is more important. We specified that there must be at least one keyword, whether it was a primary or secondary keyword, it must be presented in the title, summary, or keywords columns. In our work, we divided the keywords into three types according to the contents of the text: primary keywords, secondary keywords and non- keywords. As shown in Figure3, "text mining" was the main keyword, but for the first two texts in Figure3, "bi-text mining" and "discovery knowledge" are the secondary keywords. It is easy to see that there might be a completely different recommendation if we ignore the impact of secondary keywords. For example, although the keyword like "web mining" was totally different with the keyword "text mining", they can express some similar meaning. We converted the keyword into numeric variable, which is generally covered by the title, abstract, author's name, address and so on.

(2) Subject category. Since "WOS" has covered almost all the fields of discipline, the retrieved texts according to one certain keyword may belong to more than one subject category. For instance, if we use "text mining" as the keyword to search, besides of the computer science [24], the retrieved texts can be distributed in linguistics, psychology, Communication Sciences and the related subjects, and so on. What's more, it could be divided into a subject which has little to do with "text mining", like physics and neuroscience, if there is some blend of text mining related algorithm and the subject itself. So we need to label the subject of a text with three different types: original disciplines, related disciplines and other disciplines. Subject categories can be obtained directly from the publishing organization to which the text belongs, and it is easily distinguishable.

(3) Text authority. It is generally believed that there is a very important criterion to judge whether the authority of scientific research texts is in the level of publication of journals or conferences. There are many methods for the classification of publications at home and abroad. For the experimental data of text, we classified the texts in "WOS" into five kinds according to the authority of the texts, which is based on the commonly used classification methods of the academic journals in China, namely core A, core B, core C, regular D and the others.

(4) Cited time. The number of citations per text is easily available. The data of this parameter should be bigger. But if the difference between the two texts is small, the importance of these two texts is hard to determine. Moreover, the number of citations is also relative value, in which there is no uniform unit. For numerical variable, the date should be processed with the discrete way according to width and frequency or another settled method, as the cited times can be processed basing on the times rank 0, 1~10, 11~15 and >15. We could also use Gaussian distribution to set out a mean of the cited number, and then divided the cited degree according to the variance. However, we should mention that there was not enough theoretical foundation for this method. In this paper, we used the cited degree to distinguish the cited times, and divided the cited degree into four grades basing on the cited times: strong the citation intensity, medium citation intensity, week citation intensity and no citation intensity.

4.3.   Classification of label value

As mentioned above, the label value of texts can be divided into four categories, but these four types of label information need to be converted to numerical data. This section discussed the assignment of the four categories of label value. As we have defined the label value {keyword, subject category, text authority, citation degree} as $\{\mathbf{g}_1(\mathbf{X}), \mathbf{g}_2(\mathbf{X}), \mathbf{g}_3(\mathbf{X}), \mathbf{g}_4(\mathbf{X})\}$.

(1) Assignment of keywords. For the importance of keywords, we uses tf-idf method to assign this label value. Using r_i to express the key word of x_i, when the number of keywords is R, the label value of the keyword is expressed as $\mathbf{g}_1(\mathbf{X}) = \sum_R \mathrm{tf}(r_i) \times \mathrm{idf}(r_i)$, and the label value of the keyword between 0 and 1. Here, we assume that there is only one primary keyword in the text. The purpose of this processing is to find the recommendable text in a wider scope.

(2) Assignment of subject category. For the subject category, in general, most texts have gotten a clear distinction. We could use the "rule of thumb"[‡] to assign this parameter, which means, we could define the connection of two variables as the correlation strength. If the value of correlation strength equaled 0-0.05, it means non-correlation; if the value of correlation strength equaled 0.05-0.25, it means week correlation; if the value of correlation strength equaled 025-0.60, it means medium correlation; if the value of correlation strength equaled 0.6-1, it means strong correlation. According to the definition of the subject category, referring to the thumb rule, the text is related to the subject, the intensity of which is as follows: the original disciplines = 0.60; the related disciplines = 0.25; and other disciplines = 0.05. Under a special circumstance, some texts will appear in two or even three categories in the same time, then we should accumulate all of the related categories strength.

(3) Assignment of authority parameter. There was long-hold dogma that the definition of authoritative parameters had a strict standard and with no ambiguity. Therefore, we had directly assigned it with accurate numerical data. We also use the thumb rule to do the assignment. Based on the connection of the text and its authority of affiliated institutions, the correlation strength should be assigned as, core A = 1, core B = 0.65, core C = 0.25, general D = 0.05 and other = 0. What needs to be considered is that the authority of the text would change over time, that means, for two texts x_i and x_{i+1}, even though they had the same authority parameters, when the published time of x_i was earlier than x_{i+1}, then generally came to a authority value comparison as $x_i \geq x_{i+1}$. Here, let the original authority parameter x_i to be $\tilde{\mu}^i$, after adjusting the time parameter, its authoritative value should be $\mathbf{g}_3(\mathbf{X}) = \tilde{\mathbf{g}}_3(\mathbf{X}) \cdot t(\mathbf{X})$, here $t(\mathbf{X}) = t / (t+1)$.

(4) Assignment of cited time. The four levels of citations can also correspond to the strength of the four categories of thumb rules. Assuming that the four levels of texts are denoted as {Π1, Π2, Π3, Π4}, they denoted respectively the {strong citation intensity, medium citation intensity, week citation intensity and non-citation intensity}. After disposing the grade of Π4, the correlation coefficient of three remaining categories can be adjusted according to their mutual citation intensity. Specifically, according to the law of the famous economics budget allocation, namely the law of 60:30:10, we could

---

[‡] https://en.wikipedia.org/wiki/Rule_of_thumb

distribute the correlation strength according to this law and then adjust the citation correlation strength basing on the citation condition. The number of the strong correlation text accounts for 10% of the total number, and the number of the medium correlation text accounts for about 30% of the total number, and the number of the weak correlation text is 60%.

The citation intensity is adjusted as follows: when P1 refers to P2, P2 is adjusted to a strong citation intensity from the medium citation intensity, when P1 refers to P3, P3 is adjusted by the weak citation intensity to the medium citation intensity; when P2 refers to P3, then the weak correlation strength value of P3, then the weak correlation strength value of P3 is multiplied by P2. In this way, the citation intensity is assigned according to this law, and then the existing citation intensity is adjusted and assigned according to the citation situation of the text.

5.   Experiment

In this chapter, we first calculate the correlation coefficient of the text through the experiment, and then analyze the influence of our methods on the text recommendation. At last, we compare the difference between our method and other methods in text recommendation. Here, the method in this paper is abbreviated as TSLI method.

5.1.   Data Pre-Processing

Firstly, basing on the introduction in section 4.3, we use the tf-idf method to obtain the similarity of similar keywords. To save the calculation time, $tf(r_i)$ should be limited to the extraction of titles and abstracts for each text x_i, at the same time, we obtained the value of keywords $R = 4$. The value of each R would be discussed in the next section. Moreover, we should calculate the values of $g_2(\mathbf{X})$, $g_3(\mathbf{X})$, $g_4(\mathbf{X})$ with the method mentioned in section 4.3. Among the more than 30 thousands selected texts, there was a total of 26993 ones with no correlation, and these texts would not be considered in the TSLI method. For the rest texts, after being arranged in the descending order of citation degree, the rest qualified texts could be divided into strong, medium and weak correlation according to the 60:30:10 rules and their citation degrees were shown in figure 4.

As shown in figure 4, the degree of the correlation has been subject to a long tailed distribution, which indicted that some week correlation texts might also meet the preferences of users, so it was necessary to take these three kinds of correlation degrees into consideration all time. Therefore, we carried out the label values of more than 4 thousands texts to obtain the value they have brought to the user through the correlation analysis. Basing on the above condition, we randomly selected a text x_i, and we obtained the label value of its 31 cited text's original value and normalization value. The result of this experiment is shown in figure 5. This figure is the set of four kinds of label value of $\{g_1(\mathbf{X}), g_2(\mathbf{X}), g_3(\mathbf{X}), g_4(\mathbf{X})\}$. Figure 6 showed the label value of $\mathbf{x}_i$. As all the four vectors $\{g_1(\mathbf{X}), g_2(\mathbf{X}), g_3(\mathbf{X}), g_4(\mathbf{X})\}$ have different metrics and units, and that would

always lead to an impact on the results of data analysis. In order to eliminate the dimensional effect between the indexes, it is necessary to standardize the data and make
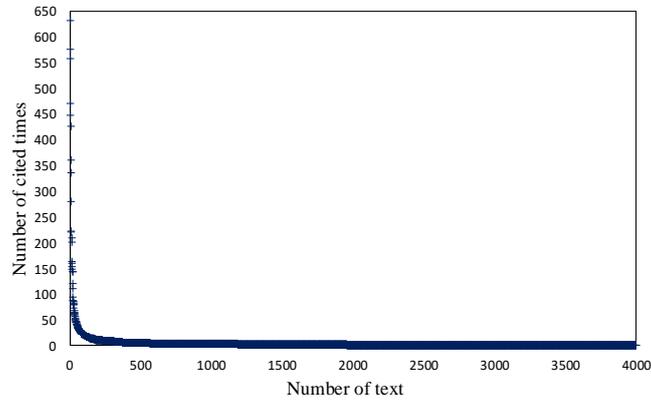
Fig. 4. The citation times of the three kinds of efficient citation degree

the comparability between the data indexes. After the original label information being standardized by the data, the indexes are in the same order of magnitude, so that the whole citation network can be operated and compared comprehensively.

In this paper, we used the Z-score standardization method. We have given the mean and standard deviation of the original data, and carried out the standardization of the data.

The processed data are in accordance with the standard normal distribution, with mean = 0 and variance = 1. The transformation function is:

$$\mathbf{g}^*(\mathbf{x}_i) = \frac{\mathbf{g}(\mathbf{x}_i) - \mu(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)}$$

μ is the sample mean, σ is the sample standard deviation. The converted $\mathbf{x}_i$ is shown in figure 6.
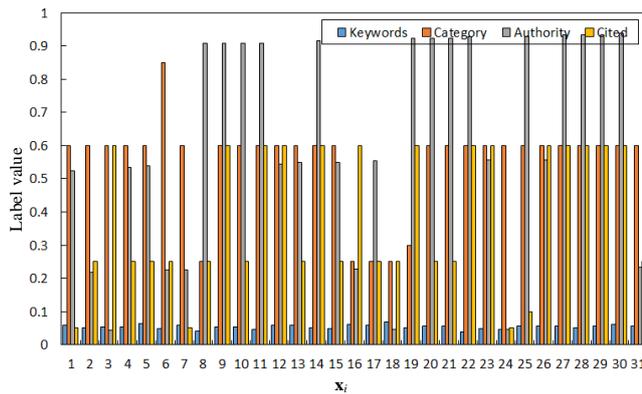
Fig. 5. The label value of text x_i
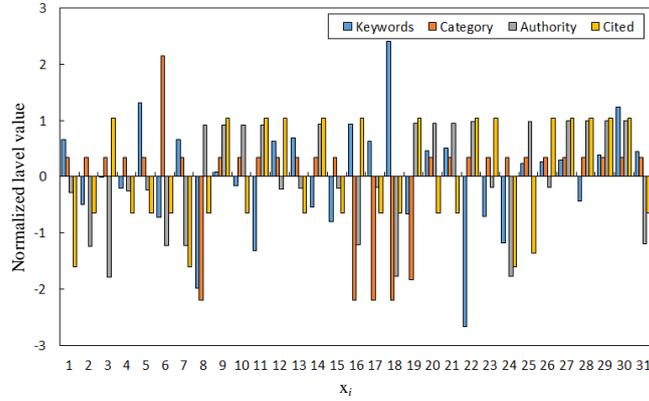
Fig. 6. Standardization of label value of text x_i

The reason why we had used the Z-score instead of using other standard methods is that the standardized data were subject to the standard normal distribution, after using this method for normalization, the vast majority of samples would be concentrated near to the average, which was conducive to the selection of samples. So with the same method, we could also select the text $\mathbf{x}_j$ referred by x_j. In our experiment, we selected out 16 qualified texts from the whole cited texts of x_j, these 16 texts contained some texts in $\mathbf{x}_i$ at the same time. The label value of $\mathbf{x}_j$ is shown in figure 7.

In this way, we finally found out all of the correlation coefficients between the text and its references. After we obtaining the values of $\mathbf{g}(\mathbf{x}_i)$ and $\mathbf{g}(\mathbf{x}_j)$, we could subsequently calculate the values of $\mathbf{s}(x_i\mathbf{x}_i)$ and $\mathbf{s}(x_j\mathbf{x}_j)$ according to the formulas (6) and (8).

5.2.   Calculation of Similarity

We have discussed the method of obtaining correlation coefficient above, but there should be uncertain parameters also needed to be discussed, which included the number of the keywords R, the constraint parameter of label value $\boldsymbol{\alpha}$ and the number of the recommendable texts k.

First, consider the case where the number of keywords R and the parameter $\boldsymbol{\alpha}$ affect each of the importance coefficients. The influence of the keyword is more important than other factors in the text recommendation method, which can be seen from most of the pretreatment process. In this paper, we had defined the influence degree of those four types of label value $\{\mathbf{g}_1(\mathbf{X}), \mathbf{g}_2(\mathbf{X}), \mathbf{g}_3(\mathbf{X}), \mathbf{g}_4(\mathbf{X})\}$ as $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$, which was exactly the parameter $\boldsymbol{\alpha}$ in formula (8), and with $\alpha_1 > \max \{\alpha_2, \alpha_3, \alpha_4\}$. Besides of the α_1, we unified the constraint parameters of the other three parameter as $\alpha_2 = \alpha_3 = \alpha_4$, and then we converted $\{\alpha_1 \cdot \mathbf{g}_1(\mathbf{X}), \alpha_2 \cdot \mathbf{g}_2(\mathbf{X}), \alpha_3 \cdot \mathbf{g}_3(\mathbf{X}), \alpha_4 \cdot \mathbf{g}_4(\mathbf{X})\}$ into $\{\alpha_1 \cdot \mathbf{g}_1(\mathbf{X}), \alpha_2 \cdot \mathbf{g}_2(\mathbf{X}), \alpha_2 \cdot \mathbf{g}_3(\mathbf{X}), \alpha_2 \cdot \mathbf{g}_4(\mathbf{X})\}$ and assuming $\alpha_1 + \alpha_2 = 1$. This assumption is to highlight the key words in the label value of the significant position, which is also in accordance with the existing traditional recommend- ation method. In addition, as for the influence of

keywords, the number of keyword R is also a non-negligible parameter, it would directly affect the assignment of keywords.
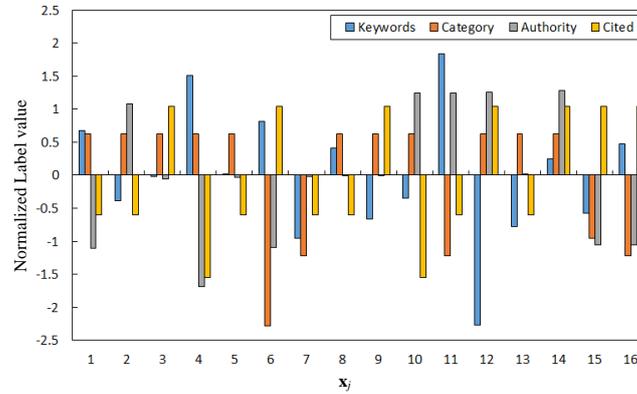
Fig. 7. Standardization of label value of text x_j

Next, we compared the correlation coefficient of x_i and its references x_i^l with the correlation coefficient of x_j and its references $\mathbf{x}_j$. Then we were able to determine the value of R and α based on the cosine similarity calculated with formula (8). We respectively valued (α_1, α_2) as $\{(0.9, 0.1), (0.8, 0.2), (0.7, 0.3), (0.6, 0.4)\}$, R as $\{2, 3, 4, 5, 6\}$ to do the test and obtain the similarity value, the results were shown in figure 8. The value of (α_1, α_2) is $\{(0.9, 0.1), (0.8, 0.2), (0.7, 0.3), (0.6, 0.4)\}$, and the value of R is $\{2, 3, 4, 5, 6\}$, which is the empirical value to verify the validity of the experiment. Considering that the values of (α_1, α_2) and R have been defined in the theory of the previous chapter, only these three parameters need to be taken as the values in the paper.

Figure 8 showed the effects of the different parameters (α_1, α_2) and R on the sim value. When $R=2$ and $R=5$, for any value of (α_1, α_2), the sim value seemed more decentralized than the other three distribution results. When (α_1, α_2) are valued as $(0.9, 0.1)$ and $(0.8, 0.2)$, their sim values also appeared decentralized, which indicated that both of the $R=2$ and $R=5$ were not the high discernible parameter in determining the recommendable text. As for $R=7$, the change of standard deviation became fairly obvious corresponding to the different value of (α_1, α_2). These results manifested that once the value of (α_1, α_2) or R do not completely adapt to certain types of samples, the recommendation method would be very unstable. As shown in the figure, when $R = 4$, no matter what the value of (α_1, α_2) was, the cosine similarity value has changed significantly, moreover, the change of standard deviation was still obvious when compared with other values of R. When $(\alpha_1, \alpha_2) = (0.6, 0.4)$, which showed that the obtained cosine similarity values were concentrated, all of the standard deviation of different R value could be differentiated obviously. Therefore, in this paper we valued (α_1, α_2, R) as $(0.6, 0.4, 4)$ to do the recommendation.

The similarity measure in text recommendation refers to calculating the similarity between texts. The larger the similarity value, the smaller the difference of text. There are many methods to calculate the similarity, cosine similarity is a mature method. For many different texts to calculate the similarity between them, a good way is to map the

labels in these texts to the vector space, form the mapping relationship between the labels and the vector data, and judge the similarity of the text by calculating the difference value of one or more different vectors.
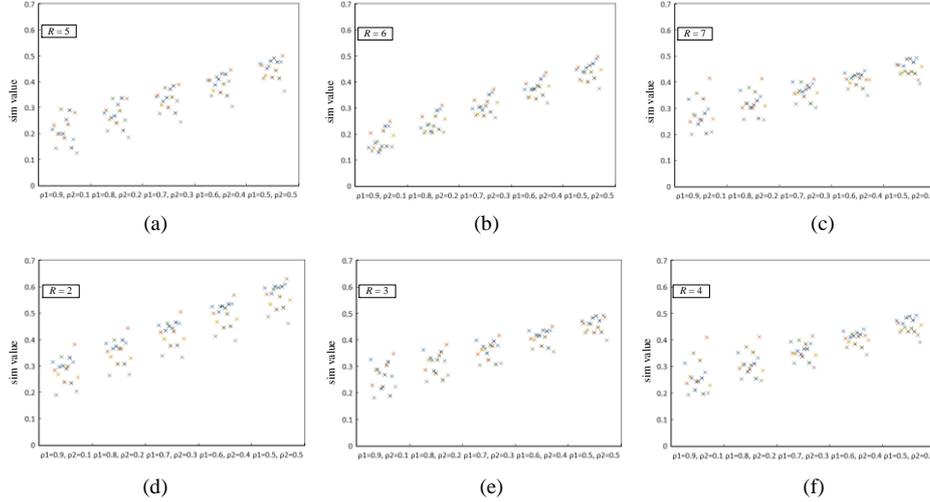
Fig. 8. The influence of different ($\alpha1$, $\alpha2$) and Ron the sim value.

5.3.   Comparison of Methods

In order to verify the correlation between the recommended texts, we compared our method with other five text recommendation methods. Firstly, we defined:

$$\text{Precision} : P@k = \frac{R_C@k}{R_T} \times 100\%$$

$$\text{Recall} : R@k = \frac{R_C@k}{k_T} \times 100\%$$

$$\text{F-measure} : F1@k = \frac{P@k \times R@k \times 2}{P@k + R@k} \times 100\%$$

Here, R_C is the number of recommendable texts, R_T is the total number of texts obtained by the user, k_T is the number of recommended texts.

The KMR [25] method is a keyword matching method that can only recommend articles with similar keywords to users by comparing the keywords of different articles.

The SoREC [26] method uses the shared user feature space to combine the social relations with the score information. By combining the two pieces of information, the SoREC identifies the users who are similar in the score and have social relations to make recommendations.

The SARSP [27] method divides users with similar interests into one class, and then the users in this class recommend each other.

The ItemKNN [28] method is to use the item's content / attributes as a vector to find a similar relationship between the users to realize the recommendation process

All this adds up to a true that our method got the most significant advantages. As mentioned before we defined the parameters as: α_1=0.6, α_2=0.4, R=4. The result is shown in figure 9, there were 5-50 texts in this experiment, our TSLI method has showed a better precision rate and recall rate than other methods. At mean time, for the average precision rate and average recall rate, there are respectively an improvement of 8.63% and 5.25%, moreover, their maximum appreciation have come to 12.76% and 7.25%. These results indicated that TSLI was able to carry out a better recommendation result for different number of texts.

(a)   (b)   (c)
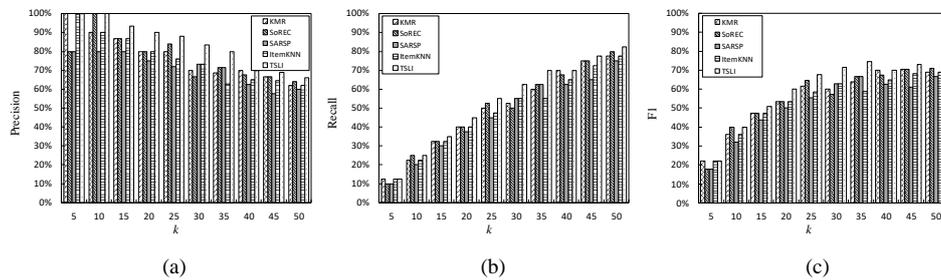
Fig. 9. Comparison of accuracy rate, recall rate and F1 value for different quantity of text.

6.   Conclusion

The correlation analysis of these similarity texts has become one of the main ideal of the text recommendation to meet the needs of different users. When the user has captured a certain text of his interest, he would like to get a series of similar texts related to it rather than searching for a large number of texts. We always capture the content of the text information basing on a number of label values, such as the research area, keywords, cited time, etc.

However, these data are very abstract, it is not easy for users to obtain the text they want through this information, and the user needs the text of its label information to be no overlap.

Basing on the information mentioned above, to model the recommendation process better and to reveal the potential influences of the text correlation on the result of the recommendation, this paper analyzed the similarity between texts and elaborated the recommendation processes from three aspects. (1) Defined and classified the label value of text. Each type of label information is introduced into the timing relationship and assigned. (2) Using the difference method to arrange the label value of a text xi and its cited text in a chronological order, then a set of correlation coefficients is formed for the text xi and the label information for each text is cited by x_i. (3) After setting the parameters of correlation coefficient, the consistent text was recommendable to the user basing on the comparison of the citation relationship of the text. Our experimental have effectively testified the quality of the recommended text by our reliable method.

References

1. Zhou J, Zeng A, Fan Y, et al. Identifying Important Scholars via Directed Scientific Collaboration Networks. Scientometrics, Vol. 114, No. 3, 1327-1343. (2018)
2. Ren Z M, Mariani M S, Zhang Y C, et al. Randomizing Growing Networks with a Time-Respecting Null Model. Physical Review E, Vol. 97, No. 5, 052311. (2018)
3. Liu X, Zhang J, Guo C. Full-text Citation Analysis: A New Method to Enhance Scholarly Networks. Journal of the Association for Information Science and Technology, Vol. 64, No. 9, 1852-1863. (2013)
4. Deng S, Huang L, Xu G, et al. On Deep Learning for Trust-Aware Recommendations in Social Networks. IEEE Transactions on Neural Networks & Learning Systems, Vol. 28, No. 5, 1164-1177. (2017)
5. Gang, L., Hanwen, Z. (2020) "An Ontology Constructing Technology Oriented on Massive Social Security Policy Documents", Cognitive Systems Research, 60, pp. 97-105.
6. Shen X L, Li Y J, Sun Y. Wearable Health Information Systems Intermittent Discontinuance: A Revised Expectation-Disconfirmation Model. Industrial Management & Data Systems, Vol. 118, No. 3, 506-523. (2018)
7. Bjork S, Offer A, Söderberg G. Time Series Citation Data: the Nobel Prize in Economics. Scientometrics, Vol. 98, No. 1, 185-196. (2014)
8. Shen X L, Li Y J, Sun Y. Wearable Health Information Systems Intermittent Discontinuance: A Revised Expectation-Disconfirmation Model. Industrial Management & Data Systems, Vol. 118, No. 3, 506-523. (2018)
9. Suominen H. Guest Editorial: Text Mining and Information Analysis of Health Documents. Artificial Intelligence in Medicine, Vol. 61, No. 3, 127-130. (2014)
10. Xue, Q., Zhu, Y., & Wang, J. (2019). Joint Distribution Estimation and Naïve Bayes Classification under Local Differential Privacy. IEEE transactions on emerging topics in computing, 1.
11. Gupta S, Varma V. Scientific Article Recommendation by Using Distributed Representations of Text and Graph. Proceedings of the 26th International Conference on World Wide Web Companion. International World Wide Web Conferences Steering Committee, 1267-1268. (2017)
12. Tellez E S, Moctezuma D, Miranda-Jiménez S, et al. An Automated Text Categorization Framework Based on Hyperparameter Optimization. Knowledge-Based Systems, Vol. 149, 110-123. (2018)
13. Mäntylä M V, Graziotin D, Kuutila M. The Evolution of Sentiment Analysis - A Review of Research Topics, Venues, and Top Cited Papers. Computer Science Review, Vol. 27, 16-32. (2018)
14. Caruccio L, Deufemia V, Esposito S, et al. Combining Collaborative Filtering and Semantic-Based Techniques to Recommend Components for Mashup Design. Computational Intelligence for Semantic Knowledge Management. Springer, Cham, 25-37. (2020)
15. Huang S, Yu Y, Xue G R, et al. TSSP: Multi-Features Based Reinforcement Algorithm to Find Related Papers. Web Intelligence & Agent Systems, Vol. 4, No. 3, 271-287. (2006)
16. Harman D. Information Retrieval: The Early Years. Foundations and Trends® in Information Retrieval, Vol. 13, No. 5, 425-577. (2019)
17. Adams B, Phung D, Venkatesh S. Social Reader: Towards Browsing the Social Web. Multimedia tools and Applications, Vol. 69, No. 3, 951-990. (2014)

18. Wrixon A, Belov A, Keller M, et al. Static Timing Analysis with Improved Accuracy and Efficiency: U.S. Patent Application 10/002,225. 2018-6-19. (2018)

19. Thorne J, Vlachos A. Automated Fact Checking: Task Formulations, Methods and Future Directions. Arxiv Preprint Arxiv:1806.07687, (2018)

20. Ling Wu, Chi-Hua Chen*, Qishan Zhang, "A Mobile Positioning Method Based on Deep Learning Techniques," Electronics, 8, no. 1, Article ID 59, January 2019.

21. Sun, S., Kadoch, M., Gong, L., & Rong, B. (2015). Integrating network function virtualization with SDR and SDN for 4G/5G networks. IEEE Network, 29(3), 54-59.

22. Fabisiak, L. 2018. "Web Service Usability Analysis Based on User Preferences," Journal of Organizational and End User Computing (30:4), pp. 1-13.

23. Bi, Zhongqin; Dou, Shuming; Liu, Zhe; Li, Yongbin. A Recommendations Model with Multiaspect Awareness and Hierarchical User-Product Attention Mechanisms. Computer Science and Information Systems, 2020, 17(3), pp. 849-865.

24. Lv, Zhihan, Dongliang Chen, Ranran Lou, and Qingjun Wang. "Intelligent edge computing based on machine learning for smart city." Future Generation Computer Systems (2020).

25. Zhao W, Wu R, Liu H. Paper Recommendation Based on the Knowledge Gap between a Researcher's Background Knowledge and Research Target. Information Processing & Management, Vol. 52, No. 5, 976-88. (2016)

26. Du M, Vidal J M, Markovsky B. SOREC: A Semantic Content-Based Recommendation System for Parsimonious Sociology Theory Construction. 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService). IEEE, 138-144. (2019)

27. Asabere N Y, Xia F, Meng Q, et al. Scholarly Paper Recommendation Based on Social Awareness and Folksonomy. International Journal of Parallel Emergent & Distributed Systems, Vol. 30, No. 3, 211-232. (2015)

28. Wang X, Sheng Y, Deng H, et al. Top-N-Targets-Balanced Recommendation Based on Attentional Sequence-to-Sequence Learning. IEEE Access, Vol. 7, 120262-120272. (2019)

Yi Yin, born in 1983, Ph. D., His research interests include machine learning, recommendation system, data mining and neural networks.

Dan Feng, born in 1970. Ph.D., professor, Ph.D. supervisor. Her research interests include computer system architecture, parallel processing, fault tolerance theory, disk array architecture, and computer storage system.

Zhan Shi, born in 1976. Ph. D., associate researcher. His research interests include massive information storage, storage service and storage management.

Lin Ouyang, born in 1974, Ph. D., His research interests include distributed system, image processing and neural networks.

Message Propagation in DTN
Based on Virtual Contact of Behavior Model

Ho-Hsiang Chan[1] and Tzu-Chieh Tsai[2]

[1]Department of Computer Science
National Chengchi University
Taipei, Taiwan
100753503@nccu.edu.tw
[2]Department of Computer Science
National Chengchi University
Taipei, Taiwan
ttsai@cs.nccu.edu.tw

Abstract. Delay Tolerant Network (DTN) is a kind of network structured to deliver message intermittently. Network connections are not persistent between nodes, instead they must rely on nodes making geographic location movements to incur contact with other nodes and establish intermittent communication sessions to allow messages delivery. We will refer to encounters via geographic location movements as "physical contact." Many DTN researches mainly focus on message delivery via physical contact. However, this paper believes that in a realistic environment, encounters between nodes not only happen geographically in nature, but also occur virtually in cyberspace. When both nodes go online on the same social media platform, it is an encounter we refer as *virtual contact*. How messages deliver for virtual contact is store-post-and-forward, just like what happens in a DTN, but it is no longer restrained by geographical locations. This paper considers a scenario in which nodes make virtual contact in cyberspace and incur message delivery based on their own behavior patterns. The verifying experiment is conducted using both survey and simulation. First of all, we handed out questionnaires for students to fill out. The questionnaire inquired them to rank their most frequent activities performed on social media platforms. According to the responses, we conclude the top 3 frequent activities when the students use social media platforms and classify them into 3 groups according to a weighted behavior pattern scheme. The classification includes *Social Group*, *Read-Only* Group and *Interest Group*. It does not matter which group a student is assigned to. In the simulation, he or she will get to decide whether to deliver/receive messages or not based on a randomized selection on 3 behavior pattern. Finally, we analyze the simulation result to determine how messages propagated in different behavior pattern groups. It is derived from the simulation that to quicken message propagation, directing messages to one of the behavior groups yields the maximize benefits. This provides the basis for further researches on collecting data of desired scenarios to establish respective propagation models.

Keywords: Delay Tolerant Network, physical contact, virtual contact, behavior pattern

1.   Introduction

1.1   Background

Nowadays the Internet is flooded with information, and it comes in different flavors. As different kinds of information spread across the world each day, people are busy receiving in and sending out information on various social platforms. For example, economic news, political comments, and sport coverage are among the most covered ones and get forwarded repetitively. To understand how quickly information is dispersed, it is necessary to evaluate the number of forwards/shares made and the frequency of forwarding done by each social platform user. As cellular network progresses and mobile smart devices popularizes, people can log in to social platforms to forward messages agilely whenever they can and wherever they want. Consequently, if there are urgent messages that must be known by the masses in a short time, transferring them through other users on social platforms or on the Internet will likely to induce higher probability of quick message propagation.

Delay Tolerant Network (DTN) is an instance of Opportunistic Network Environment, in which no single route serves as a persistent end-to-end connection, and it requires users carrying smart mobile devices to move between same geographic locations to have the opportunities of sending out messages. The majority of previous researches focus on exploring methods of message delivery and forwarding via geographic location encounters. We have also proposedan approach, NCCU Trace Data [16] which involves collecting the data of students' real geographical location movements in campus environment. When a student meets with other students, there are opportunities of forwarding messages further. However, this paper believes that current DTN researches' focuses on physical encounter scenarios rely on people's move to triggering encounters with other peoples, which no long conforms with the real-world environment.

With the advancements of Information and Communication Technology, we can easily connect to social media platforms (i.e. Facebook, Twitter, Whatsapp, etc.) via mobile networks to share information with other users. This way of transferring messages is similar to the case of DTN. The further messages are delivered, the less likely for receiving users to be online. Thus, if a user sends out messages when others happen to be online at the same time, he or she still needs to transfer messages to social media platforms before these messages are delivered. However, even if receiving users are offline, messages still remain on social media platforms waiting to be transmitted to the currently offline users when they get online.

1.2   Virtual Contact

This research believes that above-mentioned scenario is a case of *virtual contact*, as described by Figure 1 below. As illustrated by the figure, virtual contacts happen when social media users are physically located in distant locations where physical contacts are not possible, yet they can achieve virtual space encounters in cyberspace constructed on the Internet disregarding the boundaries of space and time as long as they were once

online on common social medias. Virtual contact on social platform align with one of the characteristics of DTN, store-carry-and-forward. Because users are likely to get online at different time, they are able to deliver messages on social media platforms when online, and other receiving users will accept the messages once online. If both the message-sending user and the message-receiving user are online at the same time, real-time transferring will be carried out. Message delivery is no longer limited to the classic scenario that users must be located in the same geographic region. Owing to the convenience of social media platforms, users can make *virtual contact* with each other at different geographic locations, so that the message-carrying users will be able to deliver them to the others.
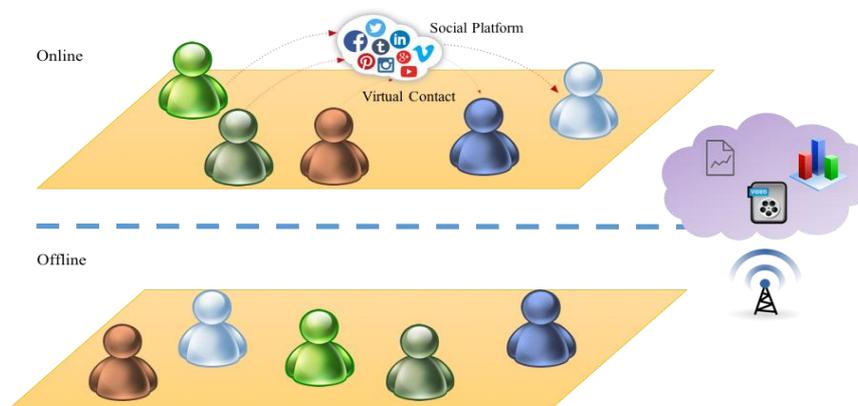
Fig.1. Overview of the virtual contact diagram

1.3   Motivation

Previous works on message delivery mostly neglect to consider message senders' delivering behavior. In a real-world environment, everyone acts out of his or her own free will and reacts to forward or receive messages. For example, some users tend to passively accept messages without any intent to share or forward them, while other users are more likely to share received messages in the hopes that even more people know about them.

This paper believes that the scenario of virtual contact is just like the characteristic of store-carry-and-forward. When users become online on social media platforms, the message-carrying users are granted with the opportunities of meeting other users, and making direct message delivery. Besides, we think people would behave differently according to their current moods and environments. For instance, one may prefer browsing messages to sharing them; one may also desire to share interesting messages to other internet users of similar interests, etc. Thus, this paper investigates how users' daily behavior patterns can influence the way messages are delivered. Considering a message delivery scenario, when an urgent message shall be sent in a way that the more people learn the better it is, we suggest a better accommodating message propagation scheme: transmitting messages in accordance with messages' properties or users' behaviors when using social media platforms.

1.4　　Purpose

This research proposes an approach that message delivery should be determined based on users' individual behaviors in a network environment. Regardless of the message delivery and acceptance, everyone will ultimately choose to deliver or receive messages based on their own interests, which aligns with the real-world scenario on message propagation. Furthermore, messages categorized to be of similar interests would make carriers behave alike. This paper attempts to resolve the problem of finding the most appropriate person to propagate messages with interests in respective fields, thus allowing them to be received by as many people as possible within a fixed amount of time. Lastly, we utilize message propagation behaviors to conduct clustering, and, therefore, are able to efficiently find a model for fast message propagation. When there is a certain type of messages needed to be dispersed quickly, we can find groups most suitable for fast message propagation via the above-mentioned message propagation model based on behavior clustering to achieve better efficiency.

2.　　Related Work

Previous social network researches based on DTN define social community according nodal geographic locations and chances of making encounters with other nodes to decide their closeness in terms of social network distances, then develop strategies of message delivery based on it. SimBet[4] utilizes nodal betweenness to computer centrality and similarity for each node to help making decision on which nodes to deliver messages.

This article believes that nodes make frequent encounters in terms of geographical locations, then there exists a social network relationship between them. It further proposes the more mobile nodes are, the nodes with better utilities can be to help deliver message. However, one drawback of such design is that messages are likely to be centralized on nodes with better utilities. If there exists less active nodes in the network, they, in the worst case scenario, might never get any message. Bubble rap[12] thinks it is better to take a prolonged observation on each node's encounters via geographical location and turn the observation results into a simulated social network. Each node should belong to at least one community or multiple communities, and each node should have information on global ranking and local ranking of the whole network system. Routing algorithm then utilizes estimated community and centrality of the social network. According the ranking done by the global community, messages are forwarded to the node with highest centrality in the community until the target node and the forwarding node belong to the same community. SANE routing [8] depends otherwise on interests and similarities as estimation basis for message delivery. The author believes that nodes with similar interests are likely to make encounters with each other. If a message is to be forwarded to the target node, forwarding the message to the nodes with similar interests will yield a better transmission efficiency. A previous research by us, NCCU Trace Data[16] is about collecting students' movements in a campus. As a laboratory collective effort, we develop a APP which can be utilized to track students' movement traces when attending classes at school, as showed in Figure 2, and propose a message delivery method via similarities between people's interests in

a real-world campus environment. This method performs better than traditional routing algorithms.

Fig.2. NCCU Trace Data of the screenshot

Besides, our previous research [17] also suggests that people will often move, according to their interests, to buildings of similar interest properties. It may happen that people with similar interests are likely to have similar routes of geographical location displacements. The message delivery method based on interest properties has a better performance compared to other routing algorithms.

The aforementioned works are all about investigating the fact that when nodes make frequent encounters via geographical location movements, there may exist social network relation and exhibit close betweenness among nodes. If a message is to be delivered, this leads to better message delivery ratio. However, this kind of social network estimation method tends to be shallow. Frequent encounters via geographical location might not necessarily mean close betweenness, but only show that nodes happen to be neighbors or have the same moving directions. This paper believes that personal information should be utilized to confirm whether there exists social network relationships between nodes, then making estimates of nodes' social network relationships to develop a method of message delivery that aligns better with real-world environments. Table 1 is the comparison among different strategies adopted by referenced works.

In a conventional Delay Tolerant Network environment, encounters via geographical locations introduce opportunities for message propagation or dispersion. Compared to the conventional methods, this paper is different in the way that opportunities for message dissemination, as we proposed, occur in cyberspace, so methods suggested by previous works are merely inspirations for our approach for that there is temporarily no similar method comparable to this research.

Table 1. Comparison table of the related work

Related work	Forward Strategy	Mobility Model	Characteristic
SimBet[4]	Node Utility	MIT Reality	betweenness Centrality
Bubble rap[12]	Node Utility	MIT Reality Cambridge Infocom06	ego-central
SANE [8]	Node Utility	Infocom06	Social-aware
NCCU Trace Data [16]	Message Interest Utility	NCCU Trace Data	Interest-aware

3.   Proposed Approach

We are continuing a previously proposed work, NCCU Trace Data [16]. This research is part of the project responsible of NCCU Trace Data collection from 115 participants' Facebook online histories and friend lists, except that we were unable to collect online histories properly from 11 of them. This makes a total of 104 participants' online histories, all of whom were notified and agreed with how we handle their data. Moreover, we asked each of the 104 experiment participants to fill out a questionnaire aiming to collect personal information, interests, and rankings of frequent Facebook activities, etc. These activities include: (1) sharing messages with friends on social media platforms (referred it as *Social*); (2) sharing articles in the groups on social media platforms (referred as *Interest*); (3) refusing to share messages while only receiving them (referred as *Read Only*). According the rankings done by the participants, we would assign the most frequent activities a weight of 3, the second most frequent one a weight of 2, and the lease frequent one a weight of 1. Each participant is required to fill out at least 1 frequent activity. Part of the original data are presented by Table 2.

Table 2. Behavioral ordering

User_ID	Social	Interest	Read Only
1	3	1	1
2	3	1	1
3	3	2	1
4	3	1	1
5	3	1	1
6	3	1	2
7	3	2	1
8	3	1	1
9	1	1	3
10	3	1	2
11	3	1	3
12	3	1	1

According to the answers filled by the participants, this research would utilize K-means algorithm to achieve clustering based on each one's behavior weights, divide them into 3 groups, and derive the classification scheme from each group's behavior weight features. The results are presented in Table 3.

Table 3. K-means algorithm classification

User_ID	Group	Distance	Social	Interest	Read Only
1	0	1.337	3	1	1
2	0	1.337	3	1	1
3	0	1.669	3	2	1
4	0	1.337	3	1	1
5	0	1.337	3	1	1
6	0	0.492	3	1	2
7	0	1.669	3	2	1
8	0	1.337	3	1	1
9	1	2.642	1	1	3
10	0	0.492	3	1	2
11	1	2.717	2	1	3
12	0	1.337	3	1	1
13	0	1.669	3	2	1
14	2	2.492	1	3	1

K-means algorithm classifies the participants into 3 different groups. The first group numbered with 0 consists of 88 people. The second group numbered with 1 consists of 6 people. The third group numbered with 2 consists of 10 people. Table 4 below displays part of the data for Group 0.

Table 4. K-means algorithm classification

User_ID	Group	Distance	Social	Interest	Read Only
1	0	1.337	3	1	1
2	0	1.337	3	1	1
3	0	1.669	3	2	1
4	0	1.337	3	1	1
5	0	1.337	3	1	1
6	0	0.492	3	1	2
7	0	1.669	3	2	1
8	0	1.337	3	1	1
10	0	0.492	3	1	2

As the data K-means algorithm classified to be group 0 show, participants prefer to share messages on social media platforms, described by their responses to the questionnaires. Thus, this research defines the first group as Social Group. Although classified as part of the Social Group, participants would obviously engage in activities other than sharing messages with friends on social media platforms. They could also choose to share articles in their groups or simply receive messages without sharing them. The second group classified by K-means algorithm, as showed by their responses to the questionnaires, all prefer to only receive messages without sharing, so they are

defined as Read Only Group. The third group classified by K-means algorithm stated in their questionnaire responses that they all prefer to share articles in their groups, so they are defined as Interest Group.

Figure 3 below is a visualization of Social Group which this research attempts to represent with a tree diagram. Each route consists of three nodes representing different behavior preferences, as the participants can only engage in one certain activity during a fixed amount of time. The weights of the activity nodes would change according to the routes connecting them. The first route shows that 88 of the participants would prioritize to share messages with friends on social media platforms, 37 of them would secondly prioritize to share messages with their groups, and these 37 people are least likely to only receive messages without sharing on social media platforms. The chance of traveling along the first route is 42%. The second route shows that 88 of the participants would prioritize to share messages with friends on social media platforms, 51 of them are least likely to share messages with groups on social media platforms, and 16 of them secondly prioritize to receive messages without sharing them on social media platforms. The chance of traveling along the second route is 18%. The third route shows that 88 of the participants would prioritize to share messages with friends on social media platforms, 51 of them are least likely to share message with groups on social media platforms, and 35 of them are also least likely to only receive messages without sharing on social media platforms. The chance of traveling along the third route is 40%. Based on the descriptions above, we can conclude that if there is a message to be received or delivered, users would have 42% chance to choose the first route, 18% chance to choose the second route, and 40% chance to choose the third route while each route has different weights according to the activity preferences it represents. Each activity carries a weight. The higher the weight it is, the more likely the corresponding activity to be performed. The actual resulting route will randomly select one of the activities to be the basis of propagating or receiving messages. Figure 4 and Figure 5 are the visualizations of Read Only Group and Interest Group, respectively.
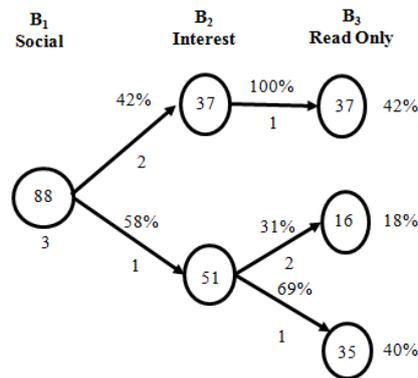
Fig.3. Social Group
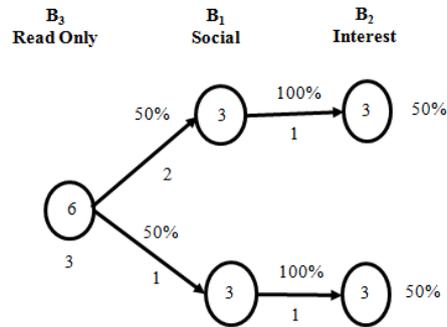
Fig.4. Read Only Group
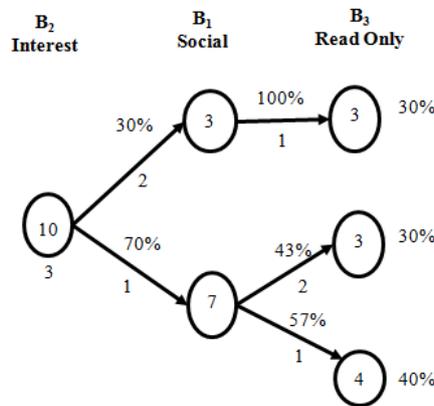
Fig.5. Interest Group

Like explained above, this research classifies the behavior patterns into three categories: (1) receiving messages from or propagating messages to social media friends, referred as Social Behavior; (2) receiving interest-provoking messages or propagate messages to interested users, referred as Interested Behavior; (3) only receiving messages without sharing to other users, referred as Read-Only Behavior.

(1) Message receiving and propagating for Social Behavior: users log into social media platforms, gain access to their friend lists of respective platforms, and invoke actions to receive or deliver messages according to the lists. This behavior pattern captures the scenario that whether users are interested at the messages or not, they will receive messages from their friends and propagate received messages to others. Such scenario is visualized by Figure 6.
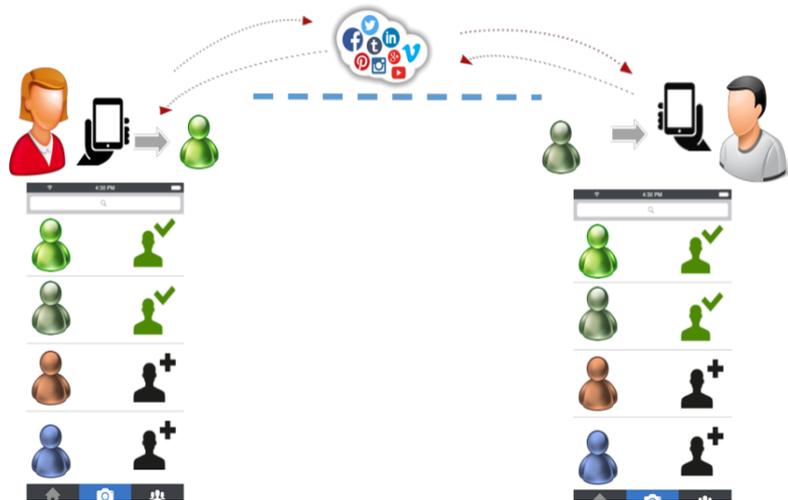
Fig.6. Message receiving and propagating for Social Behavior

(2) Message receiving and propagating for Interested Behavior: When a user U_X gets online on social media platforms, this behavior pattern captures the scenario that no matter users are friends of each other or not, they would receive interested messages or deliver interested messages to other users. Possible interest properties include: sports, reading, social activities, artistic events, community services. When a user receives messages, he or she will determine whether messages' interest properties $M_K(I_V)$ correlate with his or her own interests. To calculate its correlation with the interests, we rely on Cosine similarity as the basis. On the other hand, message delivery also depends on Cosine similarity as the basis to calculate its correlation with the interests. As shown in formula (1) below:

$$\text{Cos}\big(U_X(I_V), M_K(I_V)\big) = \frac{U_X(I_V)\cdot M_K(I_V)}{||U_X(I_V)||\cdot||M_K(I_V)||} . \tag{1}$$

(3) Message receiving and propagating for Read-Only Behavior: When users of this behavior pattern gets online on social media platforms, they will receive messages delivered from their friends or interested messages, but they would not deliver any message at all to others.

This research summarizes the mechanism of how virtual contact on social media platforms triggering receiving and delivering message as showed by Figure 7. It is further explained below:

(1)   A user U_x relies on his or her hobby when using social media platforms to decide the moment of time he or she gets online.
(2)   The user then confirms he or she shall be categorized as social group, interest group, or read-only group.
(3)   At the message-receiving stage, the user will decide the action to be taken at this point of time according to a randomized probability.

(4) After receiving a message, the users then switches to the message-propagating stage.

(5) At the message-propagating stage, the user will decide again the action to be taken at this point of time according to a randomized probability.

(6) After executing both message-receiving and message-propagating stages, the users then gets offline on the social media platform.
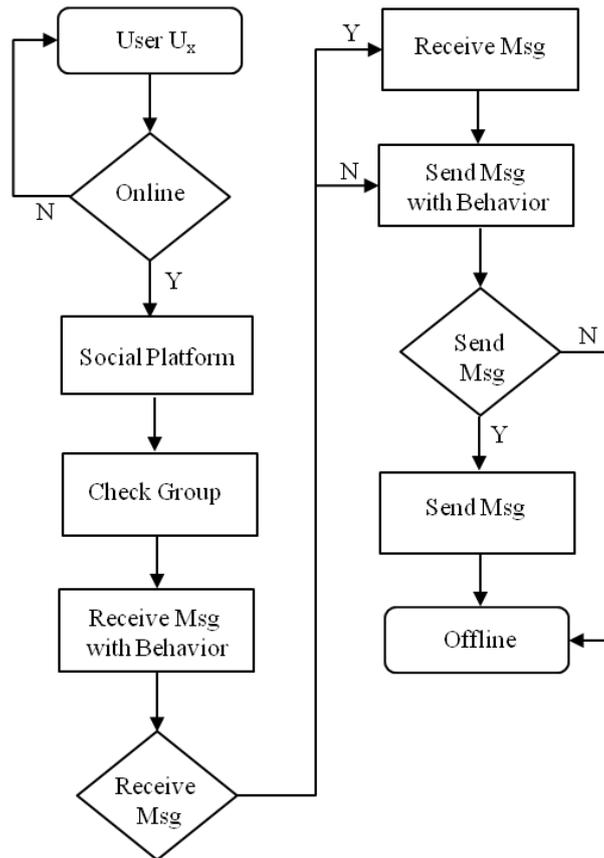
Fig.7. The Approach

In this research, we propose a message receiving and propagating mechanism constructed using the user behavior algorithm described below. When user goes online, the user will decide the action to be taken at this point of time according to a randomized probability, which is described by Algorithm 1. Message receiving and propagating mechanism is presented in Algorithm 2:

Algorithm 1 Checkout the Behavior.

Input: a dataset of User $U=\{u_1, \ldots, u_n\}$, Behavior $B=\{b_1, \ldots, b_3\}$, User Behavior Tree $T=\{t_1, \ldots, t_3\}$, Group $G=\{g_1, \ldots, g_3\}$

Output: User Behavior

1.   **foreach** $u_i U$do   /* Receive Message*/
2.   **for** (x, 1 to |G|)
3.   **if** $u_i g_x$
4.   **select** one of route from Tby
5.   **if** random(1,n)$<t_{gx}^{r1}$. leaf $\vee$
6.   **switch** (random(1,j))
7.   **case** $<t_{gx}^{r1}$. weightofb_1
8. u_i belongs to (t_{gx}^{r1}. behaviorofb_1)
9.   **case** $<t_{gx}^{r1}$. weightofb_x + t_{gx}^{r1}. weightofb_2
10. u_i belongs to (t_{gx}^{r1}. behaviorofb_2)
11.   **case** $<t_{gx}^{r1}$. weightofb_1 + $\cdots$ +t_{gx}^{r1}. weightofb_3
12. u_i belongs to (t_{gx}^{r1}. behaviorofb_3)
13.   **else if** random(1,n) $<=t_{gx}^{r1}$. leaf $\vee$ +t_{gx}^{r2}. leaf $\vee$
14.   **switch** (random(1,j))
15.   **case** $<t_{gx}^{r2}$. weightofb_2
16. u_i belongs to (t_{gx}^{r2}. behaviorofb_2)
17.   **case** $<t_{gx}^{r2}$. weightofb_2 + t_{gx}^{r2}. weightofb_1
18. u_i belongs to (t_{gx}^{r2}. behaviorofb_1)
19.   **case** $<t_{gx}^{r2}$. weightofb_2 + ... + t_{gx}^{r2}. weightofb_3
20. u_i belongs to (t_{gx}^{r2}. behaviorofb_3)
21.   **elseif** random(1,n) $<=t_{gx}^{r1}$. leaf$|$+t_{gx}^{r2}. leaf$|$ + t_{gx}^{r3}. leaf $\vee$
22.   **switch** (random(1,j))
23.   **case** $<t_{gx}^{r3}$. weightof(t_{gx}^{r3}. behaviorofb_3)
24. u_i belongs to (t_{gx}^{r3}. behaviorofb_3)
25.   **case** $<t_{gx}^{r3}$. weightofb_3 + t_{gx}^{r3}. weightofb_1
26. u_i belongs to (t_{gx}^{r3}. behaviorofb_1)
27.   **case** $<t_{gx}^{r3}$. weightofb_3 + ... + t_{gx}^{r3}. weightofb_2
28. u_i belongs to (t_{gx}^{r3}. behaviorofb_2)

Algorithm 2 Proposed Algorithm.

Input: a dataset of User Behavior $U=\{u_1,...,u_n\}$, a message M with Interest i $M=\{m_1^i,...,m_j^i\}$, Behavior $B=\{b_1,...,b_3\}$, Group $G=\{g_1,...,g_3\}$

Output: send and receive message behavior

1.   **Initial:**
2.   **Checkout the Behavior**$B=\{b_1,...,b_3\}$, for all User $U=\{u_1,...,u_n\}$
3.
4.   **foreach**$u_i U$**do**   /* Receive Message*/
5.   **while** check all message m_j^i from source node
6.   **if** $u_i b_1$
7.   **if** source in the friend list of u_i
8.   **receive** the message m_j^i to user u_i
9.   **if** $u_i b_2$
10.   **if** $S_c\left(m_j^i,u_1\right)>$ *Threshold Constant*
11.   **receive** the message m_j^i to user u_i
12.   **if** $u_i b_3$
13.   **receive** the message m_j^i to user u_i
14. **foreach**$u_i U$**do**   /* Send Message*/
15.   **while** check all received message m_l^i from source
16.   **if** $u_i b_1$
17.   **if** source in the friend list of u_i
18.   **send** the message m_l^i to source node
19.   **if** $u_i b_2$
20.   **if** $S_c\left(m_l^i,u_1\right)>$*Threshold Constant*
21.   **send** the message m_l^i to source node
22.   **if** $u_i b_3$
23.   **do nothing**

---

4.   SimulationResult

4.1.   Simulation Setting

The simulation conducted by this research has been run in a campus environment, by importing students' online histories on social media platforms and utilizing those as inputs a program written to realize the probabilistic model of students' daily behavior pattern on the platforms. This adopts The ONE Simulator to simulate the number of the actual experiment participants as show in Figure 8:
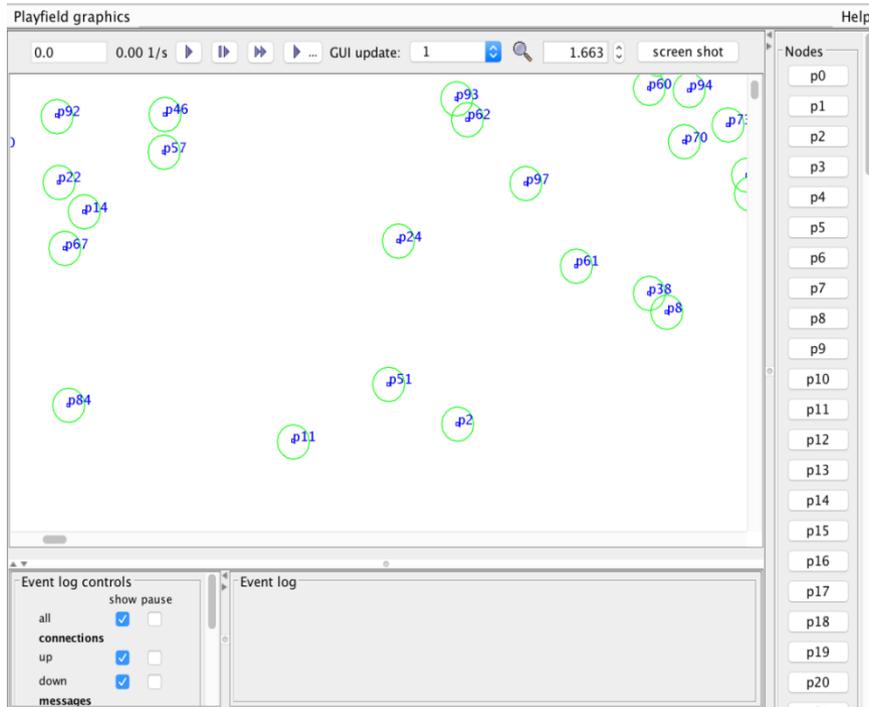
Fig.8. The ONESimulator

Using 104 nodes as the representation. The simulation duration is set to be one day, and only one message to be delivered during the session. The message is to be generated by randomly choose one node from the three groups including Social Group, Interest Group, and Read-Only Group. Nodes and message are attached with properties of interests. Each node has a friend list from collected participant data. The experiment parameters set are listed in Table 5 below:

Table 5. Simulation setting

Simulation parameters	Description
Simulation times	86400 sec
Number of nodes	104
Message Size	500K~1MB
Number of message creation	1
Buffer size	500MB
Time To Live	Unlimited
Virtual Contact	The Users behavior

4.2.   Simulation Results

This research adopts K-means clustering as the method to conduct clustering based on behavior patterns. In the experimental simulation, we aim to incorporate different clustering algorithms to repeat behavior pattern clustering, which allows us to compare and analyze performances achieved by varied clustering algorithms. Furthermore, the result will help us to evaluate whether using a different clustering algorithm impacts the performance of message propagation. The final simulation result will include results obtained from employing Gaussian mixture model and Mean shift clustering algorithms in respective experiments.

The simulation result is presented in Figure 9. We can clearly interpret the message dissemination speeds from the graph that the result suggests the fastest one is Interest Group, the second fastest one is Social Group, and the last one is Read-Only Group. In the result, it is easy to find out no matter which group a student node is part of, the message propagation ratio is steadily increasing. Of the three groups, Interest Group has a better dissemination ratio because messages contain properties of interests. When a message is interesting to most people, it is easily disseminated at a speed even faster than between friends. This kind of situation resembles how the masses utilize social media platforms as we tend to join groups or follow fan pages according to our interests, such as star chasing, group buying, technical news, etc. Thus, when obtaining information about relevant interests, we are more likely to share articles in those groups. If friends of us happen to share similar interests, it is possible for us to share those messages to them.

In Read-Only Group, the students assigned to this group tend to passively accept messages without sharing. This type of behavior is called by the general public as "Lurker." It is relatively easy to infer from the simulation result that the message propagation for the Lurker group was not as active. We can conclude that in order for a message to be disseminated within a short amount of time, it is necessary to forward it first to the students highly favoring the message's interest property, thus speeding up the message dissemination.

Figure 10 is the visualization obtained from behavior pattern clustering by Gaussian mixture model (GMM) algorithm. We can tell from the simulation result that Interest Group still achieves the best dissemination message ratio, close to 80 % of people received the message. Read Only Group remains the most poorly performed one, gaining nearly 40% of people receiving the message. After message propagation lasted for about a day, Social Group and Interest Group eventually share very similar final dissemination ratios, which shows that adopting Gaussian mixture model algorithm as the approach for clustering would make it more difficult to highlight the obvious gap among different message dissemination ratios. However, employing K-means algorithm as the approach for behavior pattern clustering will make it easier to tell Interest Group has a significantly better message dissemination ratio when compared to Social Group and Read Only Group.

Figure 11 shows behavior pattern clustering by Mean shift algorithm. The simulation result indicates that Interest Group has the best message dissemination ratio, close to 70% of people received the message. Social Group has a message dissemination ratio slightly lower than Interest Group, and Read Only Group remains the worst performed one. The simulation result suggests that no matter which clustering algorithm is used, the behavior pattern of Read Only Group is comparatively the most difficult one for

facilitate fast message propagation. As respective simulation results obtained from three different clustering algorithms adopted for behavior clustering, if there is a message characterized by multiple interest properties to be quickly dispersed, K-means clustering algorithm is a better option to distinguish clearly the differences of message dissemination ratios achieved by each behavior pattern clustering, thus identifying the most appropriate group for fast message propagation.
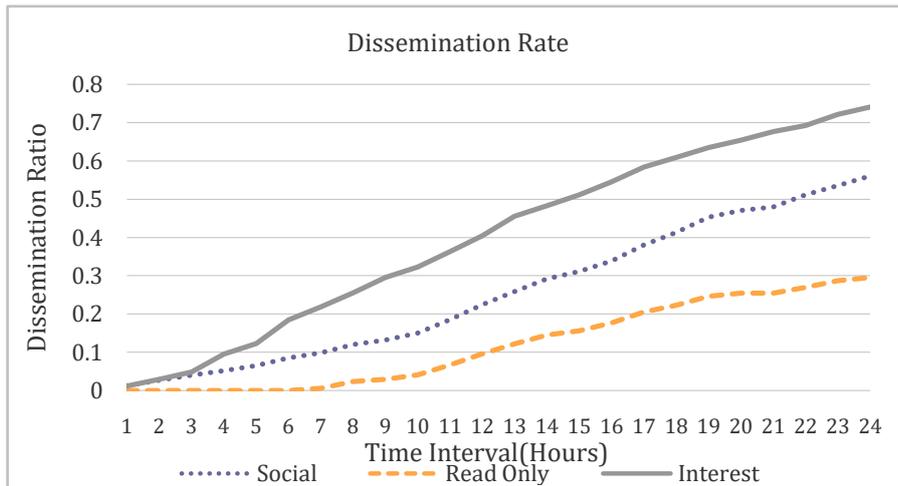
Fig.9. Dissemination Rate (K-means)
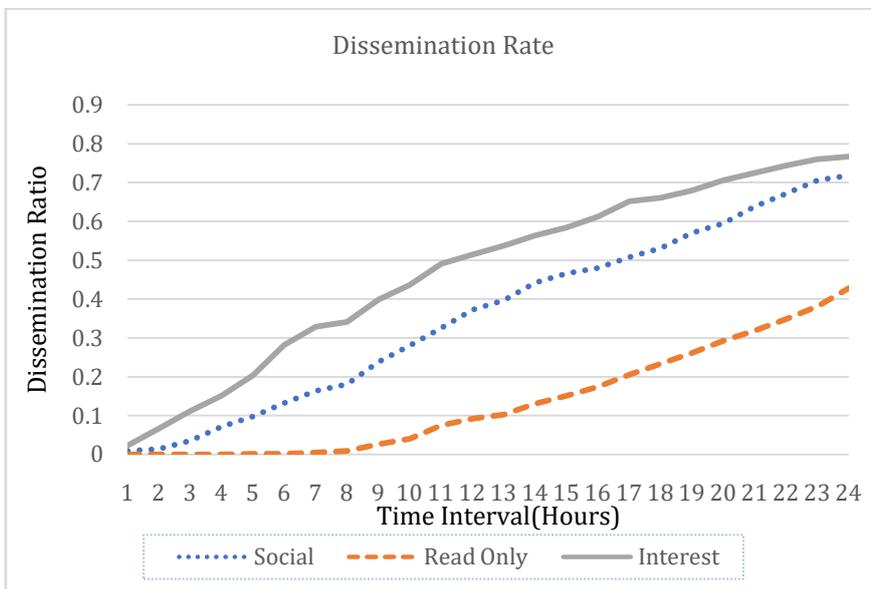
Fig.10. Dissemination Rate (GMM)
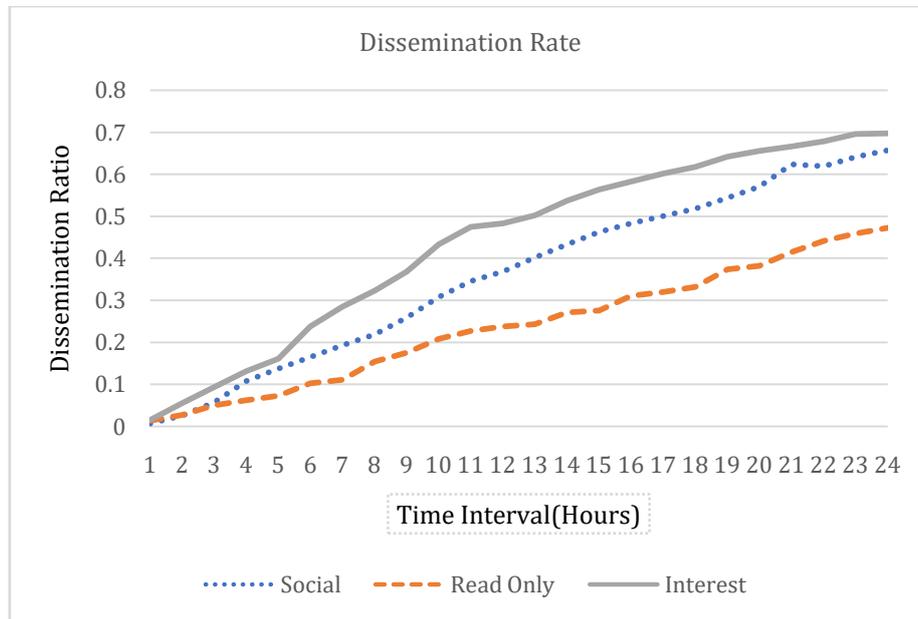
Fig.11. Dissemination Rate (Mean shift)

The simulation results from three different clustering algorithms indicate that people classified as Interest Group tend to have a faster dissemination rate. The major reason is probably due to the fact that multiple interest properties carried by messages will attract more people into dispersing and receiving, and every one's friend list shall have a limit. If a person wanting to disperse messages has few friends, the propagation speed of messages will tend to be slower when compared to messages interested the masses. This simulation result aligns very closely with the real scenario, in which messages are quickly dispersed because either receivers follow commonly interested community groups or fan pages rather than simply shared by friends.

We can compare utilizing different clustering algorithms on behavior patterns to evaluate the simulation of message propagation. From the simulation result, it can be told that people classified as Interest Group will achieve the best dissemination ratio when engaging in message propagation. This indicates the if there is a message carrying multiple interest properties to be quickly dispersed that most of the others obtain information, we can easily target people classified as Interest Group to initiate dispersion, therefore achieve great message dissemination ratio.

The simulation scenario mentioned above requires randomly assigning 5 different interest properties for each dispersed message. Consider another simulation scenario as described by Figure 12, the simulation result is conducted by dispersing messages whose interest properties are characterized as only few of the masses may pay attention to. The simulation result indicates that message dissemination ratio for Social Group is much better compared to Interest Group. We think it may be attributed to that messages' interest properties are not concerned by the masses, thus these messages cannot be easily dispersed for Interest Group. In this simulation scenario, the message dissemination ratio for Interest Group tends to be worse, while the message

dissemination ratio for Social Group maintains approximately similar efficiency like the above-mentioned simulation result. Based on this simulation result, we can tell that if messages carrying obvious interest properties are to be dispersed, finding a member of Interest Group to disperse messages will result in a better message dissemination ratio. On the other hand, if messages are of implicit interest properties, finding a member of Social Group to disperse messages will yield a more stable message dissemination ratio.
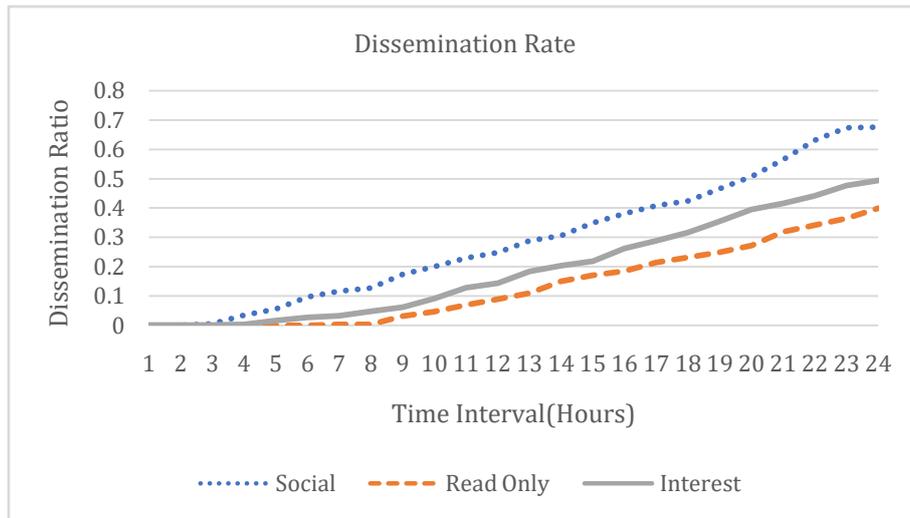
Fig.12. Dissemination Rate

We utilize different clustering algorithms to perform clustering on message dissemination behaviors. It can be told from the simulation result that when dispersing messages with clear interest properties, the message dissemination model obtained from behavior clustering shows that finding Interest Group to perform message propagation will yield better message dissemination ratio. In contrast, when messages' interest properties are not as clear, having Social Group to perform message propagation will result better efficiencies compared to Interest Group. The simulation result shows the method this paper proposed will be able to find an appropriate message dissemination group for messages being dispersed, achieving a satisfactory message dissemination ratio. Further, if the data collected were any different, situations describing message propagation in different environments would be possible. This paper proposed a message propagation model based on virtual contact of personal behavior in DTN. Future work can focus on models developed with other behavior pattern groups in varied environments.

5.   Conclusions and Future work

This research, as the continuation of previous work of NCCU TRACE DATA, proposes an approach to propagate messages via user behaviors in a virtual environment. We believe users' encounters on social media platforms are just like the case of Delay

Tolerant Network. Both of them depend the mechanism of store-carry-and-forward. This research takes users' online histories on social media platforms and their hobbies when using the platforms into consideration. By conducting a simulation resembling users' real-world activities on social media platforms, it can be clearly told from the simulation result that if a message is to be propagated in a campus, relying on a group of people sharing similar interests to disperse them will surely result in more effective coverages.

We also point out two future research directions. The first is the integration of message propagation via physical and virtual contact. In addition to users' geographical location encounters with other users, it is possible to adopt users' message exchanges with other users on social media platform in a way that both physical and virtual contact are employed are the same time to propagate messages in a more realistic way. The second is to upscale the simulation model of message propagation. So far this paper has come up with a message propagation model in a campus. If the dissemination model is to be verified in different environments, only the data of to-be-verified environment are required to be collected as input of the presented simulation in order to find out the dissemination model for users in different environments.

References

1. A. Mtibaa, M. May, M. Ammar, and C. Diot.: PeopleRank: Combining Social and Contact Information for Opportunistic Forwarding. In Proc. IEEE Infocom 2010 Mini Conference, 1-5(2010)
2. Bulut. E, Szymanski, B.K.: Friendship Based Routing in Delay Tolerant Mobile Social Networks, in Global Telecommunications Conference (GLOBECOM) (2010)
3. E. M. Daly and M. Haahr.: Social Network Analysis for Routing in Disconnected Delay-Tolerant MANETs. in Proceedings of the 8th ACM International Symposium on Mobile Ad Hoc Networking and Computing(MobiHoc'07), Montreal, Quebec, Canada, 32-40 (2007)
4. E. M. Daly and M. Haahr.: Social Network Analysis for Information Flow in Disconnected Delay-Tolerant MANETs. IEEE Transactions on Mobile Computing, vol. 8, no. 5:606-621(2009)
5. K. Jahanbakhsh, G.C. Shoja, V. King.: Social-greedy: A Socially-Based Greedy Routing Algorithm for Delay Tolerant Networks. MobiOpp'10: Proceedings of the Second International Workshop on Mobile Opportunistic Networking, ACM, New York, NY, USA, 159–162 (2010)
6. Kopecky M., VojtasP.: Visual E-Commerce Values Filtering Frameworkwith Spatial Database metric. Computer Science and Information Systems (ComSIS) 17(3):983–1006 (2018)
7. Khattak Hasan Ali, Ameer Zoobia, Din UdIkram, Khan Muhammad Khurram.: Cross-layer design and optimization techniques in wireless multimedia sensor networks for smart cities.Computer Science and Information Systems (ComSIS) 16(1):1–17 (2019)
8. Mei, A., Morabito, G., Santi, P., Stefa, J.: Social-Aware Stateless Forwarding in Pocket Switched Networks. In: Proc. IEEE Infocom, Mini Conference (2011)
9. M. C. Chuah. "Social Network Aided Multicast Delivery Scheme for Human Contact-Based Networks." In Proceedings of the 1st Workshop on Simplifying Complex Network for Practitioners (Simplex) (2009)
10. N. Eagle and A. Pentland.: Reality mining: sensing complex social systems. Personal and Ubiquitous Computing, Vol 10(4):255–268 (2006)

11. N. Eagle, A. Pentland, and D. Lazer.: Inferring Social Network Structure using Mobile Phone Data. Proceedings of the National Academy of Sciences (PNAS),106(36), 15274-15278 (2009)
12. P. Hui, J. Crowcroft, and E. Yoneki.: Bubble Rap: Social-Based Forwarding in Delay Tolerant Networks. in Proc. ACM MobiHoc, 241–250 (2008)
13. P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot.: Pocket Switched Networks and the Consequences of Human Mobility in Conference Environments. in WDTN '05: Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking (2005)
14. Spyropoulos, T., Psounis, K., AND Raghavebdra, C. S.: Spray and Wait: An Efficient Routing Scheme for Intermittently Connected mobile networks. In proc. WDTN '05, ACM Press, 252–259(2005)
15. S.C. Nelson, M. Bakht, R. Kravets.: Contact–Based Routing in DTNs. CS, University of Illinois at Urbana-Champaign, USA, In Proc. INFOCOM, April (2009)
16. Tsai, Tzu-Chieh, Chan, Ho-Hsiang.: NCCU Trace: social-network-aware mobility trace. IEEE Communications Magazine, vol. 53, no. 10: 144-149 (2015)
17. Tsai, Tzu-Chieh, Chan, Ho-Hsiang, Han, Chien Chun, Po-Chi Chen.: A Social Behavior based Interest-Message Dissemination Approach in Delay Tolerant Networks.inProc. Int. Conf. Future Netw. Syst. Secur.:62–80 (2016)
18. Ying Zhu, Bin Xu, Xinghua Shi, and Yu Wang.:A Survey of Social-based Routing in Delay Tolerant Networks: Positive and Negative Social Effects. IEEE Communications Surveys & Tutorial, Issue: 9 (2012)
19. Infocom'06.   Retrieved   June   5,   2013,   from http://crawdad.cs.dartmouth.edu/meta.php?name=cambridge/haggle#N10116
20. Zhensheng Zhang.: Routing in Intermittently Connected Mobile Ad Hoc Networks and Delay Tolerant Networks: Overview and Challenges. IEEE Communications Surveys & Tutorials, 8(1):24–37 (2006)

Ho-Hsiang Chan received his Master's degree from Shih Hsin University, Taiwan, in 2010. He is currently working toward his Ph.D. degree in the Department of Computer Science at National Chengchi University. His research interests are delay-tolerant-networks and wireless communication technology. His recent focus is message propagation based on people's behavior.

Tzu-Chieh Tsai received his BS and MS degrees both in Electrical Engineering from National Taiwan University in 1988, and from University of Southern California in 1991, respectively. After that, he got a PhD degree in Computer Science at UCLA in 1996. During 2005/08~2008/07, he was the Chair of Department of Computer Science, at National Chengchi University (NCCU), Taipei, Taiwan. During 2015/09~2018/11, he serves as Vice President of Office of Research & Development at NCCU. Currently, he is a professor in Computer Science Department at NCCU. He has served as the technical committee for several international conferences, and as one of editorial board for "International Journal on Advances in Networks and Services". He is also the CEO of NCCU Pepper &IoT Lab in the NCCU Fintech Center. His recent research works include Ad hoc Networks, P2P Data Management and Computing, Delay Tolerant Networks, 5G MEC Smart Computing, and Digital Performing Arts.

Enhanced Image Preprocessing Method
for an Autonomous Vehicle Agent System

Kaisi Huang[1], Mingyun Wen[2], Jisun Park[3], Yunsick Sung[4], Jong Hyuk Park[5],
and Kyungeun Cho[6, *]

[1]Department of Multimedia Engineering, Dongguk University-Seoul,
30, Pildongro-1-gil, Jung-gu, Seoul 04620, Republic of Korea
hkshks0825@dongguk.edu
[2]Department of Multimedia Engineering, Dongguk University-Seoul,
30, Pildongro-1-gil, Jung-gu, Seoul 04620, Republic of Korea
wmy_ncut@dongguk.edu
[3]Department of Multimedia Engineering, Dongguk University-Seoul,
30, Pildongro-1-gil, Jung-gu, Seoul 04620, Republic of Korea
jisun@dongguk.edu
[4]Department of Multimedia Engineering, Dongguk University-Seoul,
30, Pildongro-1-gil, Jung-gu, Seoul 04620, Republic of Korea
sung@mme.dongguk.edu
[5]Department of Computer Science and Engineering,
Seoul National University of Science and Technology (SeoulTech),
232 Gongneung-ro, Nowon-gu, Seoul, 01811, Korea
jhpark1@seoultech.ac.kr
[6]Department of Multimedia Engineering, Dongguk University-Seoul,
30, Pildongro-1-gil, Jung-gu, Seoul 04620, Republic of Korea
cke@dongguk.edu (corresponding author)

Abstract. Excessive training time is a major issue face when training autonomous vehicle agents with neural networks by using images as input. This paper proposes a deep time-economical Q network (DQN) input image preprocessing method to train an autonomous vehicle agent in a virtual environment. The environmental information is extracted from the virtual environment. A top-view image of the entire environment is then redrawn according to the environmental information. During training of the DQN model, the top-view image is cropped to place the vehicle agent at the center of the cropped image. The current frame top-view image is combined with the images from the previous two training iterations. The DQN model use this combined image as input. The experimental results indicate higher performance and shorter training time for the DQN model trained with the preprocessed images compared with that trained without preprocessing.

Keywords: Image preprocessing, Reinforcement learning, Deep Q learning.

---

* Corresponding author

1.   Introduction

Recently, for predicting traffic and training novice drivers, transportation simulation systems have been utilized. It is necessary for such systems to be designed with as much similarity to the real world as possible. For example, autonomous driving vehicle simulations, which can advance research in automotive safety, require highly realistic transportation simulation systems.

Recently, human-driven agent systems have been issued for automatic systems. Most approaches for learning autonomous driving with neural networks use images as input. Neural networks are excellent image classifier algorithms, that perform well when large amounts of data are provided. Reinforcement learning [1] and supervised learning [2] utilize image input for model training.

Deshpande proposed a deep reinforcement learning approach to train autonomous driving vehicles [1]. This approach utilizes environment images, agent speeds, distances to road centers and angles of the heading vectors with respect to direction of the road as the network inputs. To reduce the redundant features that don't affect decision-making of network and make the training converge faster, image preprocessing is necessary.

To reduce the training time and achieve a higher performance training result, this paper proposes a method to preprocess input images for training an autonomous driving vehicle agent with a deep Q network (DQN). In the proposed method, all the information about the roads and junctions that are related to driving a vehicle agent in a virtual environment are first extracted and recorded in an environment information file. A whole top-view image of the virtual environment is redrawn and the vehicle agent is added to this image. This top-view image is then cropped according to the vehicle agent's surrounding environment and utilized to create an input image sequence for training the DQN. The outputs of the DQN are steering and acceleration, which are utilized for the control of the vehicle agent in the virtual environment. The virtual environment returns a reward for training the DQN model in each iteration.

In the following, Section 2 relevant previous studies are discussed. Section 3 presents the main concepts of the image preprocessing for the DQN. Section 4 describes the detailed implementation of the proposed method and the experimental results. Finally, Section 5 concludes the study and discusses future works.

2.   Related work

2.1.   Imitation learning

Behavioral cloning (BC) is the most common approach to imitation learning. The objective of BC is to learn the relationship between states and optimal behaviors as a supervised learning problem [4–6]. A convolutional neural network (CNN) is frequently used for BC. A CNN learns features automatically from demonstrations given a suitable dataset. Bojarski [2] introduced a system of self-driving cars using a CNN that automatically learns the internal features of input images that it never explicitly trains itself to detect. However, it requires a large dataset for training to guarantee that most states that may occur are covered.

Because of environmental restrictions in real environments, end-to-end control cannot be trained to handle all situations. A virtual simulation environment [17] has been used for training to overcome this problem. In this approach, a virtual environment similar to the real environment is constructed for the virtual simulation. Captured images and the control signals of a vehicle agent are collected. A CNN is trained based on the collected images and control signals, and then used to control a vehicle in the real environment.

2.2.   Reinforcement learning

Reinforcement learning algorithms allow agents to learn how to behave differently in different situations. Its aim is to establish a policy that considers the situation and select actions that maximize a reward[7]. Reinforcement learning has been successfully applied to many different tasks such as playing relatively simple Atari games, mastering the relatively more complex game of Go, and controlling robots in real environments.

Inverse reinforcement learning is the most successful imitation learning approach [8,9]. This approach assumes that the behaviors that learners desire to imitate are generated by experts. This approach attempts to estimate a reward function to explain the behaviors of experts [14]. However, obtaining the reward function of inverse reinforcement learning is slow in terms of convergence speed.

A reinforcement learning based driving policy for autonomous road vehicles was proposed in [17]. The DQN is trained to control a vehicle by changing its heading, acceleration, and deceleration. The input of this DQN is a vector that is constructed from vehicle sensors and includes the longitudinal velocity and vehicle position in the lane. However, a virtual environment can occasionally be too complex to extract features for a network.

3.   DQN-input image preprocessing approach

In this section, the proposed preprocessing method is described. Section 3.1 gives an overview of the proposed framework. Sections 3.2 and 3.3, describe how to extract the environmental information from a virtual environment and how to redraw the top-view image of the whole environment, respectively. Section 3.4 provides details on how the top-view image is cropped around a vehicle agent in the whole top-view image. Finally, the DQN model structure is described in Section 3.5.

3.1.   Overview

The main process of the proposed method is shown in Fig. 1. During the environment information extraction phase, all information about roads and junctions related to a vehicle agent in a virtual environment are collected and written to an environment information file. Subsequently, during the environment redrawing phase, the information in this file is utilized to redraw the whole top-view image of the virtual environment using straight lines and arcs. The location and direction of the vehicle

agent in the virtual environment is added to the whole top-view image. During the top-view image cropping phase, the top-view image is cropped to place the vehicle agent in the center of the cropped image. In each training iteration, the cropped top-view image is utilized as an input to the DQN. The outputs of the DQN are the steering and acceleration of the vehicle agent. These two values are used to control the vehicle agent in the virtual environment, and the virtual environment returns a reward according to the status of the vehicle agent. The DQN model undergoes training in each iteration.
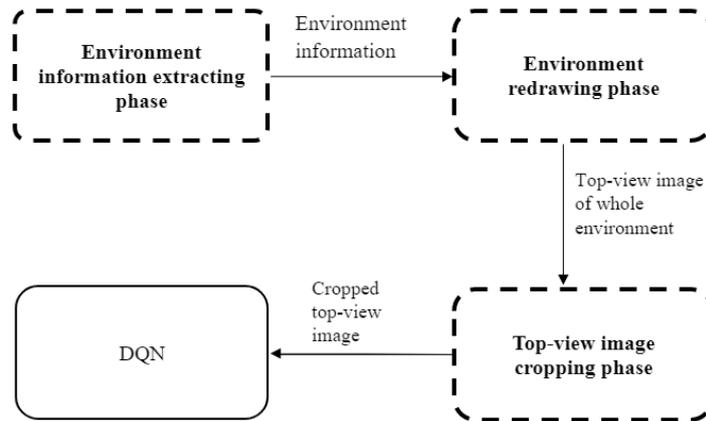
Fig. 1. Processes of the proposed method

3.2. Environmental information extraction

In a virtual environment, diverse types of information exist. The information on the roads and junctions are important for driving a vehicle agent. Roads and junctions comprise the map that cars should drive. One road can be connected to other roads or junctions, and one junction can be connected to multiple roads. All link information exists in the virtual environment. Therefore, these two types of information should be extracted.
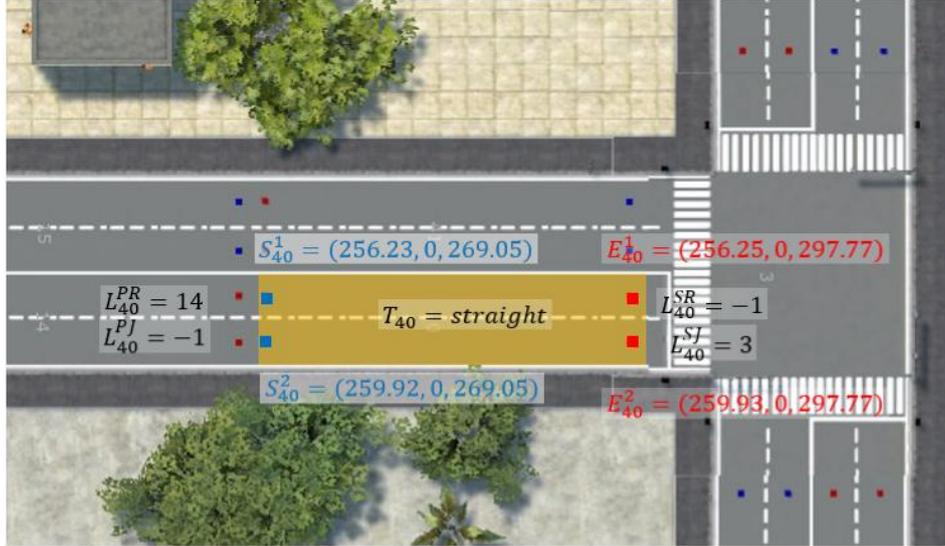
Fig. 2. Example of a 3D road

In this paper, the i-th three-dimensional (3D) road R_i is described using four elements $[i, L_i, T_i, N_i]$, defined as follows: the road ID i, link information L_i, road type T_i, and lanes N_i. The link information L_i is expressed as $[L_i^{SR}, L_i^{PR}, L_i^{SJ}, L_i^{PJ}]$, which consists of the successor road ID L_i^{SR}, predecessor road ID L_i^{PR}, successor junction ID L_i^{SJ} and predecessor junction ID L_i^{PJ}. The road type is either straight, corner or merging. The lanes N_i consists of a set of lanes, $\{N_i^1, N_i^2, ..., N_i^n, ...\}$, where n describes the order of the lanes. A lane N_i^n is denoted by $[S_i^n, E_i^n]$. Therefore, each lane contains two elements: start point $S_i^n = (x_i^n, y_i^n, z_i^n)$ and end point $E_i^n = (x_i^n, y_i^n, z_i^n)$.

Fig. 2 shows an example of a 3D road. Road R_{40}, which is indicated by the yellow area is described by $R_{40} = [40, L_{40}, straight, N_{40}]$, where $L_{40} = [-1, 14, 3, -1]$, $N_{40} = [N_{40}^1, N_{40}^2]$, $N_{40}^1 = [(256.23, 0, 269.05), (256.25, 0, 297.77)]$, and $N_{40}^2 = [(259.92, 0, 269.05), (256.25, 0, 297.77)]$.
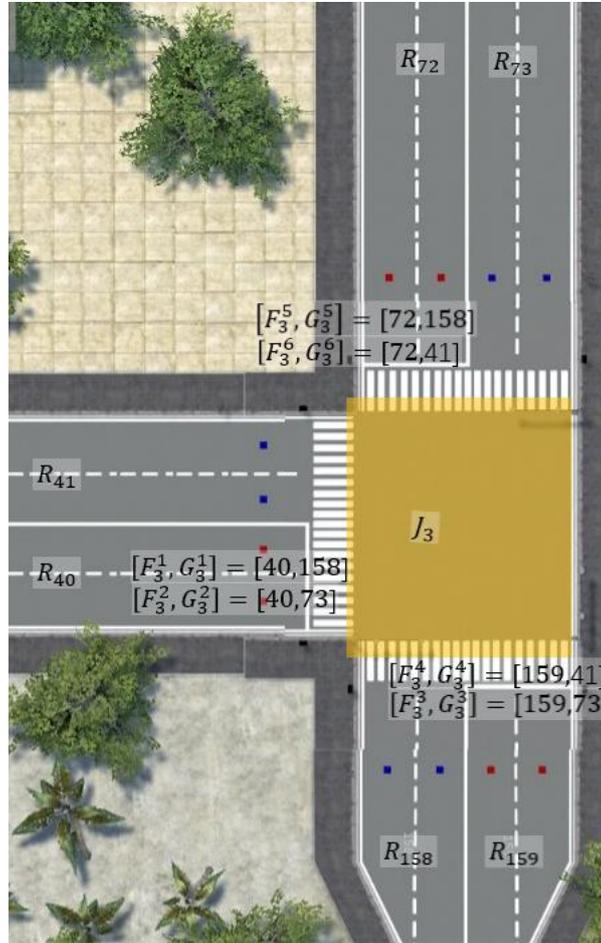
Fig. 3. Example of a 3D junction

The i-th 3D junction J_i is expressed as [i, C_i], where each 3D junction contains two elements: junction ID i, and connections C_i. Connections C_i is a set of connections {C_i^1, C_i^2, ..., C_i^c, ...}, where c denotes the order of connections, and lane C_i^c is [F_i^c, G_i^c]. Therefore, each connection contains two elements: the from-road ID F_i^c and the to-road ID G_i^c.

Fig. 3. shows an example of a 3D junction. Junction J_3, which is indicated by the yellow area, is described by J_3=[3, C_3] where C_3=[C_3^1, C_3^2, C_3^3, C_3^4, C_3^5, C_3^6], C_3^1=[$40,168$], C_3^2=[$40,70$], C_3^3=[$159,73$], C_3^4=[$159,41$], C_3^5=[$72, 158$], and C_3^6=[$72, 41$]

3.3.   Redrawing of the top-view image of the whole environment

After extracting all the information regarding the roads and junctions in a virtual environment, a top-view image of the whole environment is redrawn. First, as shown in Fig. 4, image width W and image height H are determined. According to the size of the whole environment, one pixel corresponds to a unit length α in the environment. There are three axes X, Y, and Z in the virtual environment. To redraw a top-view image, the Y axis is ignored, the image width W is the size of the environment along the Z axis, and the image height H is the size along the X axis. By considering all the start points and end points of all road information, the minima and maxima, $x_{min}=Min(x_i^n)$, $x_{max}=Max(x_i^n)$, $z_{min}=Min(z_i^n)$, and $z_{max}=Max(z_i^n)$ are extracted. To prevent cropping errors, the corresponding image edge is extended by a number of pixels β. Therefore, image width W $\frac{(z_{max}-z_{min})}{\alpha}+2\beta$ [OBJ] and image height H $\frac{(x_{max}-x_{min})}{\alpha}+2\beta$ [OBJ]. To redraw the image, start points $S_i^n=(x_i^n,\ y_i^n,\ z_i^n)$ and end points $E_i^n=(x_i^n,\ y_i^n,\ z_i^n)$ $w=((z_i^n-z_{min})\times\frac{1}{\alpha}+\beta)$ [OBJ] $h=((x_i^n-x_{min})\times\frac{1}{\alpha}+\beta)$ [OBJ].
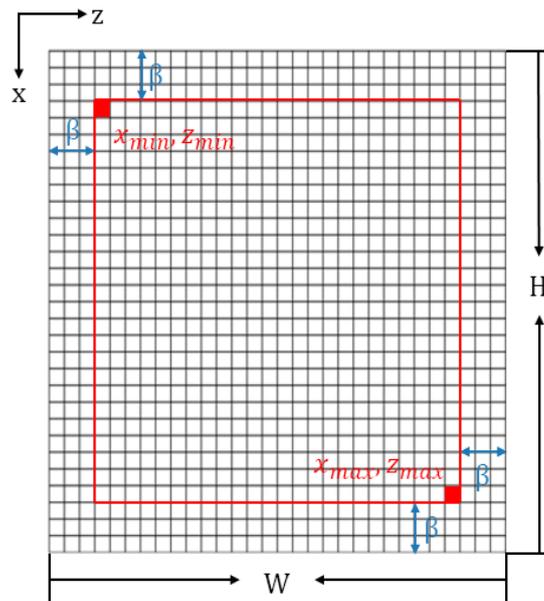
Fig. 4. Image size determination

After converting all locations of the start points and end points of all lanes to pixel coordinates, the top-view image of whole environment is redrawn. First, as shown in Fig. 5, a straight road is represented by a straight line between the start point S_i^n and the end point E_i^n of each lane (red and blue points are only for explanation). A corner road is then represented by drawing an arc between the start point S_i^n and end point E_i^n of each lane.
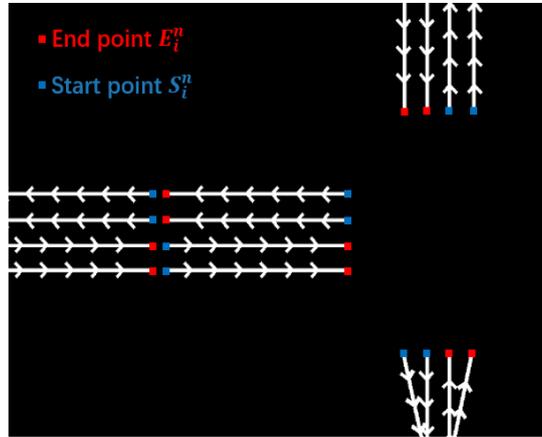
Fig. 5. Drawing road connections

Subsequently, as shown in Fig. 6,adjacent roads are connected by drawing straight lines between their end points E_i^n and start points $S_{L_i^{SR}}^n$.
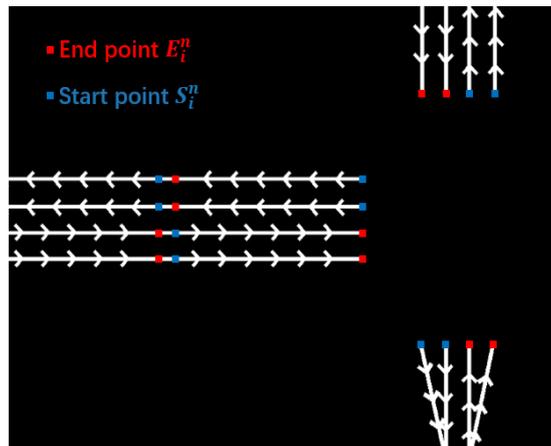
Fig. 6. Drawing connections between the roads and successor roads

Next, as shown in Fig. 7, the straight connection in junction$J_i$is connected by straight lines. The turning connections in the junction are drawn using arcs to connect end points $E_{F_i^c}^n$ and start points $S_{G_i^c}^n$.
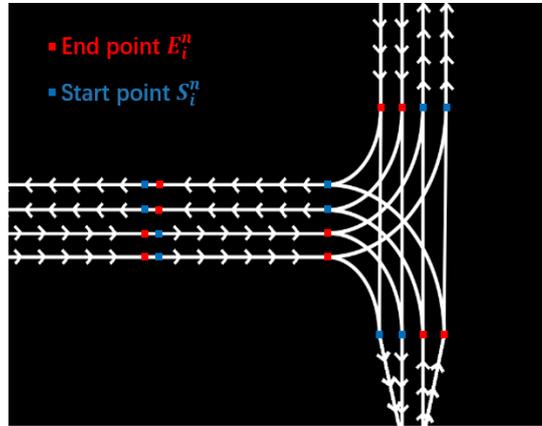
Fig. 7. Drawing junction connections
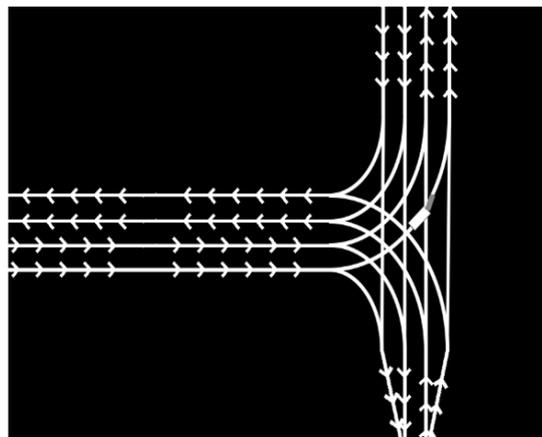
Fig. 8. Adding thevehicle agent

After the top-view image of the whole environment has been drawn, vehicle agent$v$ also needs to be added to the image as shown in Fig. 8. The vehicle agent coordinates on the x-axis x_v, and the vehicle agent coordinates on the z-axisz_v, as well as directionD_v are collected for each frame. Therefore, thevehicle agent's pixel coordinate $w_v$and h_v are

$(z_v - z_{min}) \times \dfrac{1}{\alpha} + \beta$ and $(x_v - x_{min}) \times \dfrac{1}{\alpha} + \beta$, respectively. Finally, as shown in Fig. 8, the

vehicle agent is located at pixel (w_v, h_v) and rotated to the direction$D_v$in the top-view image of the whole environment.

3.4.   Top-view image cropping

To train a DQN model, a vehicle agent only needs to consider the surrounding environments. Therefore, the top-view image of whole environment needs to be cropped to a small area centered on the location of the vehicle agent as shown in Fig. 9. To further reduce the number of features in the top-view image, the vehicle agent should always face in one direction.
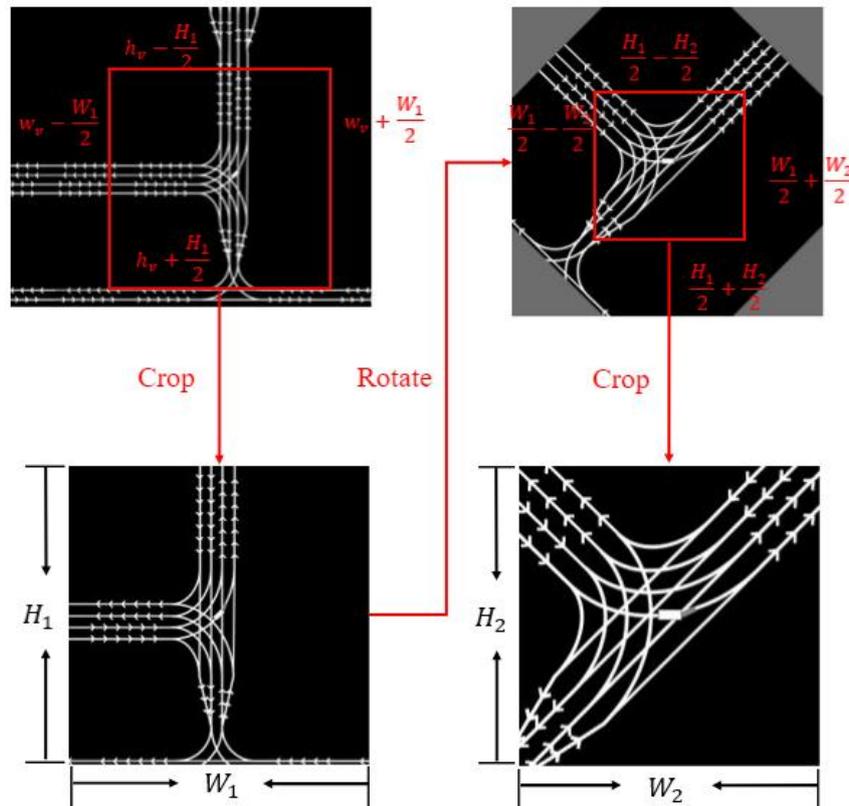
Fig. 9. Top-view image cropping process

To prevent rotation error, a two-step cropping approach is proposed. In the first crop, the width W_1 and height H_1 of the first crop are larger than the width W_2 and height H_2 of the second crop.

The signed angle θ between direction D_v of vehicle agent v and a fixed direction D_0 is utilized to rotate the top-view image after the first crop. When rotating, the resolution of the image should remain the same. The final cropped top-view image is utilized as the DQN input for each frame.

A single frame of a cropped top-view cannot contain information about the vehicle agent's speed. The cropped top-view image is converted to grayscale and stacked with the previous two frame images which have also been converted to grayscale. Fig. 10

shows a three frames top-view image. The time interval between the adjacent frames should be kept fixed.
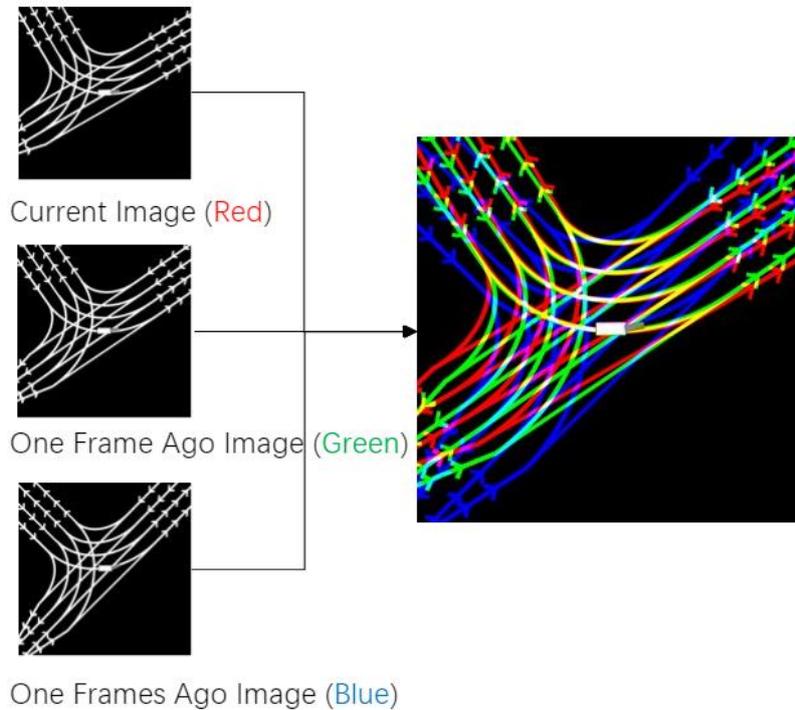
Fig. 10. Top-view image of three stacked adjacent frames

3.5.   DQN architecture with DQN-input images

The structure of a DQN is shown in Fig.11. The DQN-input image is a 240×240 top-view image and the outputs are steering and acceleration values. Steering values range from -1 to 1, in steps of 0.1. Acceleration values range from −1 to 1 in steps of 0.2. This network utilizes categorical cross entropy as the loss function. The details of the parameters of each layer are listed in Table 1.
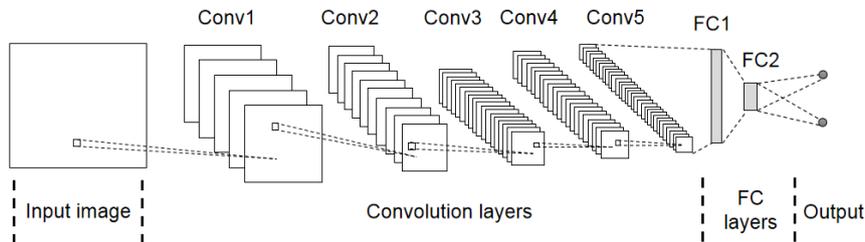
Fig. 11. DQN architecture

Table 1. Parameters of the layers of a DQN

Convolution Layer	Filters	Kernel Size	Stride
Conv1	24	5×5	2×2
Conv2	36	5×5	2×2
Conv3	48	5×5	2×2
Conv4	64	3×3	1×1
Conv5	64	2×2	1×1
Fully-connected Layer	Neurons		
FC1	36,864		
FC2	128		

The function used to calculate the DQN training reward is

$$R(V_t,\theta_t,D_t) = V_t \times (\cos(\theta_t) - \sin(\theta_t)) \quad \frac{|D_t - D_0|}{D_0} \quad (1)$$

where V_t is the vehicle agent's speed along the road direction, θ_t is the angle between the heading of the vehicle agent and the direction of the road, D_t is the distance from the road's left edge to the vehicle agent's center, and D_0 is the half of the road width.

As the number of training iterations increases, exploratory rate ε_t is decreased by

$$\varepsilon_t = Max(\varepsilon_t) \quad \frac{t \times (Max(\varepsilon_t) - Min(\varepsilon_t))}{T} \quad (2)$$

where $Max(\varepsilon_t)$ is the maximum of ε_t and $Min(\varepsilon_t)$ is the minimum of ε_t. In addition, T is the total number of frames that the vehicle agent has trained and t is current frame.

4.   Experiments

In this section, the experimental results are described. Section 4.1 introduces the experimental goal and environment. Section 4.2 presents the results of the top-view image of the whole environment. Section 4.3 presents the samples of a cropped top-view image, and Section 4.4 presents the vehicle agent obtained using a DQN.

4.1.   Experimental goal and environment

The experimental goal is to train a DQN model to control a vehicle agent in a virtual environment. To evaluate the proposed method, the results obtained using the preprocessed three-frame top-view image are compared with the results obtained using captured top-view image generated by the simulator without any preprocessing and a preprocessed single-frame top-view image. The rewards of each DQN trained with one of the three types of DQN inputs were calculated, and the number of successfully reached destinations of a vehicle agent controlled by each trained DQN was analyzed.

To train the DQN model, 300,000 frames were used. The maximum exploratory rate was 0.3 and the minimum exploratory rate was 0.01. The replay memory stored the most recent 5000 frames of data. For every frame, the batch size for the DQN was 32.

The performance of the proposed method was verified on a desktop computer with an i7-7740X 4.30GHz CPU, an NVIDIA GeForce GTX 1080 graphics card, and a 16.0GB DDR4 RAM. The virtual environment simulator shown in Fig.12 runs on Unity3D.

Fig. 12. Virtual environment simulator used for the experiment

4.2.   Results of the top-view image redrawing

In the experiment environment, there were eight junctions and a total of 162 roads: 140 straight, 7 corner, and 15 merging roads. The captured top-view image is shown in Fig. 13, and Fig. 14 shows the final redrawn top-view image of the whole environment.

Fig. 13. Capture top-view image of the whole environment
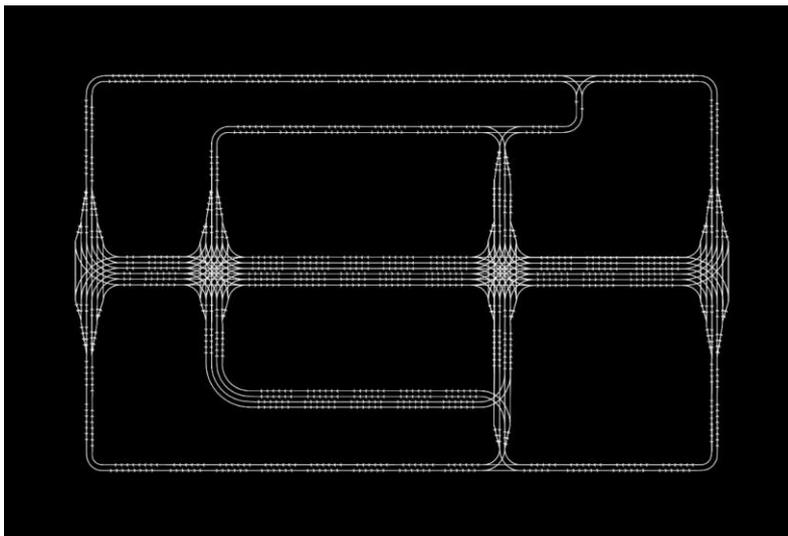
Fig. 14. Redrawn top-view image of the whole environment

In the redrawn image, the white lines represent road lanes and the arrows on the lines indicate the direction of the lanes. The areas with white crossing lines are the junctions which contain many road connections. In contrast to the captured top-view image in Fig. 13, information that is clearly not related to training a vehicle agent DQN model has been removed from the redrawn top-view image of the whole environment.

4.3.   Results of top-view image cropping

Captured top-view image samples, single-frame top-view image samples, and three-frame top-view samples are shown in Figs. 15, 16, and 17, respectively. The left panels in these figures show the vehicle agent driving on a straight road. The middle panels show the vehicle agent turning right at a corner, and the right panels show the vehicle agent turning right at a junction.
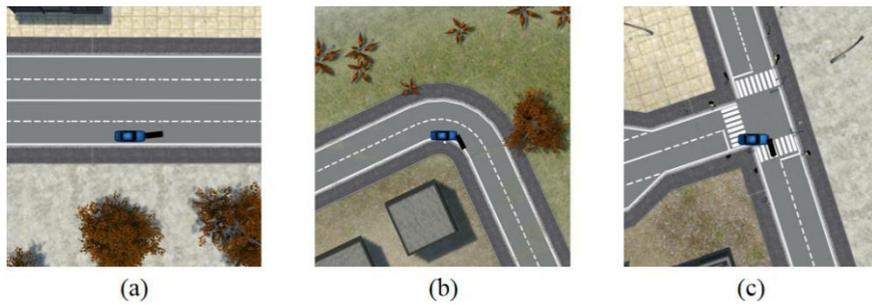
Fig. 15. Captured top-view image samples
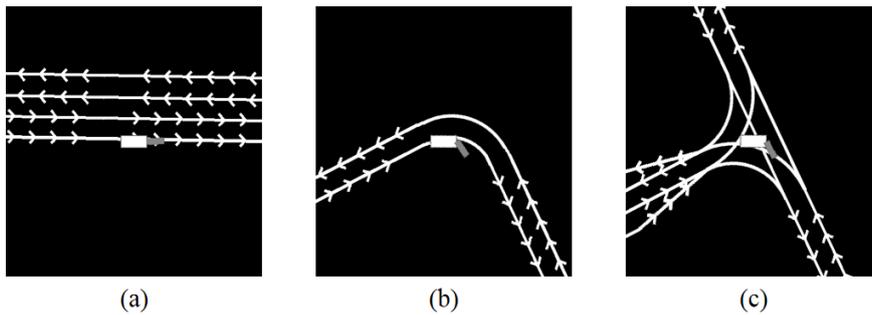
Fig. 16. Single-frame top-view image samples
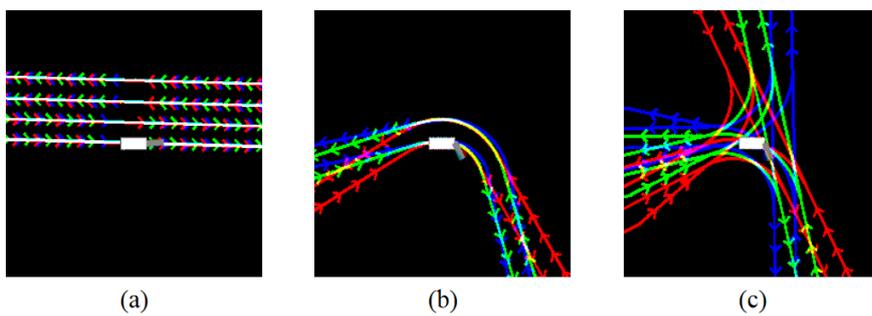
Fig. 17. Three-frame top-view image samples

4.4.   Result of training using a DQN

Three DQN models were trained with 300,000 frames of captured images, single-frame images, or three-frame images. The mean rewards of the three DQN models after 10,000 frames were processed are shown in Table 2. The mean reward of the DQN model trained with the proposed three-frame images was 221.3% higher than that of the DQN model trained with the captured top-view images.

Table 2. Comparison of the mean DQN model rewards after 10,000 frames

Input image type	Mean reward
Captured	1.6244
Single frame	4.1360
Three frames	5.2193

To evaluate the performance of the DQN models trained with the three types of input images, experiments were carried out on the same map. Each DQN model was used to control a vehicle agent to a destinations 100 times, and the number of successfully reached destination is listed in Table 3. These results show that the DQN model trained with the proposed three-frame top-view images has a success rate that is 712.5% higher than that of the DQN model trained with the captured top-view images.

Table 3. Comparison of the number of successfully reach destinations

Input image type	Example
Raw	8
Single frame	49
Three frames	65

5.   Conclusions

This paper proposed an image preprocessing method for DQN image inputs to train an autonomous vehicle agent in a virtual environment. In this method, the information of the roads and junctions related to a vehicle agent in a virtual environment is extracted and recorded in a file. Using this information, the size of the top-view images is determined. In the top-view image, roads and junctions are drawn as straight lines and arcs. The virtual environment obtained vehicle agent locations and directions for every frame. The top-view image is then cropped so that the vehicle agent is in the center of the cropped image. The outputs of the DQN are steering and acceleration values, which can be utilized to control the vehicle agent in the virtual environment.

In the experiment, after the DQN was trained for 300,000 frames, the mean training reward of the DQN model with the proposed top-view images reached 5.2193. The number of successful navigations using the DQN model trained with the proposed top-view images was 712.5% higher than tthat obtained by the DQN model trained with the captured top-view images. These results show that the proposed DQN input image preprocessing method can substantially improve the performance of the DQN model.

In future work, we will explore more applications of the proposed image preprocessing method. Further, we will test the model using other neural networks that use image as input.

References

1. Niranjan Deshpande, Anne Spalanzani: Deep Reinforcement Learning based Vehicle Navigation amongst pedestrians using a Grid-based state representation. IEEE Intelligent Transportation Systems Conference, 2081-2086. Auckland, New Zealand. (2019)
2. Mariusz Bojarski, Davide Del Testa, Daniel Dworakowski, Bernhard Firner, Beat Flepp, Prasoon Goyal, Lawrence D. Jackel, Mathew Monfort, Urs Muller, Jiakai Zhang, Xin Zhang, Jake Zhao and Karol Zieba: End to End Learning for Self-driving Car. (2016). [Online]. Available: https://arxiv.org/pdf/1604.07316
3. Ahmed Hussein, EyadElyan, Mohamed Medhat Gaber and Chrisina Jayne: Deep Imitation Learning for 3D Navigation Tasks. Neural Computing and Applications, Vol. 29, 389-404. (2017)
4. Dean A. Pomerleau: Alvinn: An Autonomous Land Vehicle in A Neural Network. Advances in Neural Information Processing Systems, 305-313. Denver, United States of America. (1989)
5. Stéphane Ross, Geoffrey J. Gordon and J. Andrew Bagnell: A Reduction of Imitation Learning and Structured Prediction to No-regret Online Learning. International Conference on Artificial Intelligence and Statistics, 627-635. Fort Lauderdale, United States of America. (2011)
6. Dean A. Pomerleau: Efficient Training of Artificial Neural Networks for Autonomous Navigation. Neural Computation, Vol.3, 88-97. (1991)
7. 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, 647–664. (2020)
8. Andrew Y. Ng and Stuart Russell: Algorithms for Inverse Reinforcement Learning. International Conference on Machine Learning, Vol. 1, 2. Palo Alto, United States of America. (2000)
9. Brian D. Ziebart, Andrew Maas, J.AndrewBagnell and Anind K. Dey: Maximum Entropy Inverse Reinforcement Learning. Association for the Advance of Artificial Intelligence, Vol. 8, 1433-1438. Chicago, United States of America. (2008)
10. Volodymyr Mnih, KorayKavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, IoannisAntonoglou, Helen King, Dharshan Kumaran, DaanWierstra, Shane Legg and Demis Hassabis: Human-level Control through Deep Reinforcement Learning. Nature, Vol. 518, 529-533. (2015)
11. Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver and DaanWierstra: Continuous Control with Deep Reinforcement Learning. (2015). [Online]. Available: https://arxiv.org/abs/1509.02971
12. Burr Settles: Active Learning Literature Survey. Computer Sciences Technical Report 1648. (2009)
13. Umar Syed and Robert E. Schapire: A Reduction from Apprenticeship Learning to Classification. Advances in Neural Information Processing Systems, 2253-2261, Vancouver, Canada. (2010)

14. Jonathan Ho, Jayesh K. Gupta and Stefano Ermon: Model-free Imitation Learning with Policy Optimization. International Conference on Machine Learning, 2760-2769. New York, United States of America. (2016)
15. Zongwei Zhou, Jae Shin, Lei Zhang, SuryakanthGurudu, Michael Gotway, and Jianming Liang: Fine-tuning Convolutional Neural Networks for Biomedical Image Analysis: Actively and Incrementally. Conference on Computer Vision and Pattern Recognition, 7340-7351. Honolulu, United States of America. (2017)
16. Makantasis, Konstantinos, Maria Kontorinaki, and IoannisNikolos: A Deep Reinforcement Learning Driving Policy for Autonomous Road Vehicles. (2019). [Online]. Available: https://arxiv.org/abs/1905.09046
17. Chenyi Chen, Ari Seff, Alain Kornhauser, Jianxiong Xiao: Deepdriving: Learning affordance for direct perception in autonomous driving. IEEE International Conference on Computer Vision, 2722-2730. Santiago, Chile. (2015)
18. Bloem, Michael, and Nicholas Bambos: Infinite time horizon maximum causal entropy inverse reinforcement learning. IEEE Conference on Decision and Control, 4911-4916. Santiago, Chile. (2014)
19. Fan, H., Xia, Z., Liu, C., Chen, Y., Kong, Q: An Auto-tuning Framework for Autonomous Vehicles. (2018). [Online]. Available: https://arxiv.org/abs/1808.04913
20. Sasaki, Fumihiro, Tetsuya Yohira, and Atsuo Kawaguchi.: Sample efficient imitation learning for continuous control. International Conference on Learning Representations.(2018)
21. Kim, B., Farahmand, A. M., Pineau, J., Precup, D: Learning from limited demonstrations. In Advances in Neural Information Processing Systems, 2859-2867. (2013)
22. Ho, Jonathan, and Stefano Ermon.: Generative adversarial imitation learning. Advances in neural information processing systems, 4565-4573. (2016)

Kaisi Huang received his B. Eng. Degree in Digital Media Technology in 2017 from North China University of Technology, Beijing, The People's Republic of China. Since September 2017, he is in the M. Eng. course at the Department of Multimedia Engineering, Dongguk University, Seoul, Republic of Korea. He has done some works mainly associated with NUI (Natural User Interface) applications, VR (Virtual Reality) applications and artificial intelligence with deep learning.

Mingyun Wen received the B. Eng. Degree in Digital Media and Art in 2016 from North China University of Technology, Beijing, China and the M.Eng. in Multimedia Engineering in 2018 from Dongguk University, respectively. She is a Ph.D. candidate in the department of Multimedia Engineering in Dongguk University now. Her main research interests include 3D simulation of large virtual environment, artificial intelligence and deep learning.

Jisun Park received her B. Eng. Degree in Multimedia in 2016 from Dongguk universitiy computer science institute, Seoul, Republic of Korea and M. Eng. Degrees in Multimedia Engineering in 2018 from Dongguk University, Seoul, Republic of Korea. Since September 2018, she is in the Dr. Eng. course at the Department of Multimedia Engineering, Dongguk University, Seoul, Republic of Korea. She has done many works mainly associated with 2D/3D artificial intelligence with deep learning.

Yunsick Sung, currently an associate professor in the Department of Multimedia Engineering at Dongguk University-Seoul, Seoul, Republic of Korea. He received his BS degree in the Division of Electrical and Computer Engineering from Pusan National

University, Busan, Republic of Korea, in 2004, the M. S. degree in Computer Engineering from Dongguk University-Seoul, Seoul, Republic of Korea, in 2006, and the Ph. D. degree in Game Engineering from Dongguk University, Seoul, Republic of Korea, in 2012. He was employed as a member of the Researcher at Samsung Electronics in the Republic of Korea between 2006 and 2009. He was a postdoctoral fellow at the University of Florida, Florida, USA, between 2012 and 2013. His research interests are focused on the areas of superintelligence in the fields of immersive softwares, UAVs, security, and music using demonstration-based learning and deep learning.

Jong Hyuk Park received Ph.D. degrees in Graduate School of Information Security from Korea University, Korea and Graduate School of Human Sciences from Waseda University, Japan. From December, 2002 to July, 2007, Dr. Park had been a research scientist of R&D Institute, Hanwha S&C Co., Ltd., Korea. From September, 2007 to August, 2009, He had been a professor at the Department of Computer Science and Engineering, Kyungnam University, Korea. He is now a professor at the Department of Computer Science and Engineering and Department of Interdisciplinary Bio IT Materials, Seoul National University of Science and Technology (SeoulTech), Korea. His research interests include IoT, Human-centric Ubiquitous Computing, Information Security, Digital Forensics, Vehicular Cloud Computing, Multimedia Computing, etc.

Kyungeun Cho is a full professor at the Department of Multimedia Engineering at the Dongguk University inSeoul, Republic of Korea since Sept. 2003. She received her B. Eng. Degree in Computer Science in 1993, herM. Eng. and her Dr. Eng. Degrees in Computer Engineering in 1995 and 2001, respectively, all from DonggukUniversity, Seoul, Korea. During 1997–1998 she was a research assistant at the Institute for Social Medicine atthe Regensburg University, Germany, and a visiting researcher at the FORWISS Institute at TU-MuenchenUniversity, Germany. Her current research interests are focused on the areas of intelligence of robot and virtualcharacters and real-time computer graphics technologies. She has led a number of projects on robotics and gameengines and also has published many technical papers in these areas.

A Study of Universal Zero-Knowledge Proof Circuit-based Virtual Machines that Validate General Operations & Reduce Transaction Validation

Soonhyeong Jeong[1] and Byeongtae Ahn[2, *]

[1] Onther Inc., 527, Gangnam-daero, Seocho-gu, Seoul, Republic of Korea
kevin.j@onther.io
[2] Faculty of Liberal & Arts College, Anyang University, 22, 37-Beongil, Samdeok-Ro,
Manan-Gu, Anyang 430-714, Republic of Korea
ahnbt@ anyang.ac.kr

Abstract. Recently, blockchain technology accumulates and stores all transactions. Therefore, in order to verify the contents of all transactions, the data itself is compressed, but the scalability is limited. In addition, since a separate verification algorithm is used for each type of transaction, the verification burden increases as the size of the transaction increases. Existing blockchain cannot participate in the network because it does not become a block sink by using a server with a low specification. Due to this problem, as the time passes, the data size of the blockchain network becomes larger and it becomes impossible to participate in the network except for users with abundant resources. Therefore, in this paper, we studied the zero knowledge proof algorithm for general operation verification. In this system, the design of zero-knowledge circuit generator capable of general operation verification and optimization of verifier and prover were also conducted. Also, we developed an algorithm for optimizing key generation. Based on all of these, the zero-knowledge proof algorithm was applied to and tested on the virtual machine so that it can be used universally on all blockchains.

Keywords: Zero-Knowledge, validation, transaction, BlockChain, Ethereum

1. Introduction

The blockchain-based distributed application market is expected to grow from about $ 3.2 billion in 2019 to more than $ 60 billion in 2024. Among them, the market with 'transaction processing' as a profit model is expected to reach 55% of the total. This means that blockchain-based distributed applications are generally provided on the basis of open source, so transaction fees rather than content usage fees are inevitably accepted by users. Therefore, the economic value of the technology to efficiently process transactions is very positive. In the past decade, numerous blockchain implementations have emerged as platforms, but there has been no significant innovation in terms of accumulating and storing transactions, but rather the burden of verifying the chain data

---

*Corresponding author

has increased as it supports complex operations. Therefore, it is necessary to lead the structural innovation of the blockchain by creating a verification module that can be commonly used in various blockchain platforms that will appear in the future.

Currently, the domestic blockchain technology is mainly biased toward mainnet-based technologies such as distributed ledgers and consensus algorithms, but the area where the domestic technology ecosystem is likely to lead in the global market is the distributed application area rather than the mainnet. And there is currently no virtual machine based on zero-knowledge proof that can efficiently verify complex operations required for distributed applications, not just bookkeeping. Blockchain is a decentralized digital ledger that secures the integrity of transaction details and allows participants to share details without the involvement of a trusted third party in a peer-to-peer (P2P) network. A typical example of applying blockchain is cryptocurrency such as Bitcoin and Ethereum. Ethereum introduced the Ethereum virtual machine (hereafter EVM). With EVM, users can program their own way, rather than performing a predefined set of tasks. However, EVM is very inefficient compared to existing virtual machines such as Java Virtual Machine (JVM). And it is difficult to support a complex application environment. Therefore, it is necessary to lead the structural innovation of the blockchain by creating a verification module that can be commonly used in various blockchain platforms.

Blockchain is classified as a simple type of blockchain made of UTXO (Unspent Transaction Output) represented by Bitcoin, and a complex type of blockchain that deals with a state tree such as Ethereum. Currently, zero-knowledge proof is used only when processing some transactions in a simple form of blockchain. Zero-knowledge proof is a system that proves to the verifier that the proofer knows that knowledge without revealing the knowledge and information he knows [1]. The proofer is the subject that proves that he / she knows the knowledge, and the verifier is the subject that verifies that the proofer knows the knowledge. When zero-knowledge proof technology is applied to storage of transaction data, data storage space can be saved by compressing data in a way that pruning actual data and leaving only proof of data. As time goes by, the data of the blockchain will gradually accumulate, and accordingly, the computing resources required to operate the full node are gradually increasing.

In the case of Ethereum, it is already difficult for an individual to operate a full node, and in the future, only a large company or large hands that can have sufficient computing resources can operate the full node. These factors will lead to the centralization of the blockchain, and this problem can be solved by reducing the resources required for data storage and verification through a virtual machine with zero knowledge proof technology. Therefore, in this paper, we developed a zero-knowledge proof algorithm capable of general operation verification and designed a zero-knowledge circuit generator capable of general operation verification. Also, by applying and testing the zero-knowledge proof algorithm to the virtual machine, the performance of the transaction can be improved. Section 2 of this paper introduces related research, and Section 3 introduces domestic and foreign cases. In section 4, an algorithm capable of transaction verification is proposed, and in section 5, a zero-knowledge circuit to be applied to a virtual machine is designed. Finally, Section 6 presents conclusions and future tasks.

2.   Related Studies

Blockchain technology can be divided into a simple type of blockchain made of UTXO (Unspent Transaction Output) and a complex type of blockchain that deals with the State Tree [2]. Currently, in the simple form of blockchain, zero-knowledge proof is used at the protocol level only in some transaction processing. However, although some complex forms of blockchain use smart contracts using smart contracts, there are limitations in terms of performance and utilization because they are implemented in the upper layer. Proof size of a single operation created through the proposed SNARKs algorithm is about 1,500 bytes (1.5 kbytes) [3].

   * Bullet Proof algorithm.
   - Transaction size of UTXO-based blockchain is measured in (in * 254 * 146 + out * 254 * 33 + 10) bytes, and increases arithmetically according to the number of * in, out used.
   * It occupies about 45,000 bytes (45kbytes) based on 1 in and 1 out.
   - Regardless of the type of transaction, the transaction size can be fixed to 1.5 kbytes, and even the simplest transaction standard is more than 70% economical.
   - The blockchain-based distributed application market is expected to grow from about $ 3.2 billion in 2019 to more than $ 60 billion in 2024 (Blockchain Market Shares, Market Strategies, and Market Forecasts, 2018 to 2024, IBM, 2018). Among them, the market with 'transaction processing' as a profit model is expected to reach 55% of the total.

Since such a blockchain-based distributed application is generally provided on the basis of open source, transaction fees rather than content usage fees are inevitably accepted by users. Therefore, the economic value of the technology to efficiently process transactions is very positive. Even if all verification nodes do not participate in block verification, the general operation is verified with the same security strength as all nodes participated and verified using zero-knowledge proof technology, thereby providing the same effect as saving the entire transaction without saving all transaction data [4].

Currently, as the value of using personal information increases, discussions on how to provide personal information have been actively conducted. Currently, one of the most common methods of providing personal information is a group that uses personal information to obtain personal consent and use personal information. However, the above method has two problems. First, information that is more than the information required by the institution for personal information is being exposed. Second, whenever a company requests personal information, there is a problem that a trusted party must provide authentication information for the information to the company. In order to solve the above problems, this paper proposes a privacy-protected personal information management method using zk-SNARK (zero-knowledge Succinct Non-interactive ARgument of Knowledge) technique and blockchain[5]. zk-SNARK is a modification of the existing ZKP to be more succinct and applicable in a non-interactive environment. This logic was first proposed in 2012, and due to its characteristics, ZKP can be implemented in a blockchain environment. In the case of a blockchain transaction using zk-SNARK, the validity of the transaction can be communicated to nodes other than the sending and receiving node without exposing information such as a receiver, a sender, and a transfer amount. ZCash is the first application of zk-SNARK, and related contents

were applied to Ethereum's Byzantium hard fork [6]. The zk-SNARK is largely divided into two parts, one is the process of converting the problem to be proved to a specific form, and the other is the process of actual proofing using the converted problem. The privacy-protected personal information management technique can guarantee the privacy and reliability of information when providing personal information through zk-SNARK. In addition, it is possible to manage personal information data while ensuring the integrity of the data through the blockchain, and sharing personal information can be performed more easily than the existing authentication method. The privacy-protected personal information management technique can guarantee the privacy while ensuring privacy when providing personal information through zk-SNARK. In addition, it is possible to manage personal information data while ensuring the integrity of the data through the blockchain, and sharing personal information can be performed more easily than the existing authentication method [7].

3.   Domestic & International cases

There are several companies with blockchain virtual machine technology. The most representative virtual machine is EVM, Ethereum's virtual machine. EVM is the first blockchain virtual machine and based on EVM, Ethereum has grown into a basic platform for smart contracts, tokens, and decentralized applications (Dapps). And many blockchain projects are using Ethereum's EVM when creating the mainnet. Currently, Ethereum is planning to upgrade to Ethereum 2.0, and when Ethereum 2.0 is introduced, the current virtual machine EVM will be converted to eWASM [7].

EOS-VM is a virtual machine created by EOS.IO and is not limited to the blockchain industry, and is expected to be used in traditional software development fields such as game engines, databases, and web frameworks. EOS-VM is a virtual machine dedicated to the blockchain system, and it can be expected to save development resources (CPU), improve blockchain scalability, and improve development efficiency compared to the first blockchain virtual machine, EVM [8].

Tron's virtual machine TVM is developed based on Ethereum's EVM and is characterized by being compatible with Ethereum. By designing a unique virtual memory mechanism, the amount of memory actually used can be greatly reduced, and the operation cost of a decentralized application can be greatly reduced by providing developers with almost unlimited memory capacity. And you can save resources by optimizing the compiler.

Table 1 shows domestic and International cases as a table.

Table 1. Domestic & International cases   (source:itfind.or.kr)

Coin name	Consensus method	Characteristic	Market cap	Remark
Ethereum	EVM	Turing completeness as the first blockchain virtual machine	$ 13 billion	Focusing on decentralization and security
EOS	EOS-VM	Consensus algorithm similar to indirect democracy	$ 2.4 billion	Value for scalability
Tron	TVM	EVM-enhanced virtual machine featuring Ethereum compatibility	$ 1.8 billion	

Currently, the domestic blockchain technology is mainly biased to the underlying technologies related to the main net, such as distributed ledgers and consensus algorithms. Due to the nature of the domestic technology ecosystem, the area that can lead in the global market is the distributed application area rather than the mainnet area. And there is currently no zero-knowledge proof-based virtual machine that can efficiently verify complex operations required for distributed applications, not just bookkeeping.

Therefore, by designing a system for improving the amount of code verification based on zero-knowledge proof applicable to various distributed applications and smart contract execution environments, it will become a distributed application-based technology with great growth potential in the future [9].

4.   Algorithm for Transaction Validation

Proof of zero knowledge must satisfy the following three conditions.

  * completeness: If a condition is true, a trusted verifier must be able to understand this by a trusted prover.

  * soundness: When a condition is false, a dishonest verifier can never convince the verifier that the condition is true by lying.

  * zero-knowledge: When a condition is true, the verifier knows nothing other than the fact that this condition is true.

The study intends to utilize zero-knowledge proofs for various types of transactions that the user wants, as well as predefined types of transactions. Circuits that can produce evidence of current zero-knowledge proofs can only perform operations in a predefined form. In order to be able to utilize this in various types of transactions desired by users, a circuit capable of verifying general operations is required. General operation means universal and various operations, not specific predefined operations. Therefore, the research team researched a circuit capable of verifying general operations and designed the method to apply it to the virtual machine. In addition, by using the zero-knowledge proof technology, even if a blockchain participant does not know the contents of the block, it can quickly verify that the contents of the block are not forged or tampered with by the node performing the proof and reporting role among all nodes. Also, by rapidly increasing the block sync speed for participants, new participants can quickly join the network [10].

Fig. 1. is about the existing transaction verification and data storage method. The verification amount increases by re-executing the block for all nodes except the minor. And as the chain data connection increases, the data size also increases linearly.
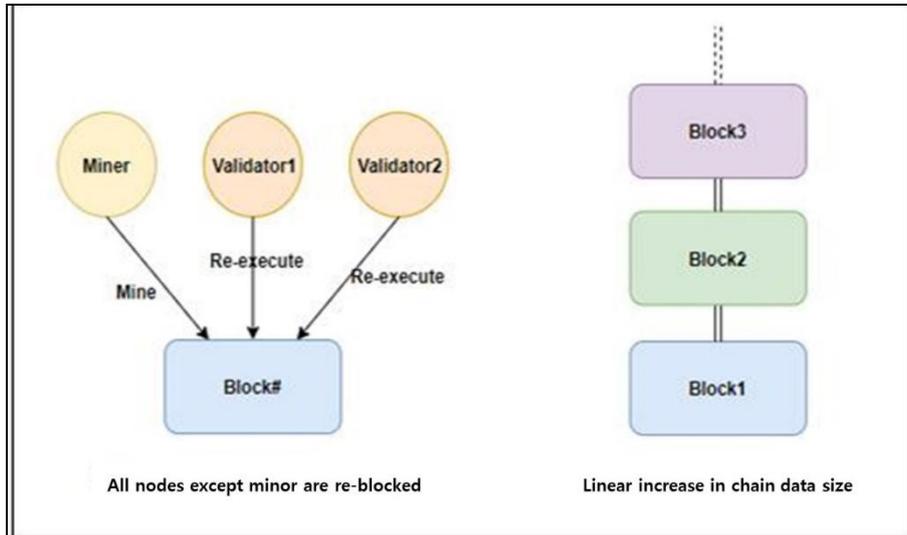
Fig. 1. Existing Transaction Validation & Data Storage Method

Fig. 2. improves the existing transaction verification and data storage method. This problem was solved by reducing the resources required for data storage and verification through a virtual machine with zero knowledge proof technology.
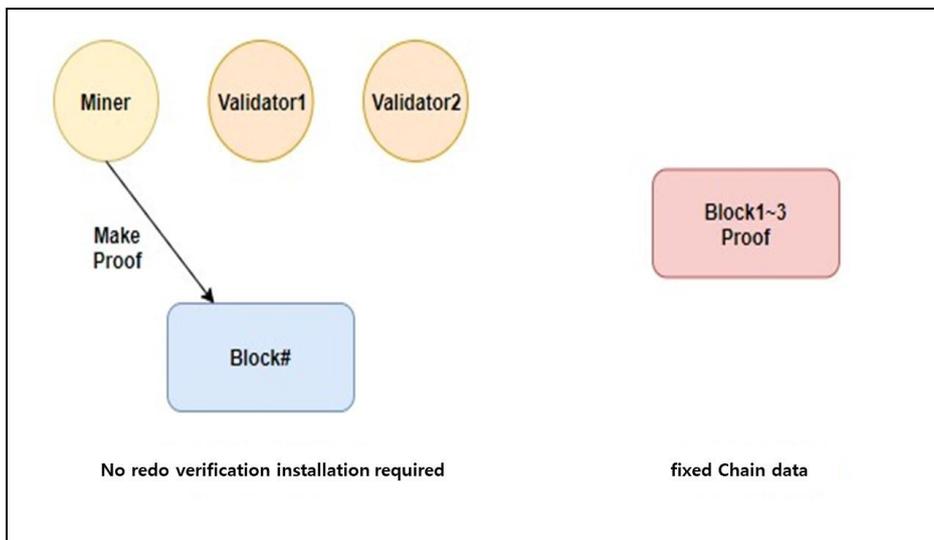
Fig. 2. Advanced Transaction Validation & Data Storage Method

The requirements of the zero-knowledge proof-based algorithm capable of general operation verification are as follows.

The circuit to be created in this study should be arithmetic. In the finite field F, the F-arithmetic-circuit takes the element in F as the input value, and the output value is also the element in F. The circuit consists of a gate and a wire, and the gate takes two numbers as inputs, adds or multiplies numbers, and outputs the result through the output wire. The wire is responsible for passing the value into and out of the gate.

Circuits must: Local Consistency Check:

Verify that all gate equations are met.

Global Consistency Check: Verify that the wires correctly connect the gates together to form the circuit.

The zk-SNARK structure considered in this study is based on cryptographic pairing.

zk-SNARK for F-arithmetic-circuit receives key generator G, attestor P, and verifier V as input values[11].

The key generator G samples the proof key pk and verification key vk using the security parameters λ and F-arithmetic-circuit C: Fn * Fh$\rightarrow$ Fl.

The above values are public parameters of the verification system that needs to be generated only once per circuit. Once set, anyone can generate non-interactive evidence using pk, and anyone can verify this evidence using vk. This is a study on a universal zero-knowledge proof algorithm that can perform general operation verification [12].

A new circuit creation method and a new zk-SNARK for the circuit are needed.

The existing zk-SNARK has the following problems.

* A new setup is required for each new program and a new key needs to be generated accordingly.

* Memory access or the number of repetitions of a loop cannot be changed depending on the input value of the program.

* Even if you allow data dependence on memory access, you have to go through heavy tasks such as verifying the Merkle Pass.

* Circuit size increases inefficiently in proportion to the program size even if an arbitrary program is supported.

The universal circuit should run on any program with less than l instructions, less than T machine steps, and less size. And it should be available in all cases with one key generation. Accordingly, the circuit that satisfies this must satisfy Cl,n,T is O(l+n+T)*log(l+n+T))gates [13].

Previously studied universal zero-knowledge proof algorithms, the size of data increases to the size of l * T according to the program size. In this case, as the program size increased, the storage cost increased significantly.

In the case of the newly created zero-knowledge proof algorithm, the size of the data increases in the form of l + T.

The circuit generator and zero-knowledge algorithm are independent of each other. If the circuit generator and the zero-knowledge proof algorithm to be applied to the circuit are independent, a more flexible system can be built [14].

Fig. 3. shows a combination of two elements, a circuit generator and a key generator, for general operation verification. The output C of the circuit is universal because it does not depend on the program or main input values, but only on the l, n, and T values. In this case, as the program size increased, the storage cost increased significantly. In the case of the newly created zero-knowledge proof algorithm, the size of the data increases.

The circuit generator and zero-knowledge algorithm are independent of each other. If the circuit generator and the zero-knowledge proof algorithm to be applied to the circuit are independent, a more flexible system can be built. Fig. 3. shows a combination of two elements, a circuit generator and a key generator, for general operation verification. The circuit's output C is universal because it depends only on the values, not on the program or the main input values. When combined with a circuit verification system such as zk-SNARK, the parameters of the verification system are also universal. In this case, all programs can be verified with a single key generation, and after that, a key suitable for a given calculation range can be selected. Therefore, the cost of generating keys for each program can be reduced [15].
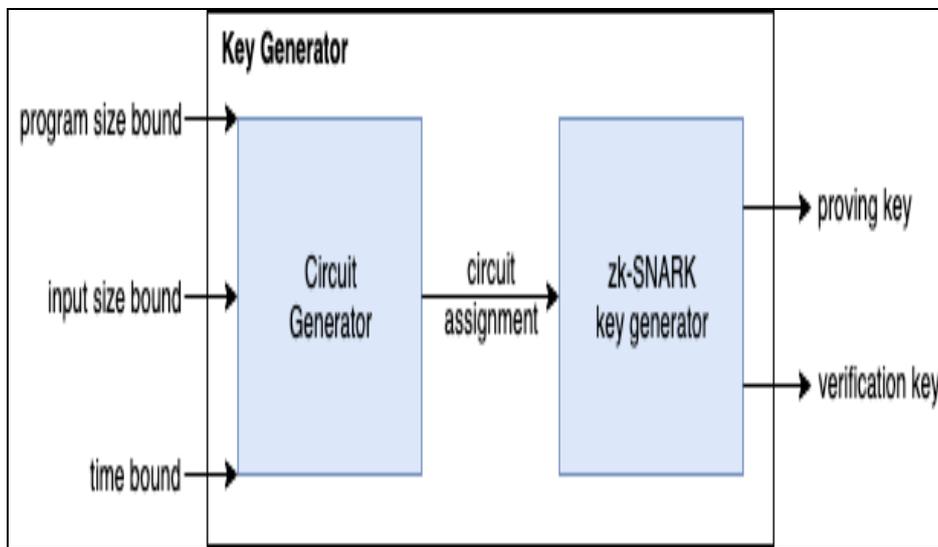
Fig. 3. Key Generator for General Operation Validation

In Fig. 4. permission for block generation is granted through a validator and a verifier for general operation verification.
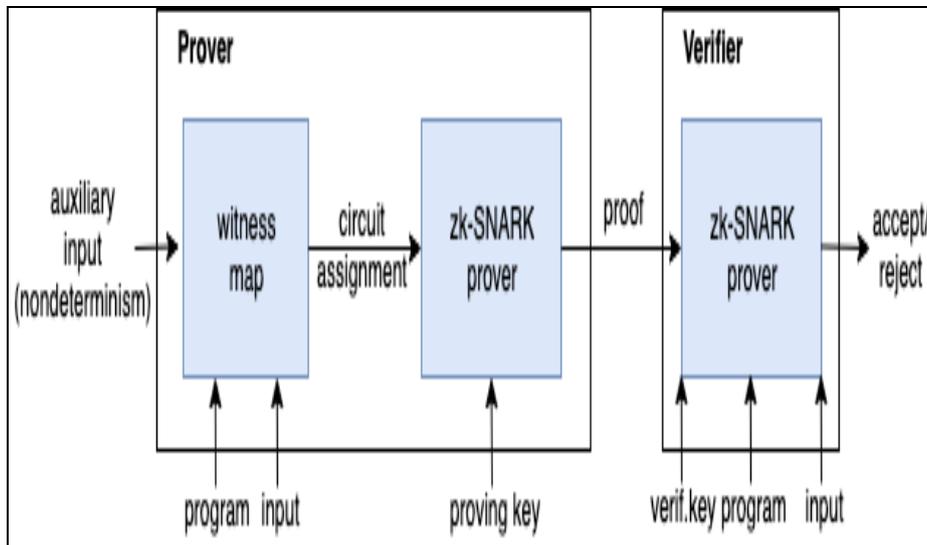
Fig. 1. Prover and verifier for general operation validation

The verifier V takes the verification key vk, $\vec{x} \in$ ı , and the evidence π as input values to verify that the evidence π is valid. The operation on V consists of two parts[16].

* Calculate $vk_{\vec{x}} := vk_{IC,0} + \sum_{i=i}^{n} x_i vk$ by entering part of vk and .

* Enter vk, v and π to be able to calculate 12 pairings and perform necessary checks .Regarding the first part of V, the variable-based multi-scalar multiplication technique can be used to reduce the amount of computation required for v calculation. With respect to the second part of V, even if the pairing evaluation takes a certain amount of time regardless of the input size n, these evaluations are very expensive and dominate for the small n. Our goal is to minimize the cost of these pairing assessments [17].

The research for the optimizer optimization is as follows.

The proofer P takes the proof key PK(including circuit C), $\vec{x} \in$ ιand the witness $\vec{a} \in$ as input values. Proofer P creates evidence π and testifies that it is $\vec{x} \in$.

The operation on P consists of two parts [18].

* Calculate the coefficient $H(z) := \frac{A(z)B(z) - (}{Z(z)}$ of the polynomial , where $A, B, C \in F_y$ is derived from the QAP instances $(\vec{A}, \vec{B}, \vec{C}, Z) := QAP_{inst}($ and QAP evidence $\vec{s} := QAP_{wit}(C, \vec{x},$.

* Calculate π by using coefficient , evidence of QAP, and public key pk. In particular, in relation to the first part of P, the coefficient T is efficiently calculated through the FFT technique of [BCGTV13a].

5.   Design of Zero-knowledge circuit

Due to the nature of the blockchain that stores all transaction data, the data on the blockchain continues to increase over time. When zero-knowledge proof technology is applied to storage of transaction data, data storage space can be saved by compressing the data by pruning actual data and leaving only proof of data. As time goes by, the data of the blockchain will gradually accumulate, and accordingly, the computing resources required to operate the full node will gradually increase [19].

In the case of Ethereum, it is already difficult for an individual to operate a full node, and in the future, it is expected that only large companies or large hands that can have sufficient computing resources can operate the full node. This will cause the centralization of the blockchain, and this problem can be solved by reducing the resources required for data storage and verification through a virtual machine with zero knowledge authentication technology [20].

Fig. 5. is designed to apply the zero-knowledge proof algorithm to the virtual machine. It shows the flow of the operation method of the Ethereum virtual machine for applying zero knowledge proof technology. Since Solidity, the smart contract language of Ethereum, is a language created for human understanding, it needs to be changed to a machine language understandable by a virtual machine in order to operate in a virtual machine. Code written in Solidity is converted to Ethereum bytecode by the compiler. This bytecode is executed by EVM, Ethereum's virtual machine. When a specific bytecode is executed, all nodes in the Ethereum network execute the same bytecode respectively to verify the transaction. At this time, if zero-knowledge proof technology that can perform general operation verification is applied to the virtual machine, even if the virtual machine does not execute the transaction, it is possible to know whether the corresponding transaction is the correct transaction by performing verification on the zero-knowledge evidence [21].
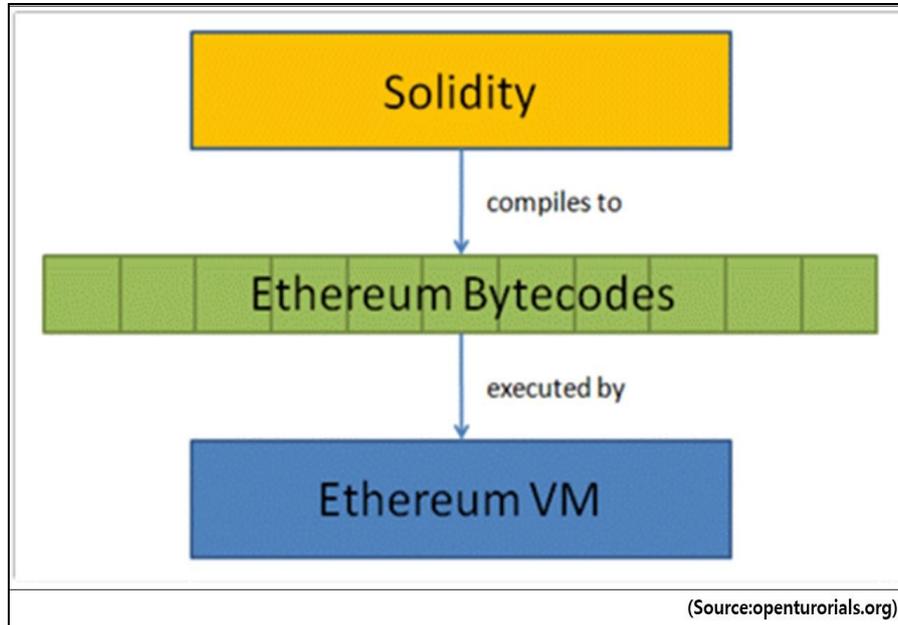
Fig.5. Execution process of Ethereum virtual machine

In order to modify the virtual machine, it is necessary to understand the structure. Therefore, Figure 6 shows the architecture and execution flow diagram of the Ethereum virtual machine. Once you understand how the virtual machine is running, you need to figure out what parts of the virtual machine need to be modified to apply zero-knowledge techniques [22].
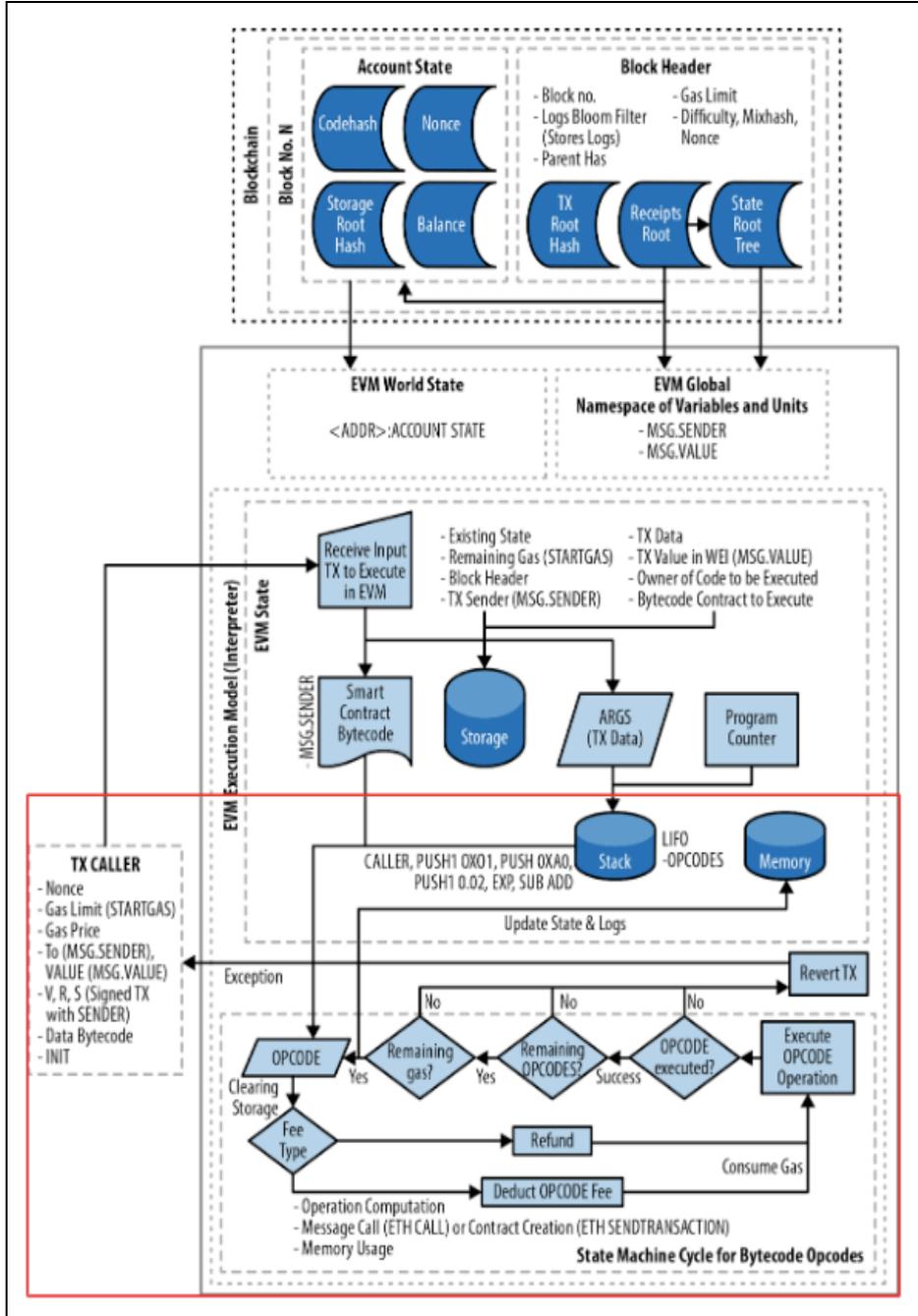
Fig. 6. Architecture &Excution Flow Chart of Ethereum Virtual Machine

In order to execute the transaction, it needs to be changed to Ethereum bytecode as mentioned. These bytecodes are decomposed into what are called opcodes, stacked on the stack and executed one by one. You must subtract the gas cost for running the virtual machine before the opcode runs. The opcodes are now executed if the gas cost is not insufficient. Fig. 7. shows the parts that need to be changed in the virtual machine when the opcodes are executed [23].
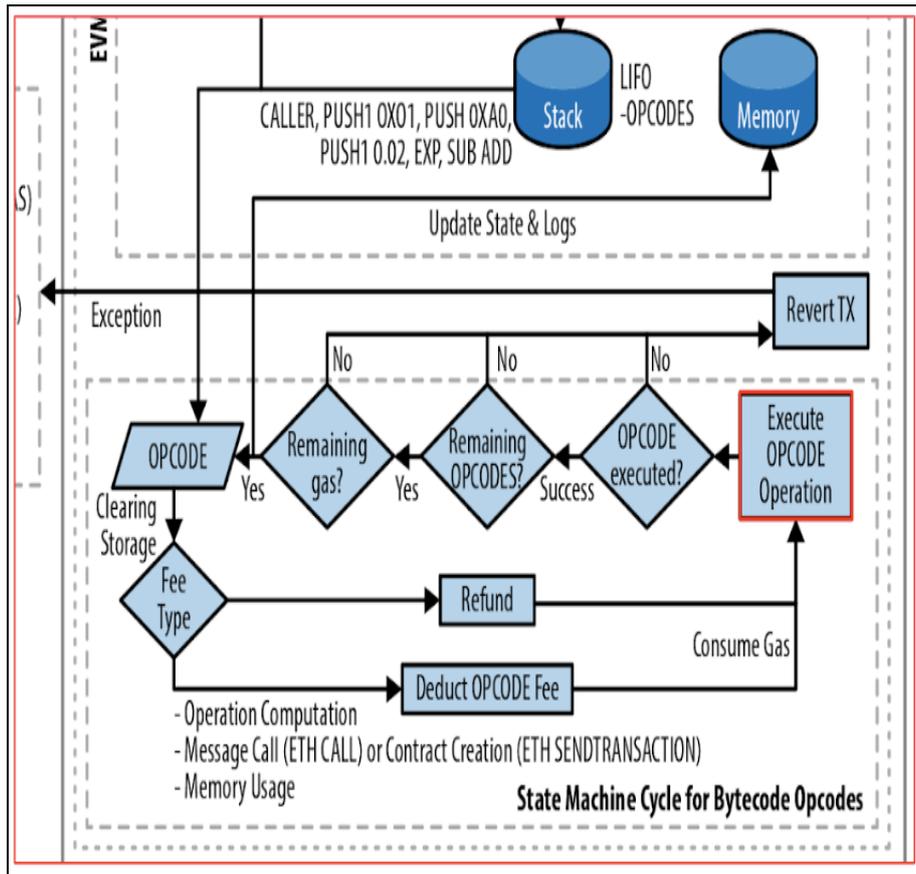
Fig. 7. Change Part in Virtual Machine

In the figure above, the part marked with a red box is the part to which zero-knowledge proof technology should be applied, and the part to create a universal circuit that can execute the opcode. Fig. 8. shows the change in data stored after the zero-knowledge proof technology is applied. After the zero-knowledge proof technology is applied to the part that performs the opcode, TX data among the data stored in the existing storage is replaced with the proof of the zero-knowledge proof. And we will create and test a virtual machine with zero knowledge proof [24].
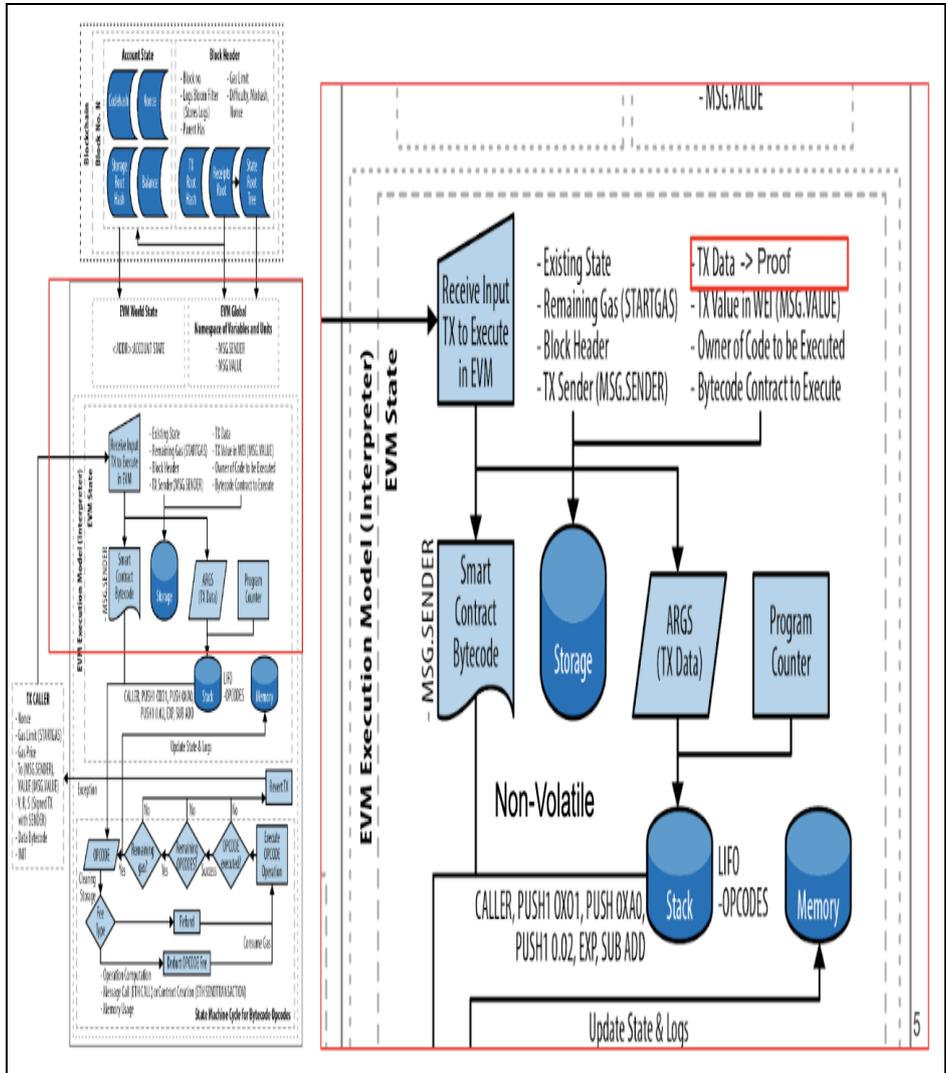
Fig. 8. Data Changes after zero knowledge proof technology

6.   Conclusion and future works

In this paper, a circuit with zero-knowledge proof algorithm for general operation verification was developed. The core of this paper solved the problem of increasing chain data size and block verification amount in the existing system. In addition, a zero-knowledge proof algorithm was designed for general operation verification [25].

Finally, a study was conducted to optimize the validator and the validator, and a study was conducted to optimize the key generation. This study is a practical example of applying the zero-knowledge proof algorithm capable of semi-operational verification, and can develop two different blockchains in the future. In addition, Crypto Currency implementation with zero-knowledge proof algorithm for general operation verification can be developed. It can also create new blockchain business opportunities, such as platform services, where DApps linked to public blockchains can be integrated with each other. However, anyone can participate in the blockchain network if a blockchain with a zero-knowledge proof-based virtual machine capable of verifying general operations is born [26].

When developing the world's first virtual machine technology based on zero-knowledge proof that can be applied to various distributed applications and smart contract execution environments, Korean companies will secure the foundation technology that can lead the global market in the distributed application market, which has great growth potential in the future. . In addition, new blockchain business opportunities are provided, such as platform services where DApps linked to the public blockchain can be integrated with each other. In addition, exports of related products and services will be expanded by revitalizing the blockchain industry and strengthening cooperation with global companies based on core technologies. The increase in expertise of domestic blockchain R & D personnel and the internalization of technology development will also create jobs for R & D personnel. Opportunities for technological innovation on the blockchain and high potential for use in other fields are also provided.

As a future task, based on this design, we will develop a zero-knowledge proof system capable of general operation verification. And based on this research, we will develop our own platform.

References

1. Eli Ben-Sasson, Alessandro Chiesa, EranTromer and MadarsVirza. "Succinct Non-Interactive Zero Knowledgefor a von Neumann Architecture", https://eprint.iacr.org/2013/879.pdf
2. Eli Ben-Sasson, Alessandro Chiesa, Daniel Genkin, and EranTromer. Fast reductions from RAMs to delegatablesuccinct constraint satisfaction problems. In Proceedings of the 4th Innovations in Theoretical Computer ScienceConference, ITCS '13, pages 401–414, 2013.
3. P. Valiant. "Incrementally verifiable computation or proof of knowledge imply time/space efficiency", In: Theory of Cryptography. Ed. by Ran Canetti. Berlin, Heidelberg: Springer, 2008, pages 1–18. isbn: 978-3-540-78524-8
4. E. B-Sasson, A. Chiesa, E. Tromer and M. Virza. "Scalable zero knowledge via cycles of elliptic curves (extended version)". In: Advances in Cryptology - CRYPTO 2014. Vol. 8617.
5. S. Bowe, J. Grigg and D. Hopwood, "Halo: Recursive Proof Composition without a trusted setup", https://eprint.iacr.org/2019/1021.pdf

6. Eli Ben-Sasson, Alessandro Chiesa, EranTromer and MadarsVirza. "Succinct Non-Interactive Zero Knowledgefor a von Neumann Architecture", https://eprint.iacr.org/2013/879.pdf

7. Eli Ben-Sasson, Alessandro Chiesa, Daniel Genkin, and EranTromer. Fast reductions from RAMs to delegatablesuccinct constraint satisfaction problems. In Proceedings of the 4th Innovations in Theoretical Computer ScienceConference, ITCS '13, pages 401–414, 2013.

8. P. Valiant. "Incrementally verifiable computation or proof of knowledge imply time/space efficiency", In: Theory of Cryptography. Ed. by Ran Canetti. Berlin, Heidelberg: Springer, 2008, pages 1–18. isbn: 978-3-540-78524-8

9. E. B-Sasson, A. Chiesa, E. Tromer and M. Virza. "Scalable zero knowledge via cycles of elliptic curves (extended version)". In: Advances in Cryptology - CRYPTO 2014. Vol. 8617.

10. S. Bowe, J. Grigg, D. Hopwood, "Halo: Recursive Proof Composition without a trusted setup", https://eprint.iacr.org/2019/1021.pdf

11. Eli Ben-Sasson, Alessandro Chiesa, EranTromer and MadarsVirza. "Succinct Non-Interactive Zero Knowledgefor a von Neumann Architecture", https://eprint.iacr.org/2013/879.pdf

12. Eli Ben-Sasson, Alessandro Chiesa, Daniel Genkin, and EranTromer. Fast reductions from RAMs to delegatablesuccinct constraint satisfaction problems. In Proceedings of the 4th Innovations in Theoretical Computer ScienceConference, ITCS '13, pages 401–414, 2013.

13. P. Valiant. "Incrementally verifiable computation or proof of knowledge imply time/space efficiency", In: Theory of Cryptography. Ed. by Ran Canetti. Berlin, Heidelberg: Springer, 2008, pages 1–18. isbn: 978-3-540-78524-8

14. E. B-Sasson, A. Chiesa, E. Tromer and M. Virza. "Scalable zero knowledge via cycles of elliptic curves (extended version)". In: Advances in Cryptology - CRYPTO 2014. Vol. 8617.

15. S. Bowe, J. Grigg, D. Hopwood, "Halo: Recursive Proof Composition without a trusted setup", https://eprint.iacr.org/2019/1021.pdf

16. Shashank Agrawal, Chaya Ganesh and PaymanMohassel, "Noninteractive zero-knowledge proofs for composite statements", Annual International Cryptology Conference, pp. 643-673, 2018.

17. Benedikt Bünz, Jonathan Bootle, Dan Boneh, Andrew Poelstra, Pieter Wuille and Greg Maxwell, "Bulletproofs: Short proofs for confidential transactions and more" in Bulletproofs: Short Proofs for Confidential Transactions and More, IEEE, pp. 0, 2018.

18. J. Katz, V. Koilesnikov and X. Wang, "Improved non-interactive zero knowledge with applications to post-quantum signatures", University of Maryland and Georgia Tech, March 2019.

19. C. P. Sah, K. jha and S. Nepal, "Zero-knowledge proofs technique using integer factorization for analyzing robustness in cryptography", Proceedings of the 10th INDIACom; INDIACom-2016 3rd International Conference on "Computing for Sustainable Global Development, 2016.

20. S. J. et al., "Blochie: A blockchain-based platform for healthcare information exchange", 2018 IEEE International Conference on Smart Computing (SMARTCOMP), pp. 49-56, June 2018.

21. X. He, S. Alqahtani and R. Gamble, "Toward privacy-assured health insurance claims", 2018 IEEE International Conference on Internet of Things (iThings), pp. 1634-1641, July 2018.

22. D. C. Snchez, Zero-knowledge proof-of-identity: Sybil-resistant anonymous authentication on permissionless blockchains and incentive compatible strictly dominant cryptocurrencies, 2019.

23. D. C. N. et al., "Blockchain for secure ehrs sharing of mobile cloud based e-health systems", IEEE Access, vol. 7, pp. 66792-66806, 2019.

24. S. Sharaf and N. F. Shilbayeh, "A secure g-cloud-based framework for government healthcare services", IEEE Access, vol. 7, pp. 37876-37882, 2019.

25. A. e. a. Al Omar, "Medibchain: A blockchain based privacy preserving platform for healthcare data" in Security Privacy and Anonymity in Computation Communication and Storage, Cham: Springer International Publishing, pp. 534-543, 2017.
26. D. Nunez, I. Agudo and J. Lopez, "Proxy re-encryption: Analysis of constructions and its application to secure access delegation", Journal of Network and Computer Applications, vol. 87, pp. 193-209, 2017.

Soonhyeong Jeong is a co-organizer of <Seoul Ethereum Meetup>, and this meetup started from November 15, 2014(even before the launching of ethereummainnet). Kevin is a CEO at the Ethereumblockchain R&D startup OntherInc in South Korea. He received a bachelor's degree in economics and a master's degree in Software from Graduate School of Software in SoongSil University in Seoul. Nowadays, he's interested in Plasma, EVM and Gas System and privacy in public blockchain.

Byeongtae Ahn received his B. Eng. Degree in Digital Media Technology in 2017 from North China University of Technology, Beijing, The People's Republic of China. Since September 2017, he is in the M. Eng. course at the Department of Multimedia Engineering, Dongguk University, Seoul, Republic of Korea. He has done some works mainly associated with NUI (Natural User Interface) applications, VR (Virtual Reality) applications and artificial intelligence with deep learning.

Image Target Detection Algorithm Compression and Pruning Based on Neural Network

Yan Sun[a] and Zheping Yan[b, *]

College of Automation, Harbin Engineering University,
Harbin 150001, China
{[a] aidenby, [b] feiyunsy1213}@163.com

Abstract. The main purpose of target detection is to identify and locate targets from still images or video sequences. It is one of the key tasks in the field of computer vision. With the continuous breakthrough of deep machine learning technology, especially the convolutional neural network model shows strong Ability to extract image feature in the field of digital image processing. Although the model research of target detection based on convolutional neural network is developing rapidly, but there are still some problems in practical applications. For example, a large number of parameters requires high storage and computational costs in detected model. Therefore, this paper optimizes and compresses some algorithms by using early image detection algorithms and image detection algorithms based on convolutional neural networks. After training and learning, there will appear forward propagation mode in the application of CNN network model, providing the model for image feature extraction, integration processing and feature mapping. The use of back propagation makes the CNN network model have the ability to optimize learning and compressed algorithm. Then research discuss the Faster-RCNN algorithm and the YOLO algorithm. Aiming at the problem of the candidate frame is not significant which extracted in the Faster-RCNN algorithm, a target detection model based on the Significant area recommendation network is proposed. The weight of the feature map is calculated by the model, which enhances the saliency of the feature and reduces the background interference. Experiments show that the image detection algorithm based on compressed neural network image has certain feasibility.

Key words: Convolutional Neural Network, Target Retrieval, Deep Learning, Algorithm Compression

1. Introduction

Images are an important way for humans to access information [1]. With the development of science and technology, image generation is getting faster and faster [2]. Computer image recognition plays an important role in many industries and is a hot topic in current research. Traditional image recognition technology uses artificial feature selection, pattern matching, linear classification and other algorithms for image recognition. The accuracy of the identification depends to a large extent on the quality of the selected features. It is difficult to extract features that express the nature of the

---

[*] Corresponding author

original data. Compared with traditional image recognition methods, deep learning has great advantages [3]. A common method is to construct a deep neural network and train it with certain data [4]. The deep neural network can automatically extract image features and achieve better recognition results. Convolutional neural networks are the key network for deep learning and image recognition [5]. By constructing a convolutional neural network, image features can be extracted layer by layer. Network construction and training methods are the key to neural network recognition. Excellent network design can achieve better training results with fewer parameters. Special network components speed up the training process. Appropriate training methods can leverage the capabilities of the network [6].

However, due to the complexity of the real world background and the diversity of scenes, as well as the occlusion and low resolution of the targets in the acquired images, the target detection technology becomes a challenging subject [7]. The main challenges facing current target detection technologies include how to reduce the impact of target size and shape on detection, how to improve the accuracy of target positioning, and how to reduce background interference. The commonly used evaluation indicators for target detection systems are detection accuracy and speed. In order to improve the detection accuracy, the target detection system needs to be able to effectively eliminate the interference of background, light, noise and other factors [8]. In order to improve the detection speed and realize real-time target detection, the target detection system needs to be able to simplify the detection process and image processing algorithms. Since the traditional target detection algorithm is manually designed [9], its accuracy cannot be adapted to various scenes, and the detection speed is slow. Considering the current research status and technical level at home and abroad, this paper will use convolutional neural network to perform object compression in algorithmic compression and pruning for deep learning [10].

Yang et al. extended the clutter model from complex feature vector to complex feature subspace, which is suitable for non-uniform patching regions, and derives extended PTD and extended GP-PNF [11]. Yang et al. proposed a novel supervised target detection algorithm that uses a single target spectrum as a priori knowledge. His proposed algorithm uses TV to maintain the spatial uniformity or smoothness of the detected output. At the same time, the constraints are used to guarantee the spectral characteristics of the unsuppressed target [12]. The final test model is the l1 norm convex optimization problem. The split Bregman algorithm is used to solve the optimization problem because it can effectively solve the l1 norm optimization problem, two synthesis and two real hyperspectral images are used for experiments [13]. Zhao et al. proposed a new hyperspectral image (HSI) target detection method, which uses St OMP reconstruction algorithm [14]. When the computational cost of the conventional sparse detection algorithm is very high, since the HSI usually has a large amount of data, St OMP can be used. The sparse representation algorithm has been successfully applied to the HSI target detection field and has achieved good results. The method improves the steps of solving the sparse coefficient, reduces the number of iterations of the process, significantly improves the detection efficiency and reduces the computational cost [15].

This paper first studies the target detection of convolutional neural networks. Its most widely used target detection deep learning model is constructed by the human visual system. Each layer of the convolutional neural network is described in detail. Of course, the convolutional neural network mainly includes forward propagation and back

propagation. After training and learning, the forward propagation mode will appear in the application of CNN network model, providing image feature extraction, integration processing and feature mapping for the model. The use of back propagation makes the CNN network model have the ability to optimize learning and The algorithm is compressed. Then the research on the Faster-RCNN algorithm and the YOLO algorithm is discussed. Aiming at the problem that the candidate frame extracted in the Faster-RCNN algorithm is not significant, a target detection model based on the attention area recommendation network is proposed. By paying attention to the model to calculate the weight of the feature map, the saliency of the feature is enhanced and the background interference is weakened. Experiments show that the image detection algorithm based on compressed neural network image has certain feasibility.

2.   Proposed Method

2.1.   Target Detection Theory Based on Convolutional Neural Network

Convolutional neural network

Convolutional neural network is the most widely used target detection deep learning model constructed by human visual system. Convolutional neural networks are a deep learning model of multi-layer neuron connections. Different features can be obtained by convoluting the input images of different convolution kernels. Convolution can be input directly from the original image without complex image preprocessing, so it is widely used. The common layer structure of the convolutional neural network includes a convolutional layer, a pooled layer, a fully connected layer, an activation function layer, and a classification layer. The structure and function of each layer are described in detail below.

As the main structure of extracting image features in CNN network model, convolutional layer plays an important role in the success of CNN network model. The sparse connection mechanism and sharing scheme in the convolutional layer can control the number of parameters and calculations of the convolutional layer to an acceptable range. It is assumed that the L layer is a convolutional layer and the L-1 layer is a pooling layer or an input layer. Then the formula for calculating the level L is (1):

$$x_j^1 = f\left(\sum_{i \in Mj} x_i^{l-1} * k_{ij}^l + b_j^l \right) \tag{1}$$

Wherein, the j-th feature map of the L-th layer of the x_j^l th type, the right side of the formula (1) is a convolution operation for all the associated feature maps x_j^{l-1} of the L-1 layer and the j-th convolution kernel k_{ij}^l of the L-th layer And, plus an offset parameter, and finally enter the excitation function.

The sampling layer, also known as the pooling layer, is an important part of the CNN network model, which can reduce the resolution of image features and the computational complexity of the model. In general, it appears after the convolutional layer to perform statistical operations on the small-scale image features extracted from the previous convolution. The sampling layer is static and its parameters do not involve correction of back propagation. There are two common sampling layers: maximum and average. Common pool operations include average pools and maximum pools. As the name suggests, the mean pool selects the average of all values in the slice as the output of the slice, while the maximum pool selects the maximum value of all values in the slice as the output of the slice. For example, maximizing the output of the previous convolutional layer retains only 25% of the activation information, which is the most commonly used pooling method in most convolutional neural networks.

The fully connected layer is usually located at the top of the entire CNN network model as a classifier for the entire model. The input of each node of the fully connected layer is connected with each output node of the previous layer, and the image features extracted by the previous one layer and the pooled layer are mapped to the label space of the sample. Due to the large number of parameters in the CNN network model, over-fitting is caused when training the CNN network model, resulting in insufficient robustness of the model. In order to avoid overtraining and fitting of the CNN network model, the Dropout operation is usually performed at the fully connected layer.

The main role of the activation layer is to give the model a nonlinear fit. In image processing, the convolution operation actually assigns a weight to each pixel in the input layer. Obviously, this is a linear model, but the high-level semantic information of the image is mostly nonlinear, so the model should introduce the ability to fit nonlinearity [16]. This is the case. There are many activation functions, such as sigmoid (Figure 1 from the network, *http://www.baidu.com*), hyperbolic tangent, rectified linear unit, and parametric rectification linear unit (PRELU).
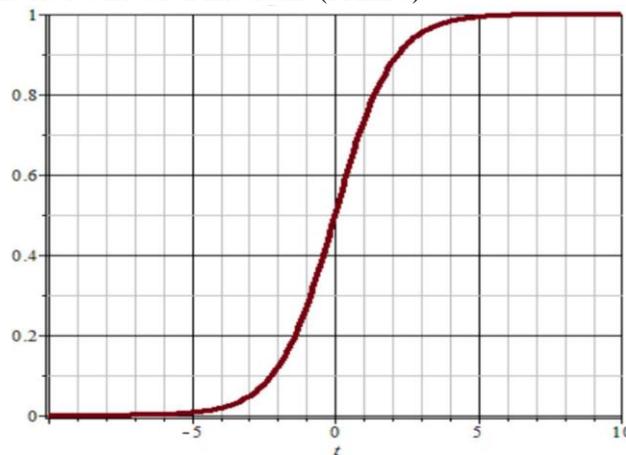
Fig. 1. Sigmoid function

The sigmoid function is a very common function in statistical learning. Its main function is that the model is relatively simple and the algorithm is easy to learn. In machine learning, for large-scale learning tasks, generally better results can be achieved. Sigmoid has also been widely used as an activation function in neural networks. The

advantage is that the derivation is simple and stable, and generally good results can be obtained. However, due to the asymmetry Sigmoid function, function value with respect to the origin is always greater than zero, the average of the output value is always greater than zero, which will reduce the speed of neural network training [17].

The ReLU activation function, as shown in (2), was proposed by Krizhevsky A. It is very popular in neural networks. The expression of the activation function in the network is sparser than the s-type function. Under a random gradient, ReLU decreases rapidly. The s-type function contains many exponential operations, which increases the computational complexity of network training. ReLU can be simpler. Most importantly, this activation function effectively alleviates the gradient dispersion problem of the s-type function. Surprisingly, it also performs well under unsupervised training.

$$y = \max(0, x) = \begin{cases} 0 & x \leq 0 \\ x & x > 0 \end{cases} \tag{2}$$

The primary function of the classification layer is to map the output of the entire connection layer to the probability that the input image belongs to a different category. The Logistic function and the Softmax function are the most commonly used classification functions in the classification layer [18]. The Logistic classification function is based on the Bernoulli distribution. It is suitable for binary classification problems. Its function expression is as follows (3).

$$\sigma(z) = \frac{1}{1 + e^{-z}} \tag{3}$$

Considering its output as the probability of t=1, t=0, you can get (4) and (5):

$$P(t = 1 \mid z) = \sigma(z) = \frac{1}{1 + e^{-z}} \tag{4}$$

$$P(t = 0 \mid z) = 1 - \sigma(z) = \frac{e^{-z}}{1 + e^{-z}} \tag{5}$$

The Softmax classification function is mainly used to solve multi-class classification problems. If the dimension of the input vector Z C will be a Softmax classification function, where C is the classification of the category and the output of the quantity is the dimension of a vector, but the elements of the probability vector at this time represent the input image, belonging to each category (6). When the denominator fill uses a regular term, the range of all outputs is limited to (0, 1).

$$y_c = \upsilon(Z)_c = \frac{e^{z_c}}{\sum_{d=1}^{c} e^{z_d}} \tag{6}$$

Compression of target detection algorithm based on convolutional neural network

Convolutional neural networks mainly include forward and backward propagation (BP). After training and learning, the application of CNN network model will appear in forward propagation mode, providing image feature extraction, comprehensive processing and feature mapping for the model [20]. The layers of the entire CNN network model and their parameters will participate in the forward propagation. Forward propagation is the main mode of propagation of the CNN network model. Forward propagation only occurs during the training and learning phases of the CNN network model. The parameters of the CNN network model are adjusted layer by layer using a correlation algorithm such as stochastic gradient descent (SGD), so that the output of the loss function in the CNN network model is gradually reduced. The use of backpropagation enables the CNN network model to have the ability to optimize learning. Backpropagation is an indispensable way of spreading the CNN network model. The function can be expressed as J (W, b), and the mathematical expression is as follows (7):

$$\mathrm{J}(W, b) = \frac{1}{n} \sum_{i=1}^{n} J(W, b, x^{i}, y^{i}) = \frac{1}{n} \sum_{i=1}^{n} (\frac{1}{2} \mid\mid h_{w,b}(x^{i}) - y^{i} \mid\mid^{2}) \quad (7)$$

Where W and b are the weights and offsets of the convolutional neural network, and the equations for updating the weight parameters and the bias terms are expressed as follows (8)（9）, where $W_{i,j}^{l}$ represents the i-th in the first layer in the convolutional neural network model. Enter the weight to connect to the Jth neuron. b_{i}^{l} denotes an offset term of the first input connected to the i-th neuron in the first layer in the convolutional neural network model.

$$W_{i,j}^{l} = W_{i,j}^{l} - \alpha \frac{\partial}{\partial W_{i,j}^{l}} J(W, b) \quad (8)$$

$$\mathrm{b}_{i}^{l} = \mathrm{b}_{i}^{l} - \alpha \frac{\partial}{\partial \mathrm{b}_{i}^{l}} J(W, b) \quad (9)$$

2.2.   Faster-RCNN Detection Algorithm

Faster-RCNN algorithm

The fast RCNN algorithm implements the real end-to-end target detection calculation process [19], which is mainly divided into three parts: convolutional neural network; regional recommendation network (RPN); Fast R-CNN target detection network. The algorithm still continues the idea that R-CNN first recommends regional reclassification. However, it successfully implements the task of using a convolutional neural network to recommend regions without using other algorithms. RPN and FastR-

CNN share the convolutional neural network for feature extraction, which reduces the number of convolution calculations and improves the speed of the whole algorithm. The R-CNN algorithm requires the extraction of features from each of the recommended candidate regions during training and testing, which requires a lot of practice and is expensive both in time and space. An important reason for these problems is that there is no shared convolution calculation [21].

R-CNN algorithm convolving all input image region, but it must be present in the overlapping part candidate region. Thus, this would create a convolution calculation of the overlap region, increasing the computational complexity of the overall testing process. Then, the FastR-CNN algorithm improved these issues. Although they all use a selective search algorithm to calculate candidate regions, the locations of convolution calculations are different. FastR-CNN first calculates the convolution of the whole image, then fuses the recommended candidate region selection search algorithm and convolution calculation feature mapping network, and obtains the candidate region of the corresponding feature vector through the RoIPooling layer, which greatly reduces the convolution calculation of operation sharing. . Reduced the number of convolution calculations. The dimensionality of these feature vectors is uniform, which facilitates subsequent classification work. FastR-CNN is inspired by SPP-Net. The proposed pooling layer pools the convolution feature and the candidate region boundary to obtain the feature vector of the corresponding region. This is equivalent to a special SP network space pyramid pool with only one pyramid structure. This is the case. In addition, in order to achieve better training results, FastR-CNN also uses some methods to accelerate, two of the most important methods are: multi-task training and minimum batch sampling.
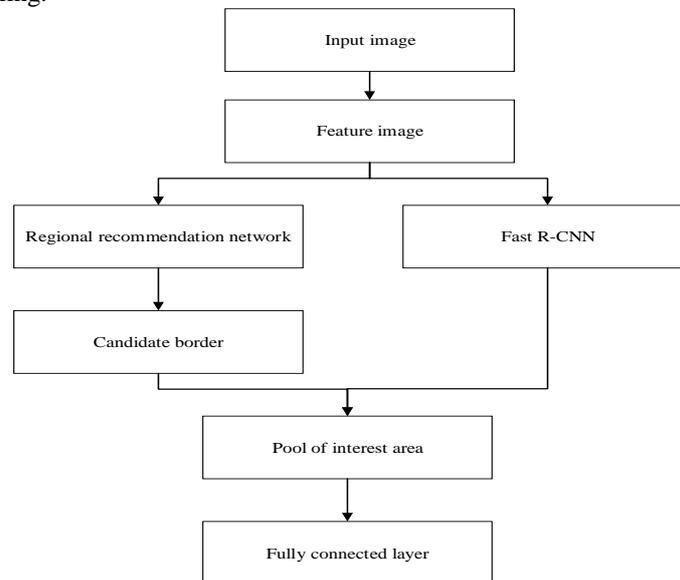
Fig. 2. Faster R-CNN algorithm flow chart

Based on the previous experience of R-CNN and FastR-CNN, the convolutional network is further used to implement the regional recommendation process. The FasterR-CNN network model does not require additional algorithms to recommend

regions. It is an end-to-end complete network model. In the FasterR-CNN algorithm, the input image is directly sent to the convolutional neural network, the feature image of the image is calculated, and then the final convolved feature image is sent to the RPN network of the regional recommendation network to recommend the candidate region. The recommended candidate region frames and their corresponding feature regions are then passed through the FastR-CNN algorithm. The proposed RoIPooling layer is merged into a feature vector of a fixed dimension. Finally, like the FastR-CNN algorithm, the feature vector is input into the classifier and the boundary regression calculator for parallel classification and boundary regression correction. The FasterR-CNN network model integrates the regional recommendation process. The RPN network shares the convolution calculation with FastR-CNN, calculates the image convolution characteristics, and forms an end-to-end target detection model, as shown in Figure 2.

It can be seen that the three algorithms of the R-CNN series are getting faster and better. One of the important reasons is the degree of sharing of convolution calculations. Representative algorithms for primary image detection based on convolutional neural networks are YOLO and SSD. The main features of the algorithm are fast, simple, real-time, but its accuracy is far lower than the R-CNN series. YOLO is one of the representatives of the first-level image detection algorithm. There are three main innovations. The first is to train end-to-end networks. Secondly, the extraction method of the regional recommendation box is improved. The third is to achieve real-time detection results. The YOLO image detection algorithm consists of 24 concave and convex faces and 2 fclayers. This is the case. The difference between YOLO and RCNN, Fast-RCNN and FasterRCNN is reflected in two aspects. The first aspect is that YOLO does not extract the region candidate box steps, saving the detection time of the model algorithm. The second aspect is to obtain the location and type of all objects in the input image and its confidence through the YOLO framework. But the shortcomings are also obvious. There are four largest pool layers in the network framework, so the image features learned through the feature extraction network are not detailed enough, which has a great impact on the accuracy of small target detection.

The YOLO detection model divides the input image into S*S grids, each of which is responsible for predicting the target information of the center. Each grid prediction target belongs to a certain class B bounding box information and C probability. The information of the bounding box includes the spatial position (x, y) and size (w, h) of the bounding box and the confidence level of the bounding box. Confidence is used to reflect whether the current border contains objects, and if there is an object, the degree of confidence is accurate. In order to transform the target detection task into a regression problem, the YOLO detection model has a vector design loss function with a special network output format, as shown in equation (9):

$$L = \lambda_1 \sum_{i=0}^{S^2} \sum_{j=0}^{B} \prod_{ij}^{obj} \left[(x_i - x_i^{'})^2 + (y_i - y_i^{'})^2 \right.$$

$$+ \lambda_2 \sum_{i=0}^{S^2} \sum_{j=0}^{B} \prod_{ij}^{obj} \left[(\sqrt{w_i} - \sqrt{w_i^{'}})^2 + (\sqrt{h_i} - \sqrt{h_i^{'}})^2 \right]$$

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \prod_{ij}^{obj} (c_i - c_i^{'})^2$$

$$+ \lambda_3 \sum_{i=0}^{S^2} \sum_{j=0}^{B} \prod_{ij}^{obj} (c_i - c_i^{'})^2$$

(9)

Optimized compression based on YOLO detection algorithm

The addition of batch normalization not only improves the convergence of the YOLOv2 model in training and learning, but also improves the generalization performance of the detection model. The high-resolution image is used to optimize the basic network of the detection model, which improves the detection accuracy of the model. The K-means clustering algorithm is used to process the PASCALVOC dataset. By analyzing the target data in the Microsoft COCO dataset, a more effective target size ratio can be obtained and the precision of the model can be improved. The prediction method of coordinate information in the boundary coordinate system is improved. In the case of not using anchor points, the information of the bounding box can be directly predicted; with fine image features, we can fuse the image features extracted from different layers of the basic CNN network through layers; using multi-scale image training detection model To make the model more adaptable to different sizes of goals; CNN's basic network has also been redesigned.

In the CNN-based image target detection model, the CNN network model is used to extract advanced image features with rich semantic information [22]. However, in order to reduce the computational complexity of the network model and to ensure that the features of the displacement characteristics, expansion and deformation invariance are extracted, the CNN network model will continuously compress the resolution characteristics of the image, which will help to improve the image recognition of the network model. Classification ability. However, the task of image object detection includes spatial location information of the target and category information of the target. Most CNN-based image target detection models use only image features extracted from the high layers of the CNN network model. CNN can extract image features with rich semantic information, but cannot extract image features with sufficient position information at the same time. The image information extracted from the low-level CNN network model may contain sufficient spatial location information of the detected object. At the same time, the image features extracted from the lower layer of the CNN network model have higher resolution than the image features extracted from the higher layers of the CNN network model. Therefore, the image features extracted from the lower layer of CNN are more conducive to detecting relatively small targets in the

image, and improve the positioning accuracy of the detection model [23]. Due to the different features of the low-level and high-level images in the CNN network model, in order to achieve better detection accuracy, the image target detection model needs to fuse the image features extracted by different layers in the model.

Therefore, the YOLOv2 detection model uses the transport layer to fuse the image features extracted from the underlying network model CNN network with the image features extracted from the upper layer of the CNN network. Then the target detection task is predicted based on the fusion feature, which greatly improves the positioning accuracy of the target detection model and the detection accuracy of the small target [24].

2.3.   Image Target Detection Algorithm Based on Deep Learning

After determining the network structure, the framework of the model is determined. The function of this model is spatial mapping. In order to determine if the mapped feature is valid, the predicted value needs to be associated with the target value. The objective function is used to relate the model output to the actual output.

The model is assumed to be represented by a function map $f(x \mid \theta)$. Where x represents the given input and θ represents the parameters of the model. The output of the model may or may not be consistent with the actual label y. In general machine learning, the objective function is usually divided into two parts. If the objective function is J, its general expression is (10).

$$J = \frac{1}{N} \sum_{i=1}^{N} L(Y_i, f(x_i \mid \theta)) + \lambda R(\theta) \tag{10}$$

The first is the loss function and the second is the regularization or penalty. The objective function determines the optimal form of the model parameters. However, how to adjust the parameters of the model to make it better and faster to approximate the ideal mapping function requires a good parameter optimization algorithm. The optimization method can be understood as a model parameter learning algorithm. First, an overview of the basic concepts of optimization is outlined. To simplify the description, the function is represented by F(x), which is the process of adjusting the function F(x) by minimizing or maximizing the value of the argument. The form of the optimization problem can be represented by minimization, since the minimization function can be converted by taking the opposite number of maximum functions. The general minimization problem can be described by a formula:

$$x^* = \arg \min f(x) \tag{11}$$

When both the independent variable and the dependent variable are real numbers, the derivative of the function is denoted as F (x). When the derivative of a point is 0, it does not provide information about the direction of the function around it. The point where F(x) = 0 is called the critical point. The global minimum point is the point at which the function has a minimum value in all defined fields, and the necessary condition is F(x) =0. The local minimum point is the point with the smallest value in the small range

around the function. It is characterized by a function point that moves above any minimum step size is greater than it. A sufficient condition for a function to have an optimal solution is that the function has convexity.

The derivative reflects the slope of the function at point x. If the value of the argument changes slightly along the slope direction, the function value of the same size as the derivative is changed. Therefore, the nature of the derivative can well control the falling point of the function, which is the principle basis of the gradient descent. The formula for updating the independent variable along the gradient direction is:

$$\mathrm{x}' \, = \, x \, - \, \eta \Delta x f(x) \tag{12}$$

Equation (12) is the basis of the gradient descent method or the steepest descent method. It should be noted that the above function approximation method is based on only one step. In fact, there are many multi-order approximation methods based on Taylor expansion functions. In addition, the reader should note that the following optimization methods are generally based on multidimensional input for ease of reading, but this article does not distinguish the dimension of the variable.

3.   Experiments

3.1.   Data Set

Currently, DCNN-based target detection training usually begins with pre-training of the network designed on ImageNet, and then fine-tunes and tests the PASCAL VOC. ImageNet Dataset ImageNet is a dataset created by the Li Fei team at Stanford University for computer vision projects [25]. The image dataset is based on the hierarchical structure of the word network. Each level of word nodes corresponds to hundreds of pictures. Currently, the average number of images per node is 500. ImageNet is currently the largest image data acquisition system. At present, there are 141997122 pictures with 2141 semantic nodes. Each picture is hand-labeled and has a certain quality guarantee. ImageNet has 103, 4,908 images for target detection tasks, 1,000 images with SIFT features, and about 1.2 million images. With the gradual expansion of the data set, ImageNet also has a corresponding image recognition algorithm to compete. The PASCAL VOC dataset will host an annual algorithm competition from 2005 to 2012. The competition provides a standardized set of excellent data sets for image recognition, object detection and image segmentation. In 2007, the database category was fixed to 20 categories. All images in the dataset have labels for classification, identification, and detection, and only some of the data have segmentation labels. This type of dataset typically uses two versions of 2007 and 2012. There are 963 pictures of VOC 2007, including 5011 training/validation groups and 4952 test groups. A total of 24,640 objects were marked. The VOC2012 dataset differs in the detection and segmentation task tags. For inspection missions, VOC2012 includes all the images from 2008-2011. The training/verification set contains 11,540 images, marking 27,450 objects. For the split task, the training/verification set contains all the images from 2007 to 2011, with 2913 images, marking 6929 objects, and the test set is only 2008. By 2011, because the 2007 test dataset did not release tags, only images.

3.2.   Experimental Platform

The hardware platform processor used in this lab is Intel (R) Core (TM) i7-6700@3.40GHz with 16.0GB RAM memory and NVIDIA GeForce GTX1080 GPU graphics card. Among them, the graphics card can provide parallel and accelerated computing for the CNN network model on a reasonable software platform [26]. The operating system used in the experiment was Ubuntu 16.04. The software platforms used are Caffeine 2 and OpenCV 3.2. The programming language used is C++, Python.

4.   Discussion

4.1.   Analysis of Image Retrieval Results

Fig. 3 shows the target boundary results detected before and after adding a location network. The first is the original map, the second is the result of the frame obtained by the FasterR-CNN algorithm, and the third is the target frame detected after optimizing the network with target location. In the first two figures, it can be seen that the position optimization network can help the boundary to more closely surround the detection target. The third group of improved maps can be seen that the position optimization network can also completely surround the target. The initial test only revolved around the bird's body, ignoring the bird's tail. After position optimization, the boundary covers the tail of the bird, closer to the label result of the data set. It is effective to expand the search area of the original candidate frame to a certain multiple.
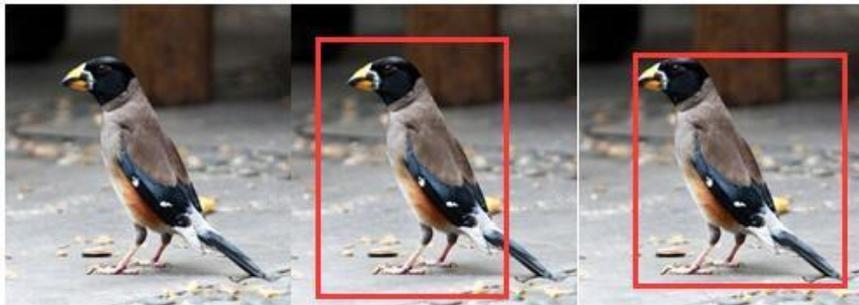
Fig. 3. Comparison of image retrieval results

4.2.   Image Detection Quantity Index Comparison

Image detection algorithms based on candidate regions, including R-CNN, Fast-RCNN, and FasterRCNN, are excellent indicators for detecting image accuracy. Another indicator of image detection is the time required to detect an image. Obviously, there is still a gap between these algorithms and real-time algorithms. It can be seen from Table 1 and figure 4 that the speed of the SSD is more advantageous in detecting the quantity quality of the picture within 1 s.

Table 1. Image detection quantity comparison

Method	Image detection
RCNN	0.03
Fast RCNN	0.49
Faster RCNN	6
YOLO	46
SSD	60
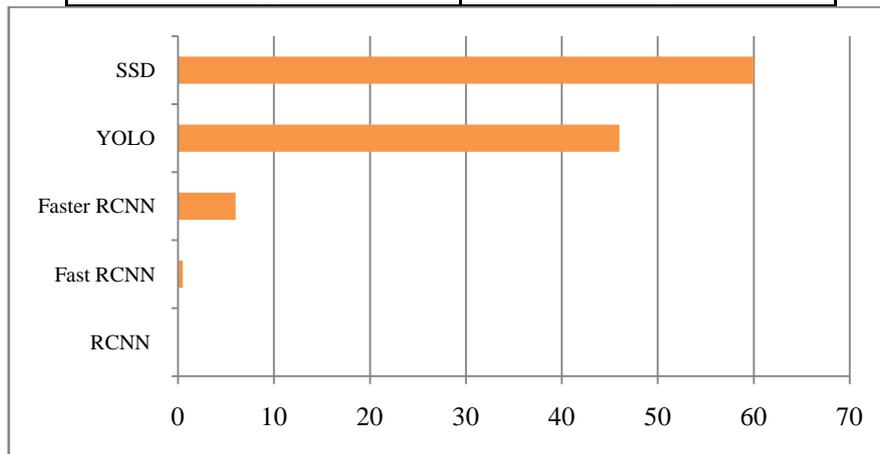
Fig. 4. Image detection quantity comparison

Image target recognition is performed using the FasterR-CNN method. Convolutional neural networks have many options. The public network can be used here as a convolutional neural network, removing the fully connected layer of the network and the pooling layer at the end of the network. The purpose of this method is to reduce the image. On a scale, each region of the image feature map has an area corresponding to the original image.

4.3.   Target Detection

The average test accuracy of all image categories from the PASCAL_VOC2012 data set can be seen from Figures 5 and 6. The MFCN algorithm proposed in this paper is superior to other target detection algorithms in mAP. The mAP of MFCN is 73.2%, which is 2.8% and 0.8% higher than Faster_R-CNN and SSD, respectively. In addition, compared with YOLO, the effective fusion of MFCN multi-features and multi-frame prediction greatly improves the detection accuracy of MFCN for small targets such as birds. The above results show that the target detection result generated by the algorithm has higher precision and can better solve the small target detection problem.
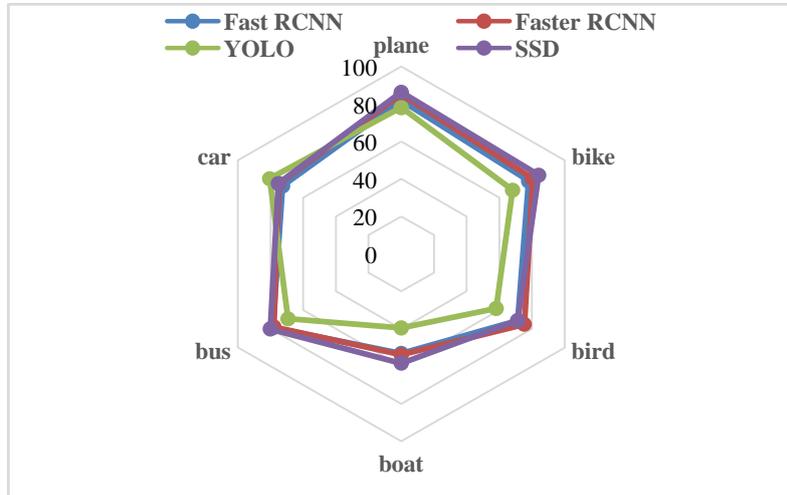
Fig. 5. Image detection results accuracy comparison 1
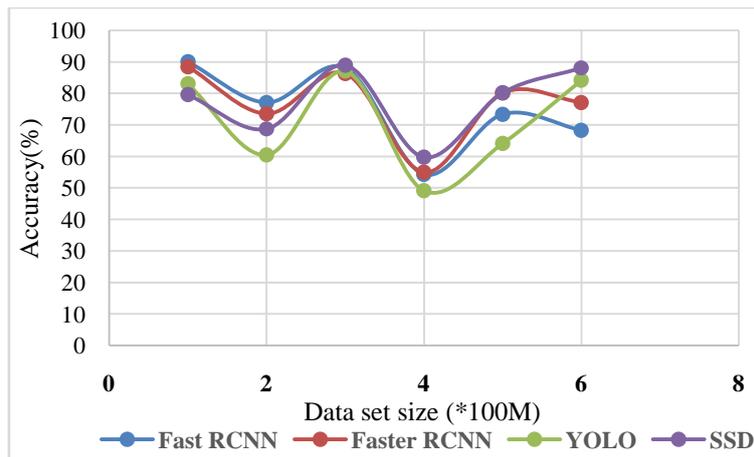
Fig. 6. Image detection results accuracy comparison 2

A target detection framework based on the regression idea of YOLO. Aiming at the problem of YOLO, a target detection method based on full convolutional network and multi-feature fusion is proposed. By establishing a convolutional neural network with no fully connected layers for target detection, since the size of the input image is not limited, the model can be used to detect multi-scale targets and can be trained using multi-scale images. The algorithm also makes full use of the feature information of different depths to obtain the feature information rich in the detected target, and improves the detection accuracy of the small-scale target.

4.4.   Comparison of Storage Costs at Different Compression Ratios

Use the method of this paper to evaluate the performance of the weight pruning method on the data set. The pruning is obtained for the ownership weight below the threshold in the network to obtain the compression model. As shown in Figure 7, the storage overhead of this method is compared with different compression ratios. At 90% compression, the maximum storage cost is reduced, and the accuracy is only reduced by 0.21%.

The performance of the convolution kernel pruning method is evaluated on the dataset using a neural network. Starting from the collection of training sets for channel selection, 10 images are randomly selected from each category in the training set to form an evaluation set. And for each input image, 10 instances are randomly sampled with different channels and different spatial locations. Therefore, there are a total of 1000 training samples used to find the best subset of channels through the greedy algorithm. Experiments have demonstrated the effectiveness of this choice (10 images per class, 10 locations per image), sufficient for neuronal importance assessment. Fine tune each layer after pruning. When all layers have been trimmed, fine-tune 10 times for greater accuracy.
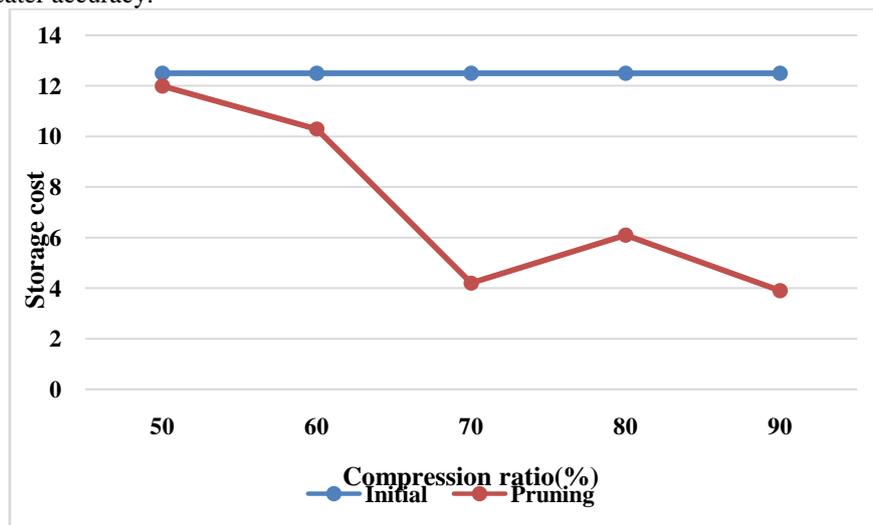
Fig. 7. Comparison of storage costs at different compression ratios

5.   Conclusions

(1) There are many redundancy in the parameters of the CNN network model. Firstly, the convolutional neural network target detection is studied. Its most widely used target detection deep learning model is constructed by the human visual system. Each layer of the convolutional neural network is described in detail. Of course, the convolutional neural network mainly includes forward propagation and back propagation. After

training and learning, the forward propagation mode will appear in the application of CNN network model, providing image feature extraction, integration processing and feature mapping for the model. The use of back propagation makes the CNN network model have the ability to optimize learning and the algorithm is compressed.

(2) The Faster-RCNN algorithm and the YOLO algorithm were discussed and discussed. Aiming at the problem that the candidate frame extracted in the Faster-RCNN algorithm is not significant, a target detection model based on the attention area recommendation network is proposed. By paying attention to the model to calculate the weight of the feature map, the saliency of the feature is enhanced and the background interference is weakened. .

(3) Construct and train deep neural networks with certain data. The deep neural network can automatically extract image features and achieve better recognition results. Convolutional neural networks are the key network for deep learning and image recognition. By constructing CNN, image features can be extracted layer by layer. The comparison test results of several algorithms show that the number of parameters of the model is greatly reduced, and the feature representation ability of the model within the acceptable range is reduced.

References

1. Zafar, Bushra; Ashraf, Rehan; Ali, Nouman; Ahmed, Mudassar; Jabbar, Sohail; Naseer, Kashif; Ahmad, Awais; Jeon, Gwanggil. Intelligent Image Classification-Based on Spatial Weighted Histograms of Concentric Circles. Computer Science and Information Systems. 15(3). 615-633. (2018).
2. Tung, Frederick, & Mori, Greg. Deep Neural Network Compression by In-Parallel Pruning-Quantization,IEEE Transactions on Pattern Analysis and Machine Intelligence,PP (99), pp. 1-1. (2018)
3. Díaz, G., Macià, H., Valero, V. et al. An Intelligent Transportation System to control air pollution and road traffic in cities integrating CEP and Colored Petri Nets. Neural Comput & Applic 32, 405–426 (2020).
4. Fan, D., Wei, Lu, & Cao, Maoyong. Extraction of Target Region in Lung Immunohistochemical Image Based on Artificial Neural Network,Multimedia Tools & Applications,75(19), pp.1-18. (2016)
5. Y Zhao, H Li, S Wan, A Sekuboyina, X Hu, G Tetteh, M Piraud, B Menze. Knowledge-aided convolutional neural network for small organ segmentation. IEEE journal of biomedical and health informatics, 23(4), pp. 1363-1373, (2019)
6. Ding, X., & Yang, Hong Hong. A Study on the Image Classification Techniques Based on Wavelet Artificial Neural Network Algorithm,Applied Mechanics & Materials,602-605, pp.3512-3514.
7. Lin, G., Wang, J., Yong, F., & Chen, N. Robust Visual Tracking Based on Convolutional Neural Networks and Conformal Predictor,Acta Optica Sinica,37(8), pp. 815003. (2017)
8. Wang, Li, Tang, Jun, & Liao, Qingmin. A Study on Radar Target Detection Based on Deep Neural Networks,IEEE Sensors Letters,3(3), pp.1-4. (2019)
9. Liu, P., Guo, J. M., Wu, C. Y., & Cai, D. Fusion of Deep Learning and Compressed Domain Features for Content-Based Image Retrieval,IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society,PP(99), pp.1-1. (2017)

10. Arun, K. S., & Govindan, V. K. A Hybrid Deep Learning Architecture for Latent Topic-Based Image Retrieval,Data Science & Engineering, 3(2), pp. 166-195. (2018)

11. Yang, Dongwen, Du, Lan, Liu, Hongwei, Wang, Yan, & Gu, Mingfei. Extended Geometrical Perturbation Based Detectors for Polsar Image Target Detection in Heterogeneously Patched Regions,IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,12(1), pp.1-17. (2019)

12. Wang, G., Yao, Y., Chen, Z., & Hu, P. Thermodynamic and optical analyses of a hybrid solar CPV/T system with high solar concentrating uniformity based on spectral beam splitting technology. Energy, 166, 256-266. (2019)

13. Yang, S., & Shi, Zhenwei.Hyperspectral Image Target Detection Improvement Based on Total Variation,IEEE Transactions on Image Processing,25(5), pp.2249-2258. (2016)

14. Liu, Hui; Li, Chenming; Xu, Lizhong. Dimension Reduction and Classification of Hyperspectral Images based on Neural Network Sensitivity Analysis and Multi-instance Learning. Computer Science and Information Systems. 16(2). 443-467.(2019)

15. Zhao, C., Jing, X.,& Li, W. Hyperspectral Image Target Detection Algorithm Based on Stomp Sparse Representation, Harbin Gongcheng Daxue Xuebao/Journal of Harbin Engineering University, 36(7), pp.992-996.(2015)

16. Ali, Munwar, Low Tang Jung, Abdel-Haleem Abdel-Aty, Mustapha Y. Abubakar, Mohamed Elhoseny, and Irfan Ali. Semantic-k-NN Algorithm: An Enhanced Version of Traditional k-NN Algorithm.Expert Systems with Applications: 113374.(2020)

17. Chi-Hua Chen, Fangying Song*, Feng-Jang Hwang, Ling Wu, "A Probability Density Function Generator Based on Neural Networks," Physica A: Statistical Mechanics and its Applications, 541, Article ID 123344, March.(2020)

18. Lakshmanaprabu SK, Mohamed Elhoseny, Shankar Kathiresan, Optimal Tuning of Decentralized Fractional Order PID Controllers for TITO Process using Equivalent Transfer Function, Cognitive Systems Research, Volume 58, December, pp. 292-303.(2019)

19. Zheng Xu, Lin Mei, Zhihan Lv, Chuanping Hu, Xiangfeng Luo, Hui Zhang, Yunhuai Liu。 Multi-Modal Description of Public Safety Events Using Surveillance and Social Media. IEEE Trans. Big Data 5(4): 529-539 (2019)

20. Xiong, Q., Zhang, X., Wang, W., & Gu, Y. A Parallel Algorithm Framework for Feature Extraction of EEG Signals on MPI. Computational and Mathematical Methods in Medicine, 2020, 1-10. (2020)

21. Junlong Chen, Xiaomeng Wang & Zhaopeng Chu Capacity Sharing, Product Differentiation and Welfare, Economic Research-Ekonomska Istraživanja, 33:1, 107-123.(2020)

22. Xu Z, Cheng C, Sugumaran V. Big data analytics of crime prevention and control based on image processing upon cloud computing. J Surveill Secur Saf 2020;1:16-33.

23. Ling Wu, Chi-Hua Chen*, Qishan Zhang, "A Mobile Positioning Method Based on Deep Learning Techniques," Electronics, 8(1), Article ID 59, January.(2019)

24. Mu Zhou, Yanmeng Wang, Yiyao Liu, and Zengshan Tian. An Information-theoretic View of WLAN Localization Error Bound in GPS-denied Environment. IEEE Transactions on Vehicular Technology. 68(4): 4089-4093.(2019)

25. de Souza, L.A., Afonso, L.C.S., Ebigbo, A. et al. Learning visual representations with optimum-path forest and its applications to Barrett's esophagus and adenocarcinoma diagnosis. Neural Comput & Applic 32, 759–775 (2020).

26. Wei, P., He, F. & Zou, Y. Content semantic image analysis and storage method based on intelligent computing of machine learning annotation. Neural Comput & Applic 32, 1813–1822 (2020).

Yan Sun received the B.S. degree in electrical information engineering in 2004, and the M.S. degrees in communication and information system, and the Ph.D. degrees in signal and information processing from Harbin Engineering University (HEU), China, in 2012.

His research interests include wireless communications, image processing and machine learning. E-mail: aidenby@163.com

Zheping Yan was born in Longyou, Zhejiang, China, in 1972. He received the B.S. degree in nuclear power equipment, the M.S. degree in special auxiliary device and system in naval architecture and ocean engineering, and the Ph.D. degree in control theory and control engineering from Harbin Engineering University, China, in 1994, 1997, and 2001, respectively. He was the Post-Doctoral Researcher in mechatronic engineering with the Harbin Institute of Technology, China, in 2004.His current research interests include identification of non-linear system, and multi-sensors data fusion and intelligent control. E-mail: feiyunsy1213@163.com

Collaborative Filtering Recommendation Algorithm in Cloud Computing Environment

Pei Tian[1]

Human Resource Center, Xi'an Peihua University,
Xi'an, 710125, Shaanxi, China
tianpei@peihua.edu.cn

Abstract. With the advent of the era of cloud computing, the amount of application data increases dramatically, and personalized recommendation technology becomes more and more important. This paper mainly studies the collaborative filtering detection algorithm in the cloud computing environment. The algorithm migrates the collaborative filtering detection technology and applies it to the cloud computing environment. It shortens the recommendation time by using the advantages of clustering. A new recommendation algorithm can improve the accuracy of recommendation, and proposes a parallel collaborative filtering recommendation algorithm based on project. The algorithm is designed with programming model The experimental results show that the proposed algorithm has shorter running time and better scalability than the existing parallel algorithm.

Keywords: collaborative filtering, Hadoop, HDFS, Map Reduce, person correlation coefficient

1.   Introduction

The rapid growth of Internet scale causes the problem of information overload [1]. It is difficult for users to obtain valuable information from massive data. As an important filtering method, personalized recommendation actively recommends articles of interest to users by analyzing their interests and historical behaviors. In order to solve the problem of Internet information interest rate overload [2], at present, different recommendation methods can be divided into There are three kinds of recommendation: content-based recommendation, collaborative filtering recommendation and hybrid recommendation [3, 4]. Recommendation based on collaborative filtering is undoubtedly the most successful personalized recommendation technology [4].

David Goldberg and others proposed a user collaborative filtering algorithm in [5]. The number of projects is far greater than the number of users, but for online e-commerce sites, the number of projects is often lower than the number of users. Location-based collaborative filtering recommendation can not only shorten the calculation time of recommendation, but also produce a variety of recommendations. Greg Linden et al. Proposed a Item-based Collaborative Filtering Recommendation Algorithm in [6]. By calculating the similarity between Items, we can find the similar

---

[1] Corresponding author

neighbors of Items and recommend them to active users. Because the number of items purchased by users only accounts for a small part of the total number of items, due to the sparsity of data, the recommendation neighborhood of collaborative filtering based on memory is not accurate. Sandvig j et al. Proposed a collaborative filtering recommendation algorithm based on Association Rule Mining in [7]. In the case of noise, the recommended accuracy shows better stability and robustness [8]. In order to improve the real-time performance of the recommendation algorithm, document [9] implements the Item-based Collaborative Filtering Recommendation Algorithm in parallel on the Hadoop platform. However, in the algorithm designed in reference [9], the calculation of all users' purchase item pairs is to find all users who purchase the item pair by searching all purchase records of the two items. The time complexity of this method is $O (m^2 * n^2)$ (M represents the number of users, n represents the number of items). At the same time, the algorithm uses two Map Reduce processes to calculate the prediction score of the user's non purchased items by regarding the user as the center: sending the items purchased by the same user to the same node, reading the similarity matrix into memory and the similarity of similar items into each item of the user when calculating the prediction, which not only reads too much redundant data , and in the massive data environment, the number of users and goods are very large, the scale of similarity matrix will lead to it can not read all, will affect the scalability of parallel collaborative filtering algorithm [9].

Based on the above problems, this paper proposes a Parallel Collaborative Filtering Recommendation Algorithm based on Item (Parallel-Item-Base Collaborative Filtering, PIB-CF), which will use Map Reduce framework and HDFS to implement the Item-based Collaborative Filtering Algorithm in Hadoop cluster. In view of the problem of algorithm design in reference [10], PIB-CF algorithm calculates that the user purchases the item pair together by sending any two item pairs as keys to the reduce node, and the items with the same item pair will be sent to the same node. The time complexity of this method is $O (n^2 * m^2)$, which will greatly reduce the search time. And this paper designs an efficient parallel strategy, designs a Map Reduce process to achieve the prediction and scoring function of two Map Reduce processes in reference [10]. In this way, the calculation results stay in memory, thus reducing the traffic, improving the efficiency of the algorithm, and achieving good real-time and scalability. Experimental results show that PIB-CF algorithm is effective.

In the first part of this paper, collaborative filtering recommendation technology and Hadoop cloud computing platform are introduced. In the second part, Project-based Collaborative filtering recommendation algorithm is proposed. In the third part, experiments are carried out. In the fourth part, the algorithm proposed in this paper is summarized.

2.   Related Research

2.1.   Collaborative Filtering Recommendation Technology

The idea of collaborative filtering technology is simple and easy to understand, and it is recommended for individuals from the perspective of groups [11]. The basic idea is that

if users had the same preferences in the past, they would have more similar preferences in the future. For example, if there are users a and B, their purchase experience is very similar. When user a recently obtains a book, but user B does not, then the recommendation idea is to recommend this book to B. Recommend books that you may like, which requires filtering out the most likely books from a large number of collections. Users are collaborating with others implicitly, so this technology is also called CF, collaborative filtering. Collaborative filtering algorithms can be divided into three categories: memory based recommendation algorithm, model based recommendation algorithm and hybrid recommendation algorithm, as shown in Figure 1.
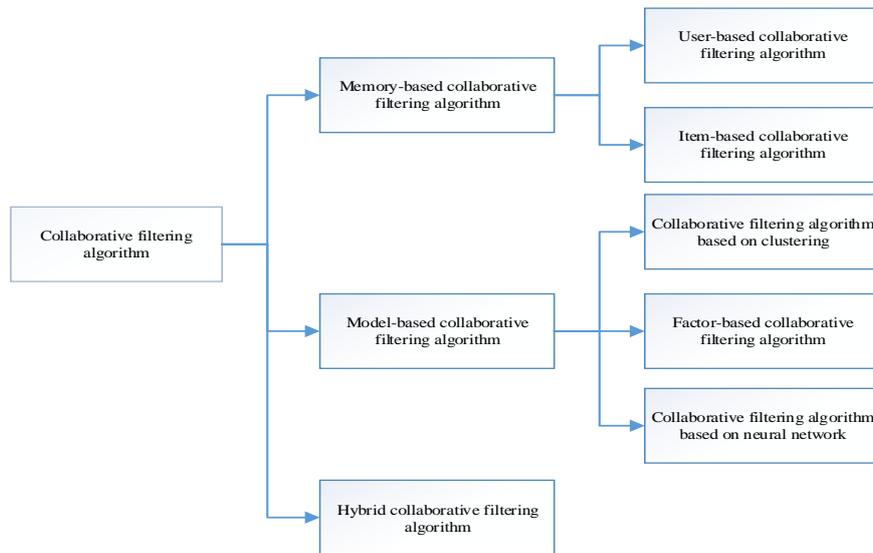
Fig. 1. Classification of collaborative filtering algorithm

Memory based collaborative filtering algorithm

Memory based collaborative filtering algorithm is a successful and practical recommendation algorithm. It is recommended by collaborative filtering through the saved historical information in the system. It analyzes and calculates the relevant historical information of the Item and users, generates a similar set of users or Item neighbors, and calculates and generates recommendations according to the scoring situation. In this process, there are two algorithms: user based recommendation algorithm (User-CF) and item based recommendation algorithm (Item-CF). The two algorithms have the same calculation flow, as shown in Figure 2:
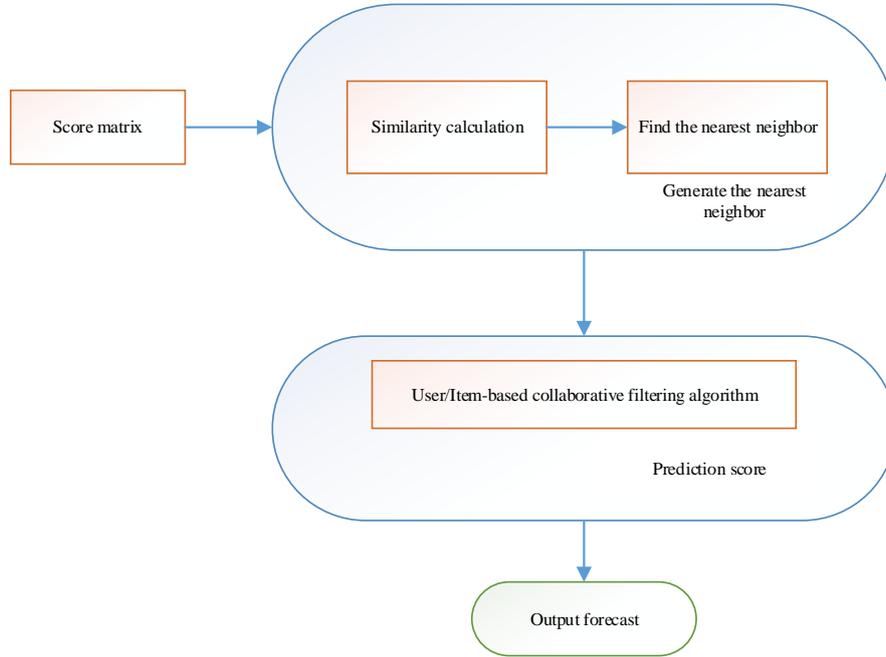
Fig. 2. Memory based collaborative filtering algorithm

User based recommendation algorithm

In real life, a person with needs often consults other people who have the same needs, and makes personal decisions by referring to other people's opinions and methods. In the same way, when a group of people with the same interests gather together and recommend their own things, it is likely that this is what other people who do not have this item need. In fact, the user-based recommendation algorithm follows this idea: if there are some users, their scoring items are basically the same, which indicates that they have the same interest preference, then their unique scoring items can be used as other users' Item scoring reference. The collaborative filtering algorithm based on users is from the perspective of users, which refers to users' ratings of items similar to the interests of the target users, to calculate the user's rating of the non rated item, namely Formula 1.

$$\hat{r} = \bar{r} + \frac{\sum_{v \in S\,(u,K) \cap N(i)} W_{uv}(r_{vi} - \bar{r_v})}{\sum_{v \in S\,(u,K) \cap N(i)} |W_{uv}|} \tag{1}$$

Where, $S\,(u, K)$ is the set of K users whose interests are most similar to those of u, and K needs to be defined according to the data set. $N\,(i)$ is a collection of users who overestimate Item I. r_{vi} is the rating of user V on Item i. $\bar{r_v}$ is the mean value of all items of user v. W_{uv} represents the similarity between user u and user v. generally.Generally, we choose one of Euclidean distance, cosine similarity measure and Pearson correlation coefficient as similarity calculation method.

Item based recommendation algorithm

Item-based recommendation is based on the degree of association between projects, which is different from user recommendation based on the similarity between users. For example: when a user likes item a, and others like item a also like Item B, then recommend Item B to the past. Simply put, "what Item is similar to what user a likes". Figure 3 shows the difference between Item-based and user-based recommendation algorithms: find similar users and know what they prefer based on user's recommendation (solid line); know user's preference based on Item recommendation (dotted line), and find similar Items.
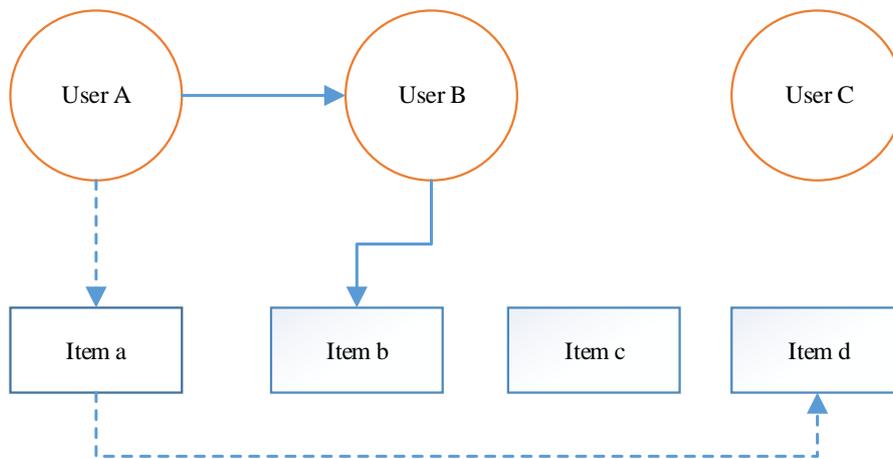
Fig. 3. Difference between item based and user based

The Item-based recommendation algorithm is to recommend from the perspective of the item. Like user based recommendation algorithm, one of Euclidean distance, cosine similarity and Pearson coefficient is used as the similarity parameter between items. Then calculate the score of the user who has not scored the item, that is, formula 2.

$$\hat{r}_b = \bar{r}_b + \frac{\sum_{a \in I \ (a,K) \cap M(i)} W_{ba} \ (r_{ai} - \overline{r_a})}{\sum_{a \in I \ (a,K) \cap M(i)} |W_{ba}|} \tag{2}$$

This formula is the score of the User I of forecast item b. $I(a, K)$ is the neighbor set of Item a, which has k neighbors. $M(i)$ is a collection of items that are overrated by user I. r_{ai} is the rating of user a on Item I. $\overline{r_a}$ is the mean value of all user scores of item a. W_{ba} represents the similarity parameter between items b and a.

2.2. Similarity Algorithm Research

For similarity algorithm, there are Euclidean distance, cosine similarity measurement, Pearson correlation coefficient similarity algorithm and so on. At present, Pearson coefficient similarity algorithm is widely recognized and applied. Next, we will introduce each algorithm and analyze its advantages and disadvantages.

European distance

This similarity method is a kind of calculation method which is widely used in many fields. It mainly uses users or projects as points on the plane to calculate the distance between each point. The calculation method is shown in Formula 3.

$$sim(a,b) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \cdots + (a_m - b_m)^2} \quad (3)$$

Where, $a, b \in U$ and a_i represent user a's rating for item i. The value range of sim(a, b) varies greatly and may be greater than 1, which is not conducive to the calculation of prediction score. Therefore, the sweight value calculated by formula 4 will be used as the weight value in the prediction calculation, so that the weight value is within [0,1].

$$sweight = \frac{1}{1+sim} \quad (4)$$

Through Euclidean distance, although we can get the similar neighbor set, it has no too high credibility as the weight of the prediction calculation, which will seriously affect the accuracy of the prediction score.

Cosine similarity

The measure of similarity is to calculate the cosine of the angle between two n-dimensional vectors. When the angle between vectors is smaller, the cosine value will be closer to 1, and the directions of the two vectors are more similar. For example, if there are two users a and B, their corresponding scoring vectors are $\vec{a} = \{a_1, a_2, \ldots \ldots\}$ and $\vec{b} = \{b_1, b_2, \ldots\}$, the similarity can be defined as formula 5.

$$sim(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| * |\vec{b}|} \quad (5)$$

The sign · represents the dot product between vectors, and $|\vec{a}|$ represents the Euclidean length of vectors, that is, the square root of the dot product of vectors themselves. The value of cosine similarity is between 0 and 1. The closer one is, the more similar it is. The accuracy of the prediction score is relatively high, but because the method does not take into account the difference between the average user score, there are still some defects. For example, in the actual score, the five values of {1, 2, 3, 4, 5} are taken as the substitute value of {worst, poor, general, good, best}. Some people are used to playing {1, 2, 3} such {bad, general, good} score, which will have a certain impact on the score prediction calculation.

Pearson correlation coefficient [15]

Pearson correlation coefficient can be used as a parameter to measure the linear correlation between two variables, such as formula 6.

$$W_{a,b} = \frac{\sum_{i \in P}(r_{a,i} - \overline{r_a}) \cdot (r_{b,i} - \overline{r_b})}{\sqrt{\sum_{i \in P}(r_{a,i} - \overline{r_a})^2 \cdot \sum_{i \in P}(r_{b,i} - \overline{r_b})^2}} \quad (6)$$

Where the $\overline{r_a}$ symbol represents the average score of user a. The range of Pearson correlation coefficient is [-1,1]. When its absolute value tends to 1, it shows that neighbors are more similar. For this method, it takes into account the fact that the user

rating standards are not the same. It can find obvious user relationships, and then get better and accurate similarity.

Compared with the other two methods, Pearson correlation coefficient is more consistent with the recommended algorithm, and it is also the most commonly used in collaborative filtering.

2.3.   Hadoop Cloud Computing Platform

Cloud computing is a pay as you go model (IT resources include networks, servers, storage, applications and services) that provide easy access to a customizable pool of IT resources through the Internet. These resources can be deployed quickly and require little management or interaction with service providers. Cloud computing centralizes the distributed computing, storage, service components and network software resources on the network. Based on resource virtualization, cloud computing provides convenient and fast services for users. It can realize the distributed and parallel processing of computing and storage. If "cloud" is regarded as a virtualized storage and computing resource pool, then cloud computing is the data storage and network computing services provided by this resource pool for users based on network platform. The Internet is the largest "cloud", on which a variety of computer resources constitute a number of huge data centers and computing centers.

The framework of Hadoop [12] includes HDFS, Map Reduce, H Base and other open-source components. At the same time, Hadoop also includes several independent subsystems, such as: hive, Chukwa, Avro, etc. So many components and subsystems make up a powerful cloud computing platform [13], Hadoop, as shown in Figure 4.
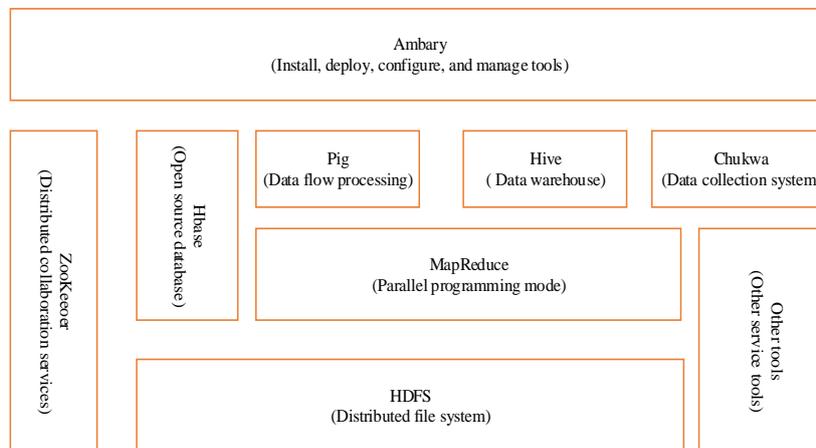
Fig. 4. Hadoop platform structure

HDFS distributed file system

HDFS is the core sub Item of Hadoop Item. HDFS is born out of Google File System (GFS), which reduces the demand for hardware. It only needs cheap hardware facilities,

and can provide a distributed file system that can store and process massive data through network link [14]. HDFS is built with master / slave structure, as shown in Figure 5.
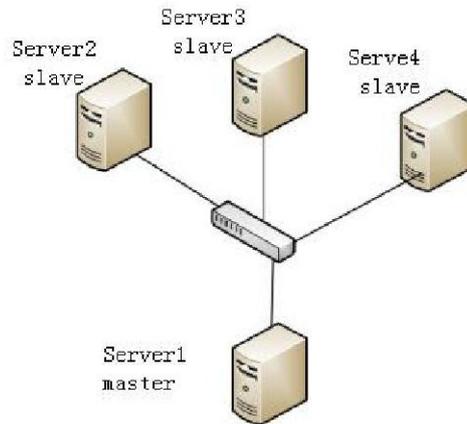
Fig. 5. Master-slave structure

For general machine deployment [16], there is only one primary node named node on the master, and each slave, there is a slave Data node on. A complete HDFS cluster consists of a name node and several data nodes. Name node is a main control node, which is responsible for the namespace of HDFS, and also manages the file tree of HDFS, metadata of all files and directories, which acts as an administrator of HDFS system. Data node is the node of data storage in file system [17]. For the storage of large-scale data files, HDFS will first split the whole data file to get multiple data blocks, and then complete redundant backup storage in multiple Data nodes. HDFS clients can complete the read and write operations of data files in the system by calling the API provided by the system. First, the client accesses the node Name node, reads all the metadata information saved, finds the corresponding Data node, and reads and writes the data. The architecture is shown in Figure 6.
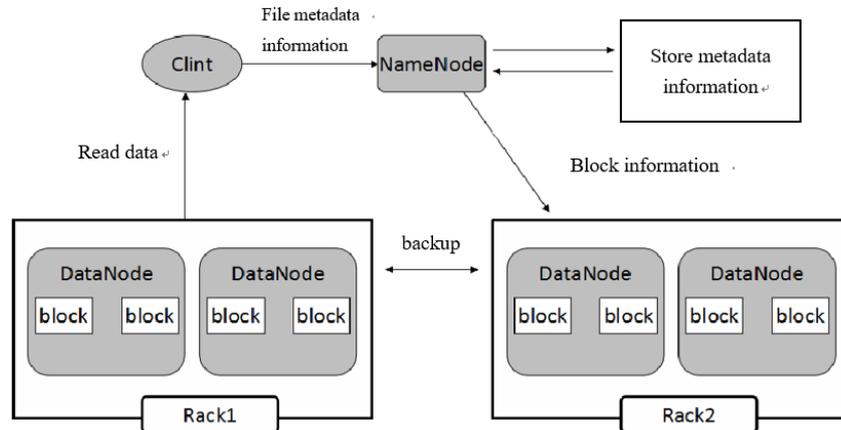
Fig. 6. Structure of HDFS

In addition, HDFS has several other functions in data security management:

(1) Data backup mode: file block is the storage form of HDFS data. Generally, there are three backups for a file block, which can greatly improve the safety factor of data;

(2) Heartbeat detection: detect Data Node frequently and in real time. If there is a problem with Data Node, use data backup to ensure the integrity of data;

(3) Data verification: CRC32 method is used for data verification. For file blocks, not only the data but also the corresponding verification information will be written; when reading file blocks, the verification will be performed first and then read.

(4) Security mode: the system in security mode will be restricted in permissions. Users are not allowed to modify and delete the contents of the system until the end of security mode. This is to detect the data on each Data Node and determine the validity of the data when the system starts.

Map reduce programming model

MapReduce is a cluster based high performance parallel computing platform (cluster infrastructure). It allows the use of common commercial servers on the market to form a distributed and parallel computing cluster with dozens, hundreds to thousands of nodes. It provides a huge but well-designed parallel computing software framework, which can automatically complete the parallel processing of computing tasks, automatically divide computing data and computing tasks, automatically allocate and execute tasks on cluster nodes and collect calculation results, and hand over many complex details of the underlying system involved in parallel computing such as data storage, data communication, fault-tolerant processing and so on The system is responsible for processing, which greatly reduces the burden of software developers.

In Hadoop, there are two carriers for executing Map Reduce tasks: Job tracker and Task tracker. Job tracker is responsible for the supervision and control of Map Reduce. According to the requirements of Map Reduce, the corresponding resources are scheduled. In a Hadoop cluster, you cannot have two Job tracker tasks at the same time.

Task tracker is responsible for running Map Reduce and responding to the request of Job tracker. Here, Map Reduce is divided into two parts: Map task and reduce task. As shown in Figure 7, the control flow of Map Reduce is as follows: Job tracker schedules tasks to Task tracker, and when Task tracker performs tasks, it will return progress reports. Job tracker will record the progress status. If a task on a Task tracker fails to execute, Job tracker will assign the task to another Task tracker until the task ends.
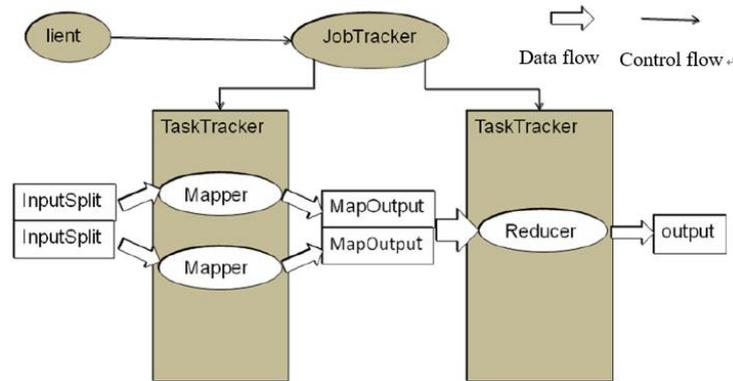
Fig. 7. Simple working principle of Map Reduce

The data processing flow of Map Reduce is as follows: the data is first processed into multiple Input Splits, and then input into the corresponding number of maps. A map program will read the data at the specified position of Input Split, then process the data in the set way, and finally write it to the local disk. Reduce will read the map output data, merge the data, and then output them to HDFS.

In Hadoop, each Map Reduce task is initialized as a job. Each job can be divided into two phases: the mapping phase and the restore phase. These two stages are represented by two functions, the Map function and the Reduce function. Map function receives input in the form of $<$ key, value $>$ and produces the same intermediate output in the form of $<$ key, value $>$. Hadoop is responsible for passing all values with the same intermediate key value to the reduce function. The reduce function receives the input in the form of $<$ key, (list of values) $>$ and then processes the value set and outputs the result. The output of reduce is also in the form of $<$ key, value $>$. In order to display conveniently, mark three $<$ key, value $>$ as $< k_1, v_1>$, $< k_2, v_2>$, $< k_3, v_3 >$, as shown in Figure 8:
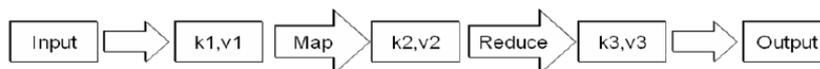
Fig. 8. Map Reduce programming model

3.   Design and implementation of project based parallel collaborative filtering recommendation algorithm

3.1.   Item based collaborative filtering recommendation algorithm

Item based collaborative filtering refers to matching the products purchased or rated by a user to similar products, and then entering similar products into the recommendation list. From the point of view of calculation, it is to take the evaluation of all users on the specified purchased goods as a vector, and each user's evaluation on the goods as the corresponding component in the vector. In this way, the similarity between items is calculated by the vector value, and the similar items of items are obtained. Then, the current user's evaluation on the non purchased items is predicted according to the user's historical shopping records, and a List of items sorted by forecast as recommendations. The algorithm can be divided into four parts:

   1. Score data preprocessing, calculate the average score of the Item.
   2. Through a certain similarity measurement method, the similarity between items is calculated.
   3. Through a certain forecast scoring model, the forecast scoring of user's UN scored items is calculated.
   4. Time performance measurement and precision measurement of the algorithm.

3.2.   Parallelization of Collaborative Filtering Recommendation Algorithm Based On Item

The most time-consuming stage of collaborative filtering recommendation algorithm is to calculate similarity matrix and predict user preferences. If the number of users is m, the time complexity of similarity calculation is $O(n^2m)$, and the time complexity of prediction is $O(nm)$. Obviously, when calculating the similarity matrix, the similarity of one item pair is independent of that of another pair. So the similarity matrix can be calculated in parallel. Secondly, when predicting user preferences, computing preferences for different users are also independent of each other, so predicting user preferences can also be done in parallel. However, the calculation of similarity must be completed before the prediction of preference, so the two must be executed serially.

   Map Reduce is a popular parallel programming model at present. Users only need to specify the calculation process of Map and function Reduce, and the system will automatically calculate in parallel on a large-scale cluster. And both implement Map Reduce.

   Using Map Reduce model to implement PIB-CF, a parallel algorithm based on collaborative filtering recommendation, requires three map and reduce processes. Map1 calculates each user's purchase item and its score. Reduce1 calculates the mean value of item ID scores. Map2 is used to calculate all pairs of items in the same user's purchase record, and reduce2 is used to calculate the similarity between items. Map3 is used to calculate all other items similar to an item, and reduce3 is used to calculate the user's forecast score for items not scored. Map1 and reduce1 calculate the Item and its mean

value and save them in the file av-file. Because the calculation process is relatively simple, we will not go into details here.

The user file of user score data read by Map2 function is stored in. Each line of data represents a user's historical purchase item record. It is read into the map node line by line in the form of key value pair < key, value >. Each key value pair represents a data record. key is the offset of the current record relative to the starting point of the input data file, and value is the item ID and its score in the current record. Map2 < key, value > calculates all pairwise item pairs purchased by the current user and their corresponding scores. The pseudo code of the function is as follows:

```
Input: key, the value corresponding to the key;
Output: key, the value corresponding to the key;
The pseudo code flow is as follows:
Map2(key,value){
1. GenItemsandRatings(value ,item,rating,len);
2. For i=0 to len - 1 do
3.   For j=i+1 to len do {
4.       item_a=Items[i]; rating_a=Ratings[i];
5.         item_b=Items[j]; rating_b=Ratings[j];
6.         key'=CombineItems(item_a,item_b);
7.       Value'=CombineItems(rating_a,rating_b);
8.         Output<key,value>;
9.   }
10.}
```

After Map2 is output, the system will send item pairs with the same key to the same Reduce node. The function of Reduce (key, value) is to calculate the similarity between two corresponding items. The pseudo code of the function is as follows:

```
Input: item label key, item corresponding score value;
Output: similarity of item to key '
The pseudo code flow is as follows:
Reduce2(key,value){
1. AssignValue(value,rating_i,rating_j);
2. ReadFileAveRat(av-file,items,aveRatings);
3. for each val in value{
4.   ReadRatings(val,rating_i[k],rating_j[k]);
5.       k++;
6.       }
7.       value'=Pearson();
8. key'=key;
9.   Output<key,value>
10.}
```

After the implementation of Reduce2, the system will get the Item similarity matrix, which will predict and grade the items that are not evaluated by each user.

The output results of Reduce2 take the form of < key, value > as the input of map3. Key is the Item pair, and value is the similarity of the Item pair. The function of Map3 (key, value) is to send all Item pairs containing a Item to the same reduce node. The pseudo code of the function is as follows:

```
Input: item code key;
Output: Code of similar item corresponding to item key
'and its similarity list value';
Map3(key,value){
1. ReadItem(key,item1,item2);
2. key'=item1;
3. value'=Combine(item2,value);
4. Output<key',value'>;
5. Key'=item2;
6. Value'=Combine(item1,value);
7. Output<key,value>
}
```

After Map3 is executed, the system will send the items similar to the items and the similarity between them to the same Reduce node. Reduce3 (key, value) calculates the predicted score of the user's non scored items. The pseudo code of the function is as follows:

```
Input: item code key, similar item list value, user
rating file av-file, item average rating file av-file
Output: Item forecast score;
The pseudo code flow is as follows:
Reduce3(key,value){
1. ReadFileAveRat(av-file,allItems,aveRating);
2. ReadFileRat(user-file,userid,items,rating);
3. for each val in value{
4.    ReadItemSimi(val,item[k],simi[k]);
5.    k++;
6. }
7. if array items no item key {
8.    value'=Prediction();
9.    key'=Combine(userid,key);
10. }
11. Output<Key',value'>
12.}
```

4.   Experimentand analysis

4.1.   Experimental platform and data set

In this paper, we use the dataset provided by the Movielens website (HTTP / /: www.grouplens. org / node / 73 /), which contains 71567 users and their evaluation of films. Each user evaluates at least 20 films with a rating of 1-5. The sparse degree of movie score in this dataset is 0.987. In this experiment, 21 (8-core) CPU servers constitute the parallel computing environment, and Hadoop is used to implement the parallel collaborative filtering recommendation algorithm proposed in this paper.

User rating data in the recommendation system includes user set $U = \{u_1, u_2, \ldots, u_m\}$ and item set $I = \{i_1, i_2, i_n\}$, the scoring matrix can be expressed as a m * n-order user scoring matrix $M = (r_{ij})_{m \times n}$, as shown in Table 1

Table 1. user Item rating matrix M

	I_1	$\ldots$	I_j	$\ldots$	I_n
U_1	$r_{1,1}$	$\ldots$	$r_{1,j}$		$r_{1,n}$
$\ldots$					
U_i	$r_{i,1}$	$\ldots$	$r_{i,j}$		$r_{i,n}$
$\ldots$					
U_m	$r_{m,1}$	$\ldots$	$r_{m,j}$		$r_{m,n}$

Each element of the matrix represents the user's rating of the Item, and the rating reflects the user's preference for the Item. The score can be used to indicate whether you like it or not, or the number can be used to indicate how much you like it.

4.2.   Measurement method

If u is used to represent the user set of item and common scoring, then the scoring similarity sim (i, j) of item i and j can be calculated by person correlation coefficient [11], and the formula is as follows:

$$sim(i,j) = \frac{\sum_{u \in U}(R_{u,i} - \overline{R_i})(R_{u,j} - R_j)}{\sqrt{\sum_{u \in U}(R_{u,i} - \overline{R_i})^2}\sqrt{\sum_{u \in U}(R_{u,j} - \overline{R_j})^2}} \qquad (7)$$

Where r_{ui} and r_{uj} represent the user's u-scores for items i and j, respectively, $\overline{R_i}$, $\overline{R_j}$ represents the average scores for items i and j.

In order to make the prediction more accurate, this paper uses an optimized prediction scoring strategy: take the average score of the items that the target user needs to score as the benchmark value, then find the neighbor set of the target items of the target user, and then use the item similarity in the neighbor set as the weight value to calculate the prediction score of the target items, and predict the calculation of the evaluation value of the user u for the item I The following expressions are as follows:

$$P_{u,i} = \overline{R_i} + \frac{\sum_{j \in S(i)} sim(i,j) * (R_{u,j} - \overline{R_j})}{\sum_{j \in S(i)}|sim(i,j)|} \qquad (8)$$

Where r_{ui} represents the user U's score for item I, $\overline{R_j}$ represents the average score of item j, and sim (i, j) represents the similarity between item i and item j.

4.3.   Experimental results and analysis

In order to test the scalability of the system, two experiments are designed. In the first group of experiments, the number of computing nodes was set, and the size of test data set was variable. The second experiment set the size of data set, and the number of computing nodes was variable. Because the process of calculating similarity in collaborative filtering recommendation based on Item can be calculated offline, only the predicted user score needs to be calculated in real time, so the test time of this experiment is the time of predicting Item score. The experimental results are shown in Figures 9 and10.
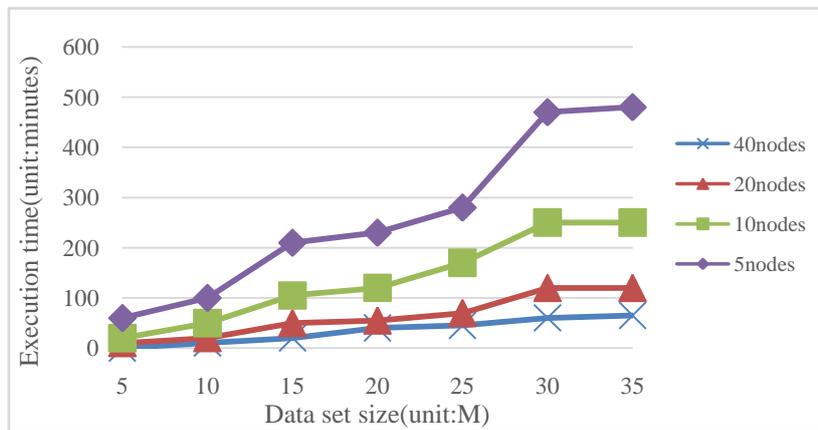
Fig. 9. Data volume and execution time
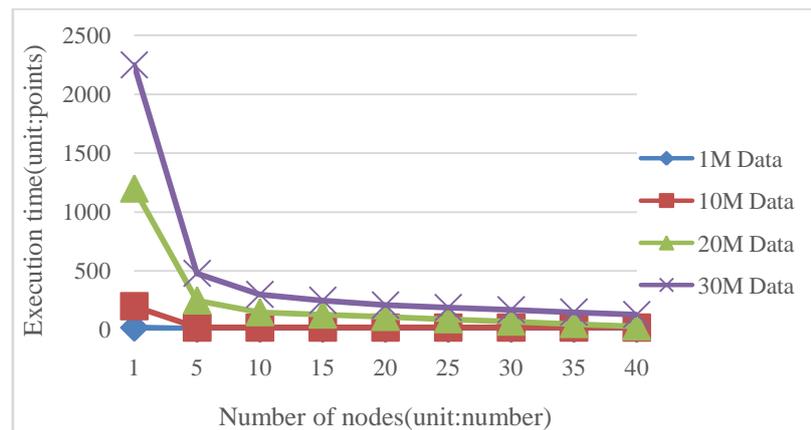
Fig. 10. Relationship between number of physical nodes and execution time

It can be seen from the figure and figure that the online prediction time can be greatly reduced after the collaborative filtering recommendation algorithm based on the Item is implemented in parallel, and good scalability can be achieved by increasing the number of computing nodes with the increasing amount of data.

In order to compare the time performance of the algorithm with the algorithm proposed in the literature, we designed a set of experiments. The experimental data set is 10 m, running on 1, 5, 10, 15, 20, 25, 30 nodes respectively. The comparison of algorithm execution time is shown in Figure 11, and the experimental results show that the performance of this algorithm is better than that of the reference [8].
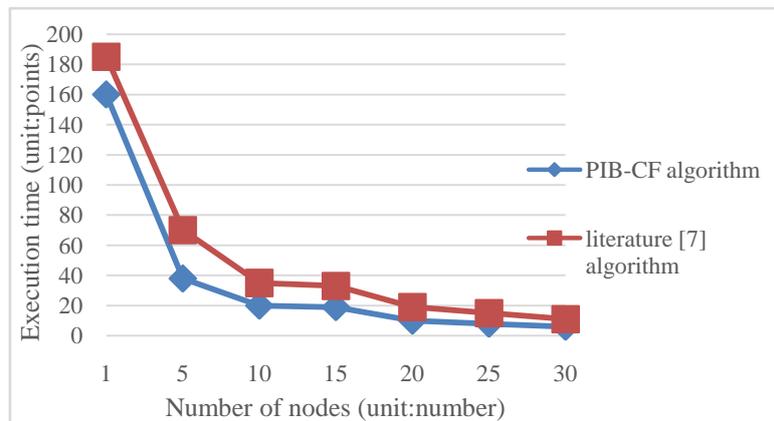
Fig. 11. Execution time comparison between PIB-CF algorithm and literature

In this chapter, a new parallel recommendation algorithm PIB-CF is proposed, which is implemented in Hadoop cluster. It can reduce the data communication and execution time by reasonably allocating the computation intensive similarity calculation process and Item prediction scoring process to each processing node. Experimental results show that PIB-CF algorithm is effective.

5. Summary

Cloud computing has become a hot technology in academia and industry. Many universities, scientific research institutions and Internet giants have invested in the research of cloud computing, which promotes the continuous upgrading from computer hardware to software, and promotes the development of computer and network technology. By purchasing cloud services, governments, universities and small enterprises do not need to purchase hardware resources, but only need a small amount of investment to purchase services from cloud service providers. This can not only reduce capital investment and make full use of existing resources, but also obtain complete solutions from cloud service providers. With enterprises turning to cloud platforms, users can enjoy the convenience of cloud computing and cloud storage anytime and anywhere through mobile terminals. The emergence of cloud computing technology has become an effective solution to solve the problem of big data. Combine cloud computing technology, recommendation technology and data mining technology.

Through the method of data mining, using the huge computing and storage advantages of cloud computing, the recommendation system can get faster and more accurate recommendation results. In this paper, a parallel collaborative filtering recommendation algorithm PIB-CF is proposed. This paper parallelizes the Item-based Collaborative Filtering Algorithm on Hadoop platform. The existing parallelization method is improved, which makes the improved algorithm run faster and more practical. Moreover, the experimental results show that the improved algorithm has good scalability.

Through cloud computing technology, the collaborative filtering algorithm based on project is parallelized in the platform. And through the hybrid recommendation technology, the recommendation accuracy is effectively improved. However, due to the limited research time, there are still some problems to be studied: the sparsity of item rating matrix is the biggest problem faced by the recommendation system. Because the items that users have purchased or browsed occupy only a small part of the total project, the similarity between users or projects cannot reflect the real relationship between them correctly. Therefore, matrix filling becomes the key problem to improve the accuracy of recommendation. The problem of cold start can not be solved effectively in the recommendation system. For a new user, due to the lack of historical information, how to obtain high-precision and comprehensive recommendation from less historical information has become a hot issue in the research of recommendation system. In this paper, we propose a hybrid recommendation algorithm to improve the recommendation accuracy. Moreover, the cold start problem is considered at the beginning of the algorithm design. However, due to the lack of research time, it was not able to complete. In the future research, we will continue to complete the research of cold start.

References

1. Lv, Zhihan, Weijia Kong, Xin Zhang, Dingde Jiang, Haibin Lv, and Xiaohui Lu.:Intelligent Security Planning for Regional Distributed Energy Internet. IEEE Transactions on Industrial Informatics, (2019)
2. Tunkelang D.: Recommendations as a Conversation with the User//Proc of the 5th ACM Conf on Recommender Systems. New York: ACM, 11-12. (2011)
3. Adomavicus G, Tuzhilin A.:Tpward the Next Generaton Od Recommender Systems: A Survey of the State-of-the-art and Possible Extensions. IEEE Transactions on Knowledge and Data Engineering, 17(6), 734-749.(2005)
4. Balabanović M, Shoham Y.: Fab: Content-Based, Collaborative Recommendation. Communications of the ACM, 40(3), 734-749.(1997)
5. Goldberg D, Nichols D, Oki B M.: Using Collaborative Filtering to Weave an Information Tapestry. Communications of the Acm, 35(12), 61-70.(1992)
6. Linden G, Smith B, York J.: Amazon. Com Recommendations: Item-to-item Collaborative Filtering. Internet Computing, IEEE, 7(1), 76-80.(2003)
7. Sandvig J J, Mobasher B, Burke R.: Robustness of the 2007 ACM Conference on Recommender Systems. Minneapolis, USA. Acm,105-112.(2007)
8. Wu, Y., Rong, B., Salehian, K., & Gagnon, G.: Cloud Transmission: A New Spectrum-Reuse Friendly Digital Terrestrial Broadcasting Transmission System. IEEE Transactions on Broadcasting, 58(3), 329-337. (2012)
9. Zhang, Y., He, Q., Xiang, Y., Zhang, L. Y., Liu, B., Chen, J., &Xie, Y.: Low-cost and Confidentiality-Preserving Data Acquisition for Internet of Multimedia Things. IEEE Internet of Things Journal, 5(5), 3442-3451. (2017)

10. Jiang J, LU J, Zhang G,.: Scaling-up Item-based Collaborative Filtering Recommendation Algorithm Based on Hadoop. Proceedings of 2011 IEEE World Congress on Services. Washington Marriott, DC, USA. IEEE, 490-497.(2011)
11. Resnick, Sushak.:GroupLens: An Open Architecture for Collaborative Filtering of Netnews. Proceedings of the 1994 Computer Supported Collaborative Conference.(1994)
12. Lu Jiaheng.: Hadoop practice. Beijing: China Machine Press.(2012)
13. Han, B., Li, J., Su, J., & Cao, J.: Self-supported Cooperative Networking for Emergency Services in Multi-hop Wireless Networks. IEEE Journal on Selected Areas in Communications, 30(2), 450-457.(2012)
14. Konstantin S, Kuang H, Radia S.: The HadoopDistributed File System. Symposium on Massive Storage SystemsandThechnologies, 23(7), 385-394.(2010)
15. Ahn H J.: A New Similarity Measure for Collaborative Filtering to Alleviate the New User Cold-starting Problem. Information Sciences,178(1), 37-51.(2008)
16. Yue, L.: Cable-driven MIS Robot with Intuitive Manipulability and Direct Force Feedback. Investigación Clínica, 60(4),813-824.(2019)
17. Han, Biao, Jie Li, Jinshu Su, Minyi Guo, and Baokang Zhao.: Secrecy Capacity Optimization via Cooperative Relaying and Jamming for WANETs.IEEE Transactions on Parallel and Distributed Systems, 26, No. 4, 1117-1128.(2014)

Pei Tian was born in Xi'an, Shaanxi. P.R. China, in 1983. He received the master's degree from Northwest University of political science and law, P.R. China. Now, he works in the human resources center, Xi'an Peihua University. His research interest include cloud computing and e-commerce collaborative filtering. E-mail: tianpei@peihua.edu.com

The Application of Virtual Reality Technology in the Digital Preservation of Cultural Heritage

Hong Zhong[1], Leilei Wang[2], and Heqing Zhang[3, *]

[1]Guilin Tourism University, Guilin 541006,
Guangxi, China
zhonghong@gltu.edu.cn
[2]Guangzhou Panyu Polytechnic, Guangzhou 511483,
Guangdong, China
Leileiwang@gzpyp.edu.cn
[3]Tourism College of Guangzhou University, Guangzhou 510006,
Guangdong, China
clear8007@163.com

Abstract: Virtual reality technology involves computer graphics, artificial intelligence, network, sensor technology and many other aspects. It can use the powerful computing and graphics processing capabilities of computers to provide alternatives to the original and express its visual, tactile, and auditory technical means. According to archaeological research data And documentary records, virtual reconstruction and simulated display of the cultural heritage that has been wiped out. "Digital protection" of cultural heritage is a new way of protection, relying on computer technology, and the use of digital equipment to collect, save, process, output and disseminate the required information, including databases established on computer systems, So as to achieve the purpose of information sharing and dissemination. This article mainly studies the application research of virtual reality technology in the digital preservation of cultural heritage. Create an immersive environment for users, display the objects realistically in the virtual reality system, thereby digitizing the technical protection of cultural heritage; secondly, use the virtual environment model of material cultural protection to build and use the terrain to generate and edit The device imitates the terrain of the natural world to achieve its position and the effect of being in it. Finally, the radial basis function is used to calculate the value in the virtual environment, so that the digital preservation of cultural heritage is more accurate. Experimental data shows that 35.54% and 64.46% of users are more likely to use the handle to interact with three-dimensional objects. They believe that the speed of the handle has changed and the control is more precise. Experimental results show that: The virtual environment reality technology specification is more efficient than the original technology in the process of digitizing cultural heritage.

Keywords: Virtual Reality Technology, Digitization of Cultural Heritage, Virtual Interaction, Virtual Environment Model, Radial Basis Function

---

[*] Corresponding author

1. Introduction

1.1.   Background and Significance

Virtual reality technology is a technology that uses a variety of interface devices such as a computer drawing system and reality control to provide a sense of engagement in a three-dimensional conversational environment generated in a computer [1]. Urban planning, interior decoration design, simulation industry, restoration monument design, bridge and road design, real estate sales, tourism education, water resources protection, electricity, geological disasters, etc. have been widely used to provide practical solutions [2]. Cultural heritage is a precious and non-renewable resource. The trend of economic globalization and accelerating the pace of modernization have brought great changes to China's cultural ecology, and its cultural heritage and living environment are seriously threatened. Many historical and cultural cities (blocks, villages and villages), ancient buildings, ruins and scenery have been destroyed. Due to excessive exploitation and unreasonable use, many important cultural heritages have disappeared or disappeared. In areas where ethnic minorities live in a relatively rich cultural heritage, due to changes in people's living environment and conditions, national or regional cultural characteristics have accelerated their disappearance. Therefore, it is urgent to strengthen the protection of cultural heritage.

The technology that cannot be achieved with existing protection methods provides a better method with the development of virtual reality technology [3]. Among the gradually disappearing cultures, some cultures can continue to live through virtual reality technology. In the virtual space, keep as many special treasures as possible, greatly reducing the number of tourists visiting and reducing the damage to murals and sculptures [4-5]. Some cultural heritage that has disappeared can be used to get a new life. Using virtual reality technology not only can better organize archaeological excavation and research work, but also can use virtual technology to present today's museums without affecting the original cultural heritage institutions. In order to let the audience understand the content of cultural heritage, important trends have also appeared in recent years [6-7].

1.2.   Related Work

Maples-Keller J L believes that VR can control the transmission of sensory stimulation by the therapist, which is a convenient and cost-effective treatment method. Sensory information is transmitted through the head-mounted display and dedicated interface devices [8]. These devices track head movements so as to naturally change motion and images with head movements, thereby making people immersive. His discussion focuses on the existing literature on the effectiveness of incorporating VR into the treatment of various psychiatric disorders, especially for exposure interventions based on anxiety disorders. In order to determine the research on the treatment of anxiety or other mental diseases based on VR, a systematic literature search was conducted [9]. Sang Y introduced an interactive truck crane simulation platform based on virtual reality technology, on which the simulation experiment of crane movement can be

completed [10]. He discussed the framework and working principle of the interactive truck crane simulation platform, taking the lifting feet and hooks as examples to illustrate the motion control mechanism of truck crane components. Interactive truck crane simulation platform utilizes browser-based structure, Java3D, virtual reality and Java Applet to develop Web3D virtual reality learning environment, which has good advantages [11]. In the past few years, Coburn J Q believes that consumer virtual reality (VR) devices have made some significant advances. Immersive VR experiences have also entered consumer homes, and their cost and space requirements are much lower than the previous generation of VR hardware. These new devices also lower the entry barrier for VR engineering applications. Past research has shown that there are great opportunities for using VR during design tasks, which can improve results and reduce development time. His work reviewed the latest generation of VR hardware and reviewed the research on VR during the design process. In addition, this work extracts the main topics from the comments and discusses how the latest technology and research affect the engineering design process. They concluded that these new devices have the potential to significantly improve some parts of the design process [12].

1.3.   Innovation in this Article

The main innovative work of this paper includes the following aspects: (1) This paper proposes a method based on VR three-dimensional scenes to obtain projection presentation data sets in the form of three-dimensional space coordinate transformation and multi-view sampling to solve the problem of model feature construction. (2) Propose and implement a method for interactively acquiring 3D models in VR 3D scenes, and specifically improve the output of traditional 3D models. View output and 3D model output have been improved to dynamically load models in 3D scenes.

2.   Virtual Technology Algorithm of Cultural Relics

2.1.   Three-Dimensional Construction of Cultural Relics Buildings Based on Virtual Interaction

Conceptual model of virtual reality

The three-dimensional virtual interactive system is a computer-generated virtual environment that gives a variety of sensory stimuli and is an advanced human-computer interactive system [13-14]. According to the definition, virtual reality consists of two parts: one part is the created virtual world (environment), and the other part is the intervene (person). The core of virtual reality is to emphasize the interaction between the two, that is, to reflect the human experience in the virtual world (environment) [15]. In this way, we can get the conceptual model of virtual reality as shown in Figure 1.
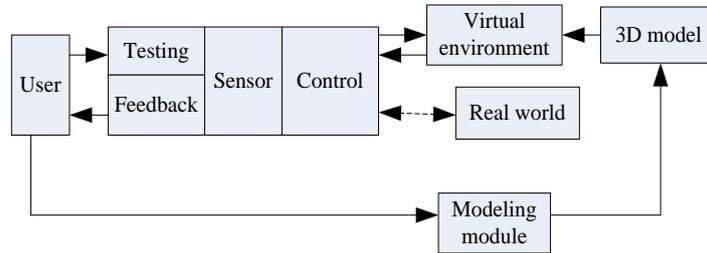
Fig. 1. Conceptual model of virtual reality

A successful three-dimensional virtual interactive system will inevitably create an environment for users to feel immersive [16]. To achieve this effect, it is necessary to realistically display the objects represented in the virtual reality system [17-18]. Not only does it resemble real objects in appearance, but it also requires good performance in terms of form, light and shadow, and texture. To achieve this requirement, the technical implementation can be divided into four steps: the first step is geometric modeling, which mainly establishes the geometric model of the three-dimensional scene; the second step is image modeling, which mainly focuses on the results of geometric modeling Perform material, lighting, color and other processing [19]; the third step is behavior modeling, which mainly deals with the behavior and motion description of objects [20-21]. As shown in Fig. 2.
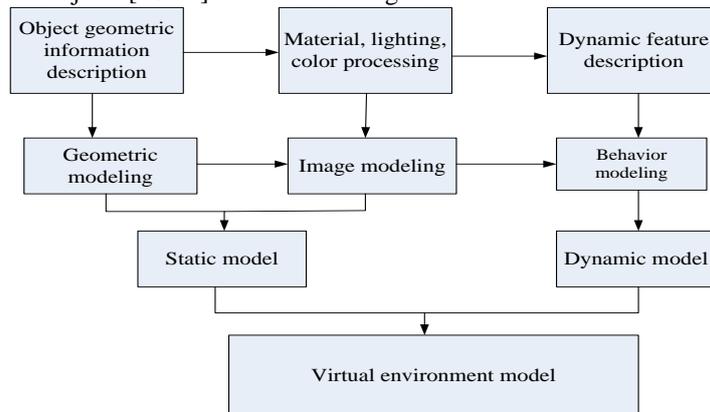
Fig. 2. Flowchart of virtual environment model

Characteristics of cultural relics building and its environment modeling

As an important field of virtual reality research, the virtual realization of architectural heritage protection has aroused great interest of researchers in recent years [22]. Compared with other graphical modeling, cultural relics architecture and its environment modeling have their own characteristics, mainly manifested in the following four aspects:

1) There can be many objects in cultural relic buildings and their environments, and it is often necessary to construct a large number of completely different types of models.

2) Most building components in cultural relics buildings must express the texture of the material surface through elaborate material texture.

3) The three-dimensional construction of the three-dimensional virtual interactive system Chinese building and its environment must be based on accurately reflecting the true spatial scale relationship of the architectural heritage and the texture characteristics of the building materials.

4) Some architectural components in the heritage buildings need to respond to the observer. When the observer interacts with the object, the object must react in some appropriate way and cannot ignore the observer's actions.

These modeling features put special requirements on the virtual environment modeling technology:

1) Reusability. There are many kinds of building components in the virtual environment, and building a model of building components often takes a lot of energy, so it is very necessary to build a standard model library that can be reused.

2) Authenticity and accuracy. The three-dimensional construction of cultural relic buildings and their environments must be based on various surveying and mapping data and historical archives.

3) When modeling, the texture mapping features of the model must be considered, and the form of its mapping coordinate mapping must be considered.

4) During the interaction, the model should provide corresponding prompts so that the interaction can proceed as intended.

Main technical indicators of virtual environment modeling

3D virtual environment modeling is one of the key technologies in virtual reality technology. Whether the model is built or not will directly affect the quality of the entire 3D virtual interactive system [23-24]. Some researchers even say that building a perfect three-dimensional model is more important than a thousand facts [25-26]. At present, many mature computer software can be used to establish virtual environment models, such as: CAD, 3DSMAX, Maya, VRML, etc[27-28]. To establish a good virtual environment should have a detailed understanding of the main technical indicators of modeling. The main technical indicators for evaluating virtual environment modeling are:

1) Precision. It is an index to measure the accuracy of the model to represent real objects.

2) Display speed. Many applications have large restrictions on the display time. In interactive applications, we hope that the shorter the response time, the better. If the response time is too long, it will greatly affect the availability of the system.

3) Manipulation efficiency. In the actual application process, the display of the model, the behavior of the motion model, and the collision detection in the virtual environment with multiple moving objects are all high-frequency operations that must be efficiently implemented.

4) Ease of use. Creating an effective model is a very complicated task. The modeler must represent the geometric and behavioral model of the object as accurately as possible. The modeling technology should construct and develop a good model as easily as possible.

5) Extensiveness. The broadness of modeling technology refers to the range of objects it can represent. Good modeling technology can provide a wide range of geometric modeling and behavior modeling of objects.

6) Real-time display. In a virtual environment, the display of the model must be above a certain limit frame rate, which often requires fast display algorithms and model reduction algorithms.

2.2. Construction of Virtual Environment Model in the Protection of Material Culture

The use of the terrain generation editor has a great advantage in imitating natural terrain, but it also has specific disadvantages [29]. Such restrictions occur in modern cities, small towns, underground palaces, etc. whose scales are beginning to appear, when relatively regular terrain prompts. When using the terrain editor to create terrain, the number of ground maps cannot be large [30-31]. The greater the number of maps selected, the lower the performance of the server. Two to three images are the best choice. In addition, the ground in the city needs a lot of textures. There are more than three highways, pedestrian crossings, lawns, etc. The streets of the city are very neat, and it is not impossible for the terrain generation editor [32-33]. Another important issue is that the ground constructed in this way is particularly space-saving. Necessary skills: The operation of 3ds max software requires proficiency. Of course, you also need to learn some specific lighting and render applications, and the function of making environment maps.

Knowledge points for constructing environment model: The visual area is the area of the scene that the audience can see in the interactive experience with virtual reality [34]. The visual area changes constantly according to the listener's movements, and the scene is divided according to this level [35]. The division of the scene area shows that no matter where the field of view is in the interactive area, the direction of the field of view sometimes only displays hundreds of scenes. If we put the whole scene together, we can interactively get the model of the whole scene to render together. Therefore, the number of models in the scene cannot be too large. Too many models will seriously affect the efficiency and execution speed of the interaction. When this happens, you can use block swapping to change work efficiency. When the scene is enlarged in the block, only the model within the display angle is rendered, and the model that is not completely displayed is not rendered. The optimization of the overall modeling scene can be obtained by transforming it into a square-like area. HDRI high dynamic range images can cover buildings and roads in the environment with a lighting information layer. The size of the HDRI high dynamic image has a decisive influence on the sharpness of the eye shadows in the scene and the atmosphere of the entire scene. In addition, HDRI high dynamic images can not only generate shadows, but also draw conclusions from the perspective of the overlay model. The new inherent color formed by the color of amber and light. The size of a bright HDRI high dynamic image needs to be adjusted according to each model. HDRI high dynamic images cannot exceed 1024 * 1024. The render can be used to render HDRI high dynamic images. With 3ds max, what kind of effect can actually be produced, what kind of result can be obtained in the virtual interactive scene. The rendering time and quality of various renders are completely different, and it takes time to achieve a better rendering effect. The role of optics is to

illuminate the scene and give it an atmosphere. The generation of lighting effects is very important for the atmosphere of the scene. In the absence of lighting effects, HDRI high dynamic images are not required. No shadow can be formed, there is no atmosphere, and the entire environment loses the depth of the three-dimensional space. The seamless texture is represented by four azimuth boundaries for each texture. Seamless texture means that even if the random side of this texture is completely merged with the other three sides, no trace of convergence can be seen.

2.3.   Radial Basis Point Interpolation Method in Virtual Environment

The radial basis function (RBF) calculates the distance between the node x and the node xi. The radial basis function has a simple structural form, good stability, and has an isotropic point. In numerical calculation and surface it has been widely used in fitting. In particular, it can help numerical calculation in a virtual environment. For two-dimensional plane problems, the stress, strain, and displacement components can be expressed as follows, where L is a differential operator:

$$\varepsilon = Lu \tag{1}$$

For a two-dimensional elastic solid, the law can be used to express its constitutive relationship. Its matrix is as follows

$$\sigma = D\varepsilon \tag{2}$$

In the formula (2), σ and ε are the components of stress and strain, respectively, and D is expressed as the constant matrix of the material. In order to make the weak form of the balance equation easy to operate and solve, we need to describe the problem defined in the problem domain, number the nodes, first solve the interpolation coefficients in the radial basis point interpolation shape function, and then form the population through mathematical derivation. Then, calculate the overall body force vector of all nodes, and finally form the overall governing equation. Through mathematical derivation and transformation, you can get

$$Ku = F \tag{3}$$

In the formula (3), K is the overall stiffness matrix; F is the overall physical force vector. For the overall physical force vector, it can be further expressed as:

$$F = F^{b} + F^{t} = \int_{\Omega} \varphi_i b d\Omega + \int_{\Gamma_t} \varphi_i t d\Gamma \tag{4}$$

By solving the system of equations, the displacement of the node can be obtained first, and then the stress and strain of the node can be further obtained by the formula.

3.   Virtual Interactive Eye Tracking Experiment

3.1.   Experimental Environment Configuration

In this experiment, we used HTC Vive combined with a Glass DK II eye movement module to record the eye movement data of the subjects. The HTC Vive virtual reality helmet integrates the tracking technology and positioning guidance system of the Steam VR platform [36], with a binocular resolution of 1980 * 1080 (monocular 1080 * 1200 resolution), its refresh frequency is 80Hz, and the tracking accuracy reaches 0.08 degree. Computer hardware and software platform: a desktop computer equipped with a GTX1080Ti graphics card, an Intel Core i7-8700K processor, and a DELL LCD monitor. The test procedures for this study were edited by Unity3D software.

3.2.   Experimental Procedure

Because each person's eye feature structure is different, in order to ensure the accuracy of the eye movement data, eye movement calibration is required for all subjects before the experiment starts. There is no time limit for the operation of this experiment. According to the needs of the experiment, several indicators as shown in Table 1 were selected to analyze the cognitive state of the subjects.

Table 1. Experimental data analysis indicators

Experimental sequence	Evaluation index	Remarks
Experiment 1, 3, 4	FT (Finish time, s)	Time taken to complete the task
	TtFF (Time to first fixation, s)	The time from the start of the experiment to the first time you are watched
	ACC (Average time needed for a correct click, s)	The average time to complete the task correctly
	AC (Accuracy)	The correct rate selected by all subjects
	VR (Visit ratio, count)	Proportion of points at which objects are fixed

The calculation method of the data analysis indicators in Table 1 is as follows:

(1) Time to complete the task (FT). This parameter is the time required for the subjects to enter the experimental test task and complete all the experimental tasks:

$$FT = t_{end} - t_{begin} \quad (5)$$

In formula (5), t_{end} represents the time stamp of task completion and t_{begin} represents the time stamp of task start.

The time from the beginning of the experiment to the first observation (TtFF). This parameter is the time from the beginning of the experiment to the first time the subject finds the target correctly:

$$TtFF = t_{find} - t_{begin} \quad (6)$$

In formula (6), t_{find} represents the time stamp when the target is found for the first time, and t_{begin} represents the time stamp at the start of the task.

The average time to complete the task correctly (ACC). This parameter represents the average time the subjects correctly choose:

$$ACC = \frac{\sum_{i=1}^{N} t_{rf}}{N} \quad (7)$$

In equation (7), t_{rf} represents the time to complete the task correctly, and N represents the number of subjects who completed the task correctly.

The correct rate (AC) selected by all the test subjects. This parameter represents the proportion of subjects who correctly complete the task under different color mapping schemes:

$$AC = \frac{N_r}{N_{total}} \quad (8)$$

In equation (8), N_r represents the number of subjects who completed the task correctly, and N total represents the number of all subjects.

The fixation ratio (VR). This parameter indicates the proportion of the subject's viewpoint falling on the target object:

$$VR = \frac{C_{object}}{C_{total}} \quad (9)$$

In equation (9), C_{object} represents the number of effective viewpoints falling on the surface of the target object, and C_{total} represents the number of all effective viewpoints.

3.3.   Data Collection

In this experiment, a total of 20 subjects were divided into two groups, each group of 10 people, receiving the same experimental test tasks. Aged between 24 and 28 years old, in order to ensure that all the tests have the same cognitive level, the subjects have no experience of virtual reality experience, and the tests have professional knowledge related to cartography and geographic information systems. The subjects' naked eyesight and myopia were not higher than 600 degrees, without color blindness.

4.   Discussion and Analysis of Experimental Results

4.1.   Analysis of Virtual Reality Tracking Technology

Analysis of task settings

A total of 10 configuration tasks are set, and the object operates according to the target position through translation and rotation. The target position is given by a line frame model with the same shape as the object being operated. These 10 tasks are timed tasks. In order not to let users get too tired, the maximum time limit for each task is 65 seconds. There are three ways to enter the next task. The user will automatically match the target location and automatically enter the next task. After a period of time, the task will automatically enter the next task and choose to abort the current task. The configuration task is set with an error threshold. When the user locates the target within the error threshold range of the target position, the color of the target changes to green, indicating that the task was successful and the next task begins. In order to improve the accuracy of the configuration, the error thresholds in the three directions of x, y, and z are set to 0.003 meters. As the time of the task elapses, if the user does not locate the target at the target position, the task will fail and the position of the last second of the target will be saved. Table 2 shows the description of each task.

Each user needs to complete the testing of the two devices separately. To prevent deviations in the test sequence, half of the users first test the interaction of the control handle, and the other half of the users first test the interaction of the multi-entity tracker. In order to maintain consistency with entity interaction, when interacting with the handle, the task starts timing when the handle grabs the object, not counting the time it takes the user to grab the object. For tasks with ten matching locations, each user performs two rounds of testing for each device. Each round of tasks is always displayed in the same order, as shown in Table 2. The target position of each task is set according to the Latin matrix to distinguish between different rounds. Record the completion time of each task, the deviation of the position of the model, and calculate the success rate of the task of each type of equipment. For the additional two tasks, only do the test of the handle interaction, do two rounds, each round of 2 tasks, each task includes 2 interactions.

Table 2. Task settings

Serial No.	Task	Description
1	x-axis translation	translate the object in the x-axis direction to match the target position
2	y-axis translation	translate the object in the y-axis direction to match the target position
3	z-axis translation	translate the object in the z-axis direction to match the target position
4	x- / y- / z-axis translation	translate the object in the xyz axis direction to match the target position
5	x-axis rotation	rotate the object around the x axis to match the target position
6	y-axis rotation	rotate the object around the y axis to match the target position
7	z-axis rotation	rotate the object around the z axis to match the target position
8	x- / y- / z-axis rotation	rotate the object around the xyz axis to match the target position
9	rotate around the z axis and translate along the xy axis	match the object to the target position
10	six full degrees of freedom	match the object to the target position

Analysis of test results

The time to complete the interaction of the handle starts when the user grabs the object using the handle and starts counting, and stops when the user releases the handle to place the object within the allowable error of the target position. The completion time of the physical tracker interaction starts when the user uses the tracker to collide with the object, and stops counting when the collision detection trigger calculates that the object has reached the allowable error of the target position. The results of experiments that did not successfully complete the task were not included in the final results. For the translation task, the rotation task and the average completion time of the tasks with 6 degrees of freedom to complete the operation are shown in the Fig. 3.
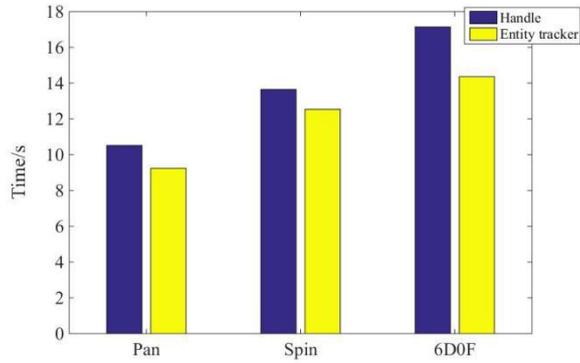
Fig. 3. Average task completion time

The needle diagram of the time taken by the entity tracker and the handle to interact with each task is shown in Fig. 4. The task completion time is in seconds. The table shows the median, upper quartile, and lower quartile, with the minimum and maximum time to complete (the length of the vertical bar).
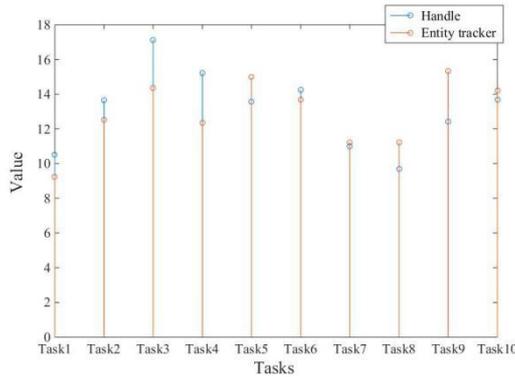
Fig. 4. Time taken by the physical tracker to interact with the handle

4.2.   Experimental Analysis of Illumination-Three-Dimensional Model Attribute Mapping

This part of the experiment includes 1 test: the qualitative test of the city 3D model attribute value under the synergy of color and light visual variables, transparency = 1. According to the attribute data, it can be known that the building model id 52, 33, 48, 36, 55 is the first 10% of the attribute value is smaller. As shown in Fig. 5, under the condition of light intensity of 0.4, the test results expressed by the isometric color mapping are relatively uniform, followed by linear color mapping, circular color mapping and divergent color mapping. The lower the value, the darker the entire test scene, giving the subject a visually low feeling, which affects the subject's judgment. It

can be seen from Fig.. 6 that the result of the test scene based on iso-illuminance color mapping is still uniform, while linear color mapping is next.
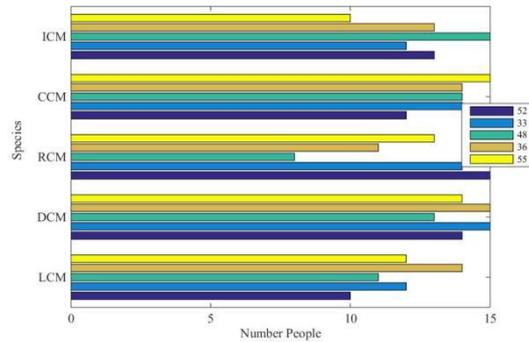
Fig. 5. Statistics of subject selection results under light intensity of 0.4
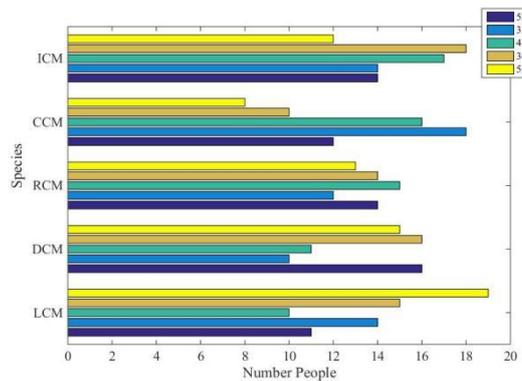
Fig. 6. Statistics of the results of subject selection under the condition of light intensity 1.0

According to the analysis of Tables 3 and 4, it is found that under the condition of light intensity of 0.4, the task completion time is longer than that of natural light, but under the condition of light intensity of 0.4, the accuracy of the corresponding illumination mapping scheme is improved, and the first attention The time is also shortened. Combined with the intuitive analysis of the statistical chart, it can be concluded that the iso-illuminance mapping scheme is effective for dark lighting conditions; in addition, when the lighting condition is 1.0, the overall task completion time becomes longer, but in the isometric color. Under the mapped scene, the choices made by the subjects are more accurate, and the time of the first attention is earlier than the natural environment. This shows that when the illumination variable changes, the isometric color mapping is the most effective.

Table 3. The average value of eye movement index of qualitative test scene with light intensity of 0.4

	LCM	DCM	RCM	CCM	ICM
FT(s)	5.32	5.43	5.62	5.64	5.21
TtFF(s)	0.52	0.63	0.51	0.34	0.45
ACC(s)	0.08	0.35	0.24	0.15	0.07
AC	0.65	0.35	0.24	0.62	0.54
VR	0.63	0.25	0.34	0.36	0.42

Table 4. The average value of eye movement index of qualitative test scene with light intensity of 1.0

	LCM	DCM	RCM	CCM	ICM
FT(s)	6.32	6.35	6.59	6.32	6.24
TtFF(s)	0.75	0.74	0.76	1.02	0.36
ACC(s)	0.08	0.23	0.14	0.25	0.22
AC	0.75	0.65	0.55	0.61	0.45
VR	0.66	0.35	0.31	0.25	0.66

The light-three-dimensional model attribute mapping experiment conducts extensive research on the visual variables of the three-dimensional model in the form of user cognition experiments. Through the cross combination of visual variable conditions such as color, transparency, and lighting, the user cognition experiment is designed. The indicator calculates the average value from the data generated by the experiment, and combines it with the statistical chart analysis of the results of the subject's answer selection to find the most suitable color mapping under each condition. According to the user's cognitive experiment results, combined with the process determination method, a three-dimensional model symbol visual variable mapping model is constructed, and appropriate model evaluation indicators are selected to evaluate the rationality of the constructed model through cognitive experiments.

5.   Conclusions

Based on the development of the three-dimensional virtual simulation of cultural heritage protection, this paper has studied and researched most of the three-dimensional virtual interactive technology and the three-dimensional virtual interactive development tool platform at home and abroad, and determined the realization method and technical route of the research object. According to the actual situation and technical conditions, the virtools platform uses the interactive functions of the system to implement the virtools behavior interaction module and establish a virtual interactive program. This greatly shortens the development cycle of the cultural heritage protection 3D virtual simulation system and optimizes the system development process.

Aiming at the digital protection of architectural models of cultural heritage monuments, this paper uses VR-based 3D scenes as research tools and VR-based 3D feature extraction and search as research goals to realize the cultural heritage digital application. This article overcomes the differences in the user's interpretation of the model. It directly searches the 3D 3D model with the 3D space shape manually drawn in the VR environment, avoids the overall route of the large building complex, and loads the searched consistent pattern into a monument dynamically with the VR scene model. Realize the virtual interaction of the digitization of historical monuments. The visual and historical information of cultural heritage is recorded and preserved in digital form, and the shape data is reproduced and stored in the form of three-dimensional digital model. With the help of virtual reality and human-computer interaction technology, it provides an intuitive and vivid visual image and modern media platform.

The rapid development of digital technology has brought unprecedented opportunities for the development and protection of China's cultural heritage. It is promoting the full realization of the sharing and inheritance of cultural heritage in an unprecedented manner and speed. The digital system of cultural landscape heritage relies on the network platform and uses a variety of digital technologies to create virtual cultural heritage scenarios. It integrates text, pictures, audio, video, animation and virtual interactive experience to establish a cultural resource protection and sharing system.

References

1. Lv, Zhihan, Dongliang Chen, Ranran Lou, and Houbing Song. "Industrial Security Solution for Virtual Reality."IEEE Internet of Things Journal(2020)
2. Y. Sun, H. Song, A. J. Jara and R. Bie, "Internet of Things and Big Data Analytics for Smart and Connected Communities," inIEEE Access, vol. 4, pp. 766-773, 2016.
3. Lv, Z. (2019). Virtual reality in the context of Internet of Things. Neural Computing and Applications, 1-10.
4. Donghui C, Guanfa L, Wensheng Z, et al. Virtual reality technology applied in digitalization of cultural heritage. Cluster Computing, 22(4):1-12. (2019)
5. Huang Y , Huang Q , Ali S , et al. Rehabilitation using virtual reality technology: a bibliometric analysis, 1996–2015. Scientometrics, 109(3):1547-1559. (2016)
6. Z. Liang, R. Shuang. Research on the value identification and protection of traditional village based on virtual reality technology. boletin tecnico/technical bulletin, 55(4):592-600. (2017)
7. Zhang H, Zheng H. Research on interior design based on virtual reality technology. Boletin Tecnico/Technical Bulletin, 55(6):380-385. (2017)
8. Lv, Zhihan, Liang Qiao, Qingjun Wang, and Francesco Piccialli. "Advanced Machine-Learning Methods for Brain-Computer Interfacing."IEEE/ACM Transactions on Computational Biology and Bioinformatics(2020).

9. 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. Harvard Review of Psychiatry, 25(3):103-113. (2017)

10. Krishnaraj, N, Elhoseny, M, Lydia, EL, Shankar, K, ALDabbas, O. An efficient radix trie-based semantic visual indexing model for large-scale image retrieval in cloud environment. Software Practice and Experience,2020; In Press.

11. Sang Y, Zhu Y, Zhao H, et al. Study on an Interactive Truck Crane Simulation Platform Based on Virtual Reality Technology. International journal of distance education technologies, 14(2):64-78. (2016)

12. Coburn J Q, Freeman I J, Salmon J L. A Review of the Capabilities of Current Low-Cost Virtual Reality Technology and Its Potential to Enhance the Design Process. Journal of Computing & Information Science in Engineering, 17(3):031013.1-031013.15. (2017)

13. T N Chen, X T Yin, X G Li. Application of 3D virtual reality technology with multi-modality fusion in resection of glioma located in central sulcus region.. zhonghua yi xue za zhi, 98(17):1302-1305. (2018)

14. Zeming L. Design and implementation of a Korean language teaching system based on virtual reality technology. Agro Food Industry Hi Tech, 28(1):2156-2159. (2017)

15. Yan, Zuohao, and Zhihan Lv. "The Influence of Immersive Virtual Reality Systems on Online Social Application."Applied Sciences10, no. 15 (2020): 5058.

16. Ciatto, Giovanni; Mariani, Stefano; Omicini, Andrea. ReSpecTX: Programming Interaction Made Easy. Computer Science and Information Systems. 2018. 15(3), 655-682.

17. Milica Lajbenšperger, Marija Šegan, Sanja Rajić. The use of the modern technology in educational work: user study for digitization of cultural heritage. Journal of Education Culture & Society, 4(2):71-78. (2020)

18. Stefano, A. de, Tausch, R, Santos, P. Modeling a virtual robotic system for automated 3D digitization of cultural heritage artifacts. Journal of cultural heritage, 19(9):531-537. (2016)

19. Hou, Y., & Wang, Q. (2018) "Research and Improvement of Content-Based Image Retrieval Framework", International Journal of Pattern Recognition and Artificial Intelligence, 32(12), 1850043.

20. Montagnani, Maria Lillà, Zoboli L. The making of an 'orphan': Cultural heritage digitization in the EU. International journal of law and information technology, 25(3):196-212. (2017)

21. Budak I, Santoi E, Stojakovi V, et al. Development of Expert System for the Selection of 3D Digitization Method in Tangible Cultural Heritage. Tehnicki Vjesnik, 26(3):837-844. (2019)

22. Ciurea C, Filip F G. New Researches on the Role of Virtual Exhibitions in Digitization, Preservation and Valorization of Cultural Heritage. Informatica Economica, 20(4/2016):26-33. (2016)

23. Shaw E F. Making Digitization Count: Assessing the Value and Impact of Cultural Heritage Digitization. Archiving Conference, 2016(1):197-201. (2016)

24. Huda M, Siregar M, Ramlan, et al. From live interaction to virtual interaction: Addressing moral engagement in the digital era. Journal of Theoretical and Applied Information Technology, 95(19):4964-4972. (2017)

25. Wang X, Ong S K, Nee A Y C. Real-virtual components interaction for assembly simulation and planning. Robotics & Computer Integrated Manufacturing, 41(OCT.):102-114. (2016)

26. Hartwick, Peggy. Investigating research approaches: Classroom-based interaction studies in physical and virtual contexts. ReCALL,1-16. (2018)

27. Kim M, Lee D. Improving transparency of virtual coupling for haptic interaction with human force observer. Robotica, 35(2):354-369. (2017)

28. Muenzer S, Zadeh M V. Acquisition of spatial knowledge through self-directed interaction with a virtual model of a multi-level building: Effects of training and individual differences. Computers in Human Behavior, 64(nov.):191-205. (2016)

29. Bevacqua E, Richard R, Loor P D. Believability and Co-presence in Human-Virtual Character Interaction. IEEE Computer Graphics and Applications, 37(4):17-29. (2017)

30. Williams T, Szafir D, Chakraborti T, et al. Report on the First International Workshop on Virtual, Augmented, and Mixed Reality for Human-Robot Interaction. AI Magazine, 39(4):64-66. (2018)
31. Nascimento Araujo T, Moreira Salles R. A Model for Management of Virtual Networks in OpenFlow Environment. IEEE Latin America Transactions, 14(10):4318-4326. (2016)
32. Li Y, Liu Q, Tan S R, et al. High-resolution time-frequency analysis of EEG signals using multiscale radial basis functions. Neurocomputing, 195(jun.26):96-103. (2016)
33. Christelis V, Mantoglou A. Pumping Optimization of Coastal Aquifers Assisted by Adaptive Metamodelling Methods and Radial Basis Functions. Water Resources Management, 30(15):5845-5859. (2016)
34. Li, Xingchen; Zhang, Weizhe; Wang, Desheng; Zhang, Bin; He, Hui. Algorithm of Web Page Similarity Comparison Based on Visual Block. Computer Science and Information Systems, 2019, 16(3), pp. 815-830.
35. Lv, Zhihan,Liang Qiao, Amit Kumar Singh, Qingjun Wang. "Fine-grained Visual Computing Based on Deep Learning."ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)(2020).
36. M.Elhoseny, Multi-object Detection and Tracking (MODT) MachineLearning Model for Real-Time Video Surveillance Systems, Circuits, Systems, and Signal Processing, First Online: 20 August 2019 39, pp. 611–630.

Hong Zhong was born in Lipu County, Guilin, Guangxi Province in 1971.She received PH.D degree from Beijing Forestry University and works at Guilin Tourism University now. Her research interests include Tourism Management, Tourism Culture.
E-mail: zhonghong@gltu.edu.cn

LeileiWang was born in Kaifeng, Henan P.R. China, in 1985. She received the master's degree from Guangdong University of Finance and Economics, P.R. China. Now, her research interest include Tourism Management, Tourism Culture and big data analysis. E-mail: Leileiwang@gzpyp.edu.cn

Heqing Zhang was born in, Xinning, Hunan P.R. China, in 1967. He received the PH.D from Sichuan University, P.R. China. Now, He works in School of Tourism of Guangzhou University, his research interests include Tourism Management, Tourism Culture. E-mail: lyzhq8007@ gzhu.edu.cn

Extraction of Mosaic Regions through Projection and Filtering of Features from Image Big Data

Seok-Woo Jang[1]

[1] Department of Software, Anyang University
22, 37-Beongil, Samdeok-Ro, Manan-Gu, Anyang 14028, Korea,
swjang7285@gmail.com

Abstract. When uploading multimedia data such as photos or videos on social network services, websites, and so on, certain parts of the human body or personal information are often exposed. Therefore, it is frequent that the face of a person is blurred out to protect the portrait right of a particular person, and that repulsive objects are covered with mosaic blocks to prevent others from feeling disgusted. In this paper, an algorithm that detects mosaic regions blurring out certain blocks based on the edge projection is proposed. The proposed algorithm initially detects the edge and uses the horizontal and vertical line edge projections to detect the mosaic candidate blocks. Subsequently, geometrical features such as size, aspect ratio and compactness are used to filter the candidate mosaic blocks, and the actual mosaic blocks are finally detected. The experiment results showed that the proposed algorithm detected mosaic blocks more accurately than other methods.

Keywords: Video Analysis, Edge Extraction, Geometrical Feature, Filtering, Candidate Verification

1. Introduction

Recently, with the rapid development of mobile sensing, computer networks, high-definition cameras and artificial intelligence technologies, the amount of social multimedia data available is increasing explosively [1-5]. In particular, as smart mobile devices and location-based services are expanded, real-time location-based video data are actively used, and such digital image data are creating large-scale big data [6-10]. Additionally, related real application programs that can be used to effectively search and process such social multimedia data are diversely developed and utilized in various social computing fields [11-15].

In general, a color image includes valuable information but also includes information not preferred being exposed to other people. For example, Google Street View [16, 17] that can be used by ordinary people to freely search and confirm a map through an Internet browser sometimes exposes the face of a person or the license plate of a car, which is considered the personal information of a particular person. The service uses 11 lenses of 100 million pixels, which are captured in high quality to provide relatively detailed scenes such as bus routes, street structure and people.

However, in some cases, the street view service is too private, leading to a backlash and criticism of invading privacy. In fact, when she entered the address of a woman in

California into a street view search window, she saw a cat being raised through her living room window. Besides, the view of women who are on the street or sunbathing in bikinis has been shown in the street view, which has increased the controversy over whether or not human rights violations are excessive among Internet users. As such, as social video services that can be usefully utilized in real life are developed, it is expected that image big data exposed to personal information will also increase exponentially.

In addition, in the process of searching blogs on the Internet, uploaded pictures or video clips sometimes include the face or some body part of a person who does not wish to be exposed. In other words, the Internet is also causing dysfunction by providing users with indiscriminately exposed video contents that require social control. Video contents that contain personal information and exposed human body parts or harmful contents are easily exposed and distributed to children and adolescents who lack judgment and restraint without any sanctions. It is a big problem. Furthermore, if various video contents, including exposed human body parts, are made available to the public, the mental damage and side effects to the parties concerned are expected to be beyond imagination.

Nowadays, some websites or mass media have access to articles on the damage of video content that exposes personal information, so the side effects of this do not seem to be apparent. However, not long after, the existing mega-class Internet will bring a quantum jump to speed and it will be replaced by a Giga-class Internet with a different communication life. Video content will also explode. Therefore, it is clearly expected that this kind of various harmful video contents will overflow in the near future because video data can be viewed in a much more natural and faster environment than now.

Accordingly, mosaic blocks are frequently used nowadays to blur out such areas [18]. In particular, particular attention is paid to induce the generated mosaic blocks to be harmonized with the surrounding blocks. In addition, in the image processing fields related to detection of harmful and adult images [19, 20] and in the real application fields related to computer vision and pattern recognition, to effectively determine whether or not the given input image includes exposed body parts not preferred by users, an attempt is being made to detect the image parts processed with grid-type mosaic blocks [21].

The existing studies associated with the field where mosaic-processed blocks are automatically detected from such digital image data can be found in the related references. In [22], the edge was detected from the input image, and the candidate mosaic blocks were generated based on the detected edge. Then, the actual mosaic blocks were finally selected through the filtering process that uses geometrical features. In [23], to accurately detect mosaic blocks, the fuzzy c-means clustering algorithm was used. As far as this method is concerned, the Sobel edge was detected from the input image, the clustering features were extracted, and the fuzzy c-means clustering was applied to distinguish between general image blocks and grid-type mosaic blocks. In [24], the edge was detected from the image, and the template matching was used to detect the candidate mosaic domains. Finally, a support vector machine (SVM) was used to select the final actual mosaic blocks. In [25], a method is proposed that analyzes mosaic features in detail using a template matching-based method and then detects mosaics in the spatial domain. This method employs two strategies to speed up the detection of mosaics: a new template matching strategy and preprocessing of edge

images. In addition to such methods, many methods are used to effectively detect mosaic blocks [26].

Although the existing approaches described above are capable of extracting mosaic-processed blocks to a certain extent, they are not sufficient in terms of accuracy. To this day, studies related to algorithms generating and detecting mosaic blocks are relatively not actively conducted. Accordingly, in this paper, a new algorithm that can be used to effectively detect grid-type mosaic blocks based on the Canny edge projection was proposed. In this paper, the mosaic blocks used to blur out particular human body parts in harmful or adult images were selected as the main detection targets. In general, although the mosaic blocks used in harmful images may vary in size depending on the block to be blurred out, such mosaic blocks tend to be based on a simple format rather than a complicated format. Although there are harmful images that include blurring or virtual object insertion instead of grid-type mosaic blocks, most harmful images tend to use mosaic blocks to this day.

The methods introduced in this study have the following main contributions compared to other existing methods: First, the proposed method detected edges from color images being entered and then newly defined an edge that is connected continuously for more than a certain length in horizontal and vertical directions as a line edge. The line edge introduced in this paper can be usefully used as a new feature in object detection and image segmentation. Second, the proposed method was to project a line edge and then compare the frequency of the projected edges to effectively detect the boundaries of the blocks forming the mosaic. Third, the proposed algorithm improves the accuracy of final mosaic area detection by extracting geometric features such as size, aspect ratio, and density and robustly filtering candidate areas of mosaic detected in the previous phase.

Figure 1 shows the overall flow chart of the mosaic detection algorithm proposed in this paper. As shown in Figure 1, the proposed algorithm initially detects the Canny edge features from the input color image, then extracts the line edge over a certain length. Candidate regions of the mosaic are then detected based on horizontal and vertical line edge projection. Subsequently, the final actual grid-type mosaic regions are detected by removing the non-mosaic region through robust filtering of candidate mosaic regions using geometrical features such as size, aspect ratio, and compactness.
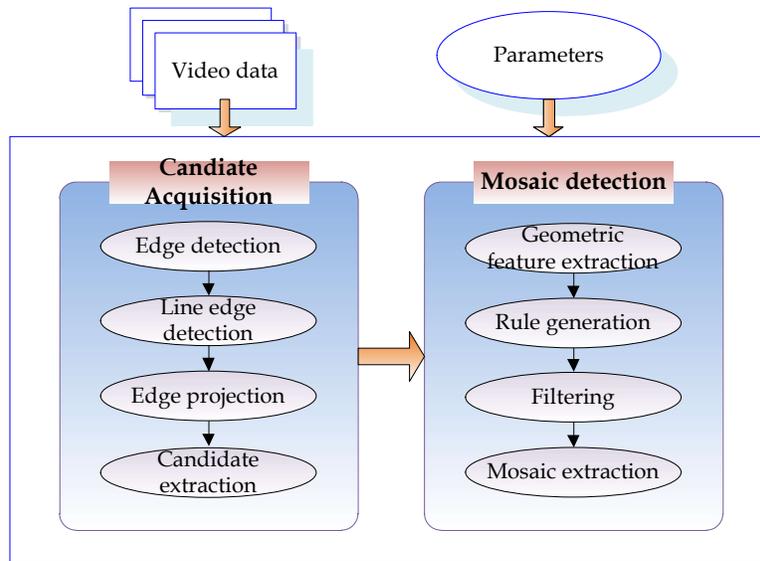
Fig. 1. Overall flowchart of the proposed method.

In this paper, we robustly detect the mosaic area in blocks to effectively protect specific parts of the human body exposed from the input images. This is to accurately detect images harmful to adolescents or children by detecting grid-type mosaic areas. Therefore, the proposed system does not recognize the detected mosaic region, nor does it remove the detected mosaic or generate a new mosaic. In other words, the suggested algorithm extracts a grid-type mosaic feature that can be used to detect harmful images along with color, edge, skin region, texture, contour, and shape features.

The remaining chapters of this paper are constructed as follows. In Chapter 2, diverse existing studies conducted in related to the mosaic detection methods used in the image processing and video analysis fields are described. In Chapter 3, the method for extracting the candidate mosaic blocks through extracting the edge and using the horizontal and vertical projections of the extracted edge is described. In Chapter 4, the method that uses geometrical features to remove the non-mosaic blocks from the candidate mosaic blocks and filtering the actual mosaic blocks is described. In Chapter 5, the results of an experiment conducted to compare and evaluate the performance of the proposed mosaic detection method are described. In Chapter 6, the conclusion of this paper is described and the future research direction is described as well.

2.   Related Work

As information communication technologies, miniaturized displays, and inexpensive storage devices make progress, the volume of available social multimedia data is increasing exponentially [27-33]. However, such video data usually include valuable

information, but also include information not preferred being exposed to other people such as the face or human body part of other people. Accordingly, to determine whether or not given video data include information not preferred by users, an attempt is being made to detect the image parts processed with grid-type mosaic blocks. Therefore, in recent years, the need for a technique for robustly detecting such mosaic regions and effectively filtering out the detected mosaic areas has been raised. Various existing methods related to the detection of grid-type mosaics, which are introduced in related documents, are as follows.

As far as the method using the edge [22] is concerned, to accurately identify the harmfulness of the input color image, a technique detecting the grid-type mosaic blocks in an adult image was proposed. In this method, to detect the grid-type mosaic blocks mainly used in adult video content, an algorithm was conducted in two major phases: a phase where the candidate mosaic blocks were detected and a phase where the candidate mosaic blocks were verified. Initially, in the phase where the candidate mosaic blocks were detected, the edge was detected from the input color image, a mosaic map was generated and binarized based on the detected edge image, and the final candidate mosaic blocks were detected. In the phase where the candidate mosaic blocks were verified, the mosaic map value of the candidate mosaic blocks and the geometrical features of the candidate mosaic blocks were used to detect the final mosaic blocks. The experimental results of this method showed that the proposed method is generally superior to the adult image, but shows slightly higher false detection on non-mosaic images. Most of the candidate regions using mosaic maps included mosaics, but false or non-detection occurred somewhat because a verification method using a simple threshold was used. Further research is needed on the part of the verification of candidate areas. The overall performance of the suggested edge-based mosaic detection algorithm was not bad. Figure 2 shows the overall flow diagram of the edge map-based mosaic detection approach.
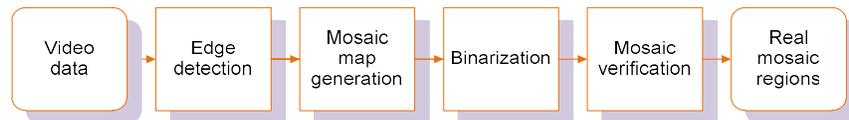
Fig. 2. Overall flowchart of the edge-based method.

As far as the method using the clustering [23] is concerned, a new mosaic block detection method that detects mosaic blocks based on the fuzzy c-means clustering was proposed. Image recognition techniques are commonly applied to detect degraded video in television viewing systems. Mosaic signals, which degrade the quality of the video, occur easily in TV signals. This approach consisted of three major phases. In the first phase, the input video is captured by the digital TV receiver and decoded into video frames. The input image is then converted into the gray-scale image and the Sobel edge detection algorithm was applied to the input color image to acquire the edge image. In the second phase, by using the extracted edge as the target, the clustering features of the fuzzy c-means clustering algorithm were extracted. In the Sobel edge detection algorithm, an adaptive thresholding is used to convert gray-scale images to binary images. Subsequently, features for the clustering algorithm are extracted by the direct

selection method. In the third phase, the fuzzy c-means clustering was applied, and the mosaic blocks were distinguished from all general image blocks. In other words, according to the classification result, mosaic blocks of the image are detected. If the number of mosaic blocks is higher than the predefined threshold, the image is considered a frame with mosaic blocks. Based on the experimental results of this method, it was specified that the algorithm used in this method was able to distinguish the actual mosaic blocks relatively well and that the incorrect detection rate was comparatively low. In particular, the existing mosaic detection method has a disadvantage that it is not possible to distinguish between Chinese characters and actual mosaic, but the proposed method detects the actual mosaic area more robustly. Figure 3 shows the overall flowchart of the fuzzy c-means clustering-based mosaic block detection method.
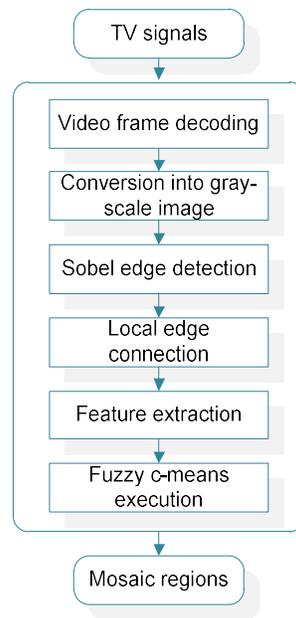
Fig. 3. Overall flowchart of the clustering-based method.

As far as the method using the macroblock [24] is concerned, a macroblock format was utilized to detect mosaic blocks. In general, as far as a digital video is concerned, mosaic blocks are expressed as one of the general phenomena resulting from video defects and are what lowers the quality of a video. Accordingly, restoring a video by detecting mosaic blocks is gradually becoming more important. In general, since a macroblock is the smallest synchronization unit used in video images, mosaic regions occur in the unit of macroblock. In this paper, a mosaic block detection method based on the macroblock's solid edge detection was proposed. The edge of an image is a collection of image pixels. The gray-scale of a pixel often varies greatly, and such collections of pixels are often closed curves. The curve concentrates most of the

information of the image. Therefore, extracting image edges is very important for recognizing and understanding the whole image. For this purpose, initially, the input color image was processed based on the Canny edge detection algorithm, and the edge image was acquired. Secondly, the binarized edge image was traversed based on the macroblock, and the solid edge of the macroblock was determined. In this study, solid edges are defined as clear edges in macroblocks and can be acted as the determination of mosaic blocks. In this method, depending on the edge detection result, the edge type of the macro block can be divided into five types. Finally, the mosaic block was determined based on the number of all the macroblocks' solid edges. In other words, if the number of mosaic macro blocks of a video frame is greater than the number threshold of macro blocks, this frame is indicated as a mosaic frame. The experimental results of this method showed the good performance of the proposed algorithm. Compared to the video mosaic detection method based on support vector machine [34, 35] and template matching [36, 37], the complexity is low, thus it is faster and does not require data training. Besides, the method is simple and efficient.

The quality of a video may be lowered by physical issues such as repeated projection, low-quality compression, decompression and defective chemical degradation of original recording data. It is gradually becoming more important to use broad-based digital media application programs to find a quality-degraded video. One general video defect is mosaic blocks where an original image information loss occurs due to a few squares combined together. As far as the template matching-based method [25] is concerned, a method for detecting mosaic defects in spatial and edge domains was proposed after specifically analyzing the characteristics of a mosaic defect. In this method, four even squares were initially detected, and the intersection points of the squares were selected as the mosaic macroblock (MMB)'s features. The mosaic consisted of various MMBs. Meanwhile, an automatic detection algorithm prompt and effective in an edge domain was proposed by analyzing the existing approach method. In this method, to increase the mosaic detection speed, two new tricks were selected: a new template matching strategy and the edge image pre-processing. The experimental results of this method showed the outstanding performance of the proposed algorithm. In the future, the idea of the proposed mosaic detection approach can be applied broadly to solve other problems for degraded video defect detection, such as digital dropout, betacam dropout, and so on.

In addition to the various methods described above, new methods related to mosaic detection are being continuously introduced [26]. Although the existing algorithms described above are capable of accurately detecting mosaic-processed blocks to a certain extent, they are relatively not sufficient in terms of accuracy. Also, existing methods have various constraints on the photographing environment or condition of the color image and the grid mosaic area. To this day, new studies related to image processing and pattern recognition algorithms that generate and detect grid-type mosaics to blur out target objects exposing personal information are not as actively conducted as other studies.

3.   Extraction of Candidate Mosaic Blocks

The grid-type mosaic areas included in the input color image consist of mosaic blocks (MB). In general, such mosaic blocks demonstrate the following four characteristics. Initially, the pixels located within a mosaic block share the same color. Accordingly, no edge exists within a mosaic block. Secondly, mosaic blocks share the same size. In other words, both the horizontal length and the vertical length of the mosaic block are the same. Thirdly, mosaic blocks are adjacent to one another and form a cluster. Fourthly, no edge exists among adjacent mosaic blocks sharing the same gray value.

In this paper, initially, the Canny edge CE(x, y) was extracted from the input image [38-42]. As far as a color image is concerned, since an edge is one of the important features for detecting mosaics, as the initial phase, to accurately extract an edge image is very important. In general, since the Canny edge demonstrates comparatively high accuracy in terms of finding the contour line, it is one of the most used edge detectors in the image processing and computer vision fields.

Canny edge extraction consists of four major steps. First, a Gaussian smoothing is applied to the input color image. Second, the Sobel operator is applied to the smoothed result image to obtain the edge strength map and the edge direction map. Third, non-maximum suppression is applied to create a thin-thickness edge map. Fourth, hysteresis thresholding is applied to remove false positives. In general, Canny edge extraction has the advantage of accurately detecting edges, while the algorithm is complicated and takes some time to execute. The horizontal and vertical masks used for Canny edge extraction in this paper are shown in Figure 4.

- 1	0	1	1	2	1
- 2	0	2	0	0	0
- 1	0	1	- 1	- 2	1

(a) Horizontal mask   (b) Vertical mask

Fig. 4. Horizontal and vertical mask.

The proposed system detects edges connected continuously over TH_{line} pixels in the horizontal direction (0° or 180°) from the extracted canny edge, which are called horizontal line edge $LE_h(x, y)$ in this paper. Besides, edges continuously connected over TH_{line} pixels in the vertical direction (90° or 270°) are detected and named as vertical line edge $LE_v(x, y)$.

Then, as shown in (1), the vertical line edge $LE_v(x, y)$ is projected in the direction of the x axis. A set of x coordinates in which Proj(x), which is the number of pixels accumulated through projection, is equal to or greater than TH_{accum} is obtained as shown in (3).

$$Proj(x) = \sum_{x=0}^{W} LE_v(x, y) \tag{1}$$

$$Proj(y) = \sum_{y=0}^{H} LE_h(x, y) \tag{2}$$

Similarly, as shown in (2), the horizontal line edge $LE_h(x, y)$ is projected in the y-axis direction. Then, a set of y coordinates in which $Proj(y)$, which is the number of pixels accumulated through the projection, is equal to or greater than TH_{accum} is acquired as shown in (4). In (1) and (2), W and H represent the width and height of the input image.

$$X_{accum} = \{x_1, x_2, ..., x_m\} \tag{3}$$

$$Y_{accum} = \{y_1, y_2, ..., y_n\} \tag{4}$$

X_{diff}, the difference among the adjacent x coordinates in group X_{accum}, and Y_{diff}, the difference among the adjacent y coordinates in group X_{accum} are calculated as shown in (5) and (6), respectively. Then, of X_{diff} and Y_{diff}, the values having the highest frequency and the value having the lowest frequency are selected and determined as N and M of the mosaic blocks having a size equivalent to the N×M pixel size.

$$X_{diff} = \{x_k - x_{k-1} \mid x_k \in X_{accum}, \ k = 2, 3, 4, ..., m\} \tag{5}$$

$$Y_{diff} = \{y_k - y_{k-1} \mid y_k \in Y_{accum}, \ k = 2, 3, 4, ..., n\} \tag{6}$$

We select x coordinates where the difference between adjacent x coordinates in the set X_{accum} is N, and y coordinates where the difference between adjacent y coordinates in Y_{accum} is M, as shown in (7) and (8).

$$X_{mosaic} = \{x_k \mid x_k \in X_{accum}, (x_k - x_{k-1}) == N\} \tag{7}$$

$$Y_{mosaic} = \{y_k \mid y_k \in Y_{accum}, (y_k - y_{k-1}) == M\} \tag{8}$$

The minimum enclosing rectangle (MER) [43, 44] regarding the domain including all the locations (x, y) having a value included in X_{mosaic} as the input image's x coordinate and a value included in Y_{mosaic} as the input image's y coordinate are selected as shown in (9), and the defined MER is selected as the candidate mosaic block to be detected in this paper. In (9), (x_p, y_p) and (x_q, y_q) refer to the initial and final coordinates of the MER, respectively. The candidate regions of the grid-type mosaic defined as in (9) are used as the input data of the actual mosaic detection step, which is the next step.

$$MER_{mosaic} = \{x_p, y_p, x_q, y_q\} \tag{9}$$

Figure 5 shows the main process flow diagram for extracting the candidate mosaic blocks mentioned above.

```
┌──────────────────┐
│   Input image    │
└──────────────────┘
         │
┌──────────────────┐        ┌──────────────────┐
│   Canny edge     │◄──────►│    CE(x, y)      │
│   extraction     │        └──────────────────┘
└──────────────────┘
         │
┌──────────────────┐        ┌──────────────────┐
│  Extraction of   │◄──────►│   LE_h(x, y)     │
│horizontal line edge│      └──────────────────┘
└──────────────────┘
         │
┌──────────────────┐        ┌──────────────────┐
│  Extraction of   │◄──────►│   LE_v(x, y)     │
│ vertical line edge│       └──────────────────┘
└──────────────────┘
         │
┌──────────────────┐        ┌──────────────────┐
│  Projection of   │◄──────►│    Proj(x)       │
│ vertical line edge│       └──────────────────┘
└──────────────────┘
         │
┌──────────────────┐        ┌──────────────────┐
│  Projection of   │◄──────►│    Proj(y)       │
│horizontal line edge│      └──────────────────┘
└──────────────────┘
```

$Proj(x) > TH_{accum}$ ◄──────► X_{accum}

$Proj(y) > Th_{accum}$ ◄──────► Y_{accum}

$x_k - x_{k-1}$ ◄──────► X_{diff}

$y_k - y_{k-1}$ ◄──────► Y_{diff}

$x_k - x_{k-1} == N$ ◄──────► X_{mosaic}

$y_k - y_{k-1} == M$ ◄──────► Y_{mosaic}

$\{x_p, y_p, x_q, y_q\}$ ◄──────► MER_{mosaic}

```
┌──────────────────┐
│ Candidate mosaic │
│     regions      │
└──────────────────┘
```
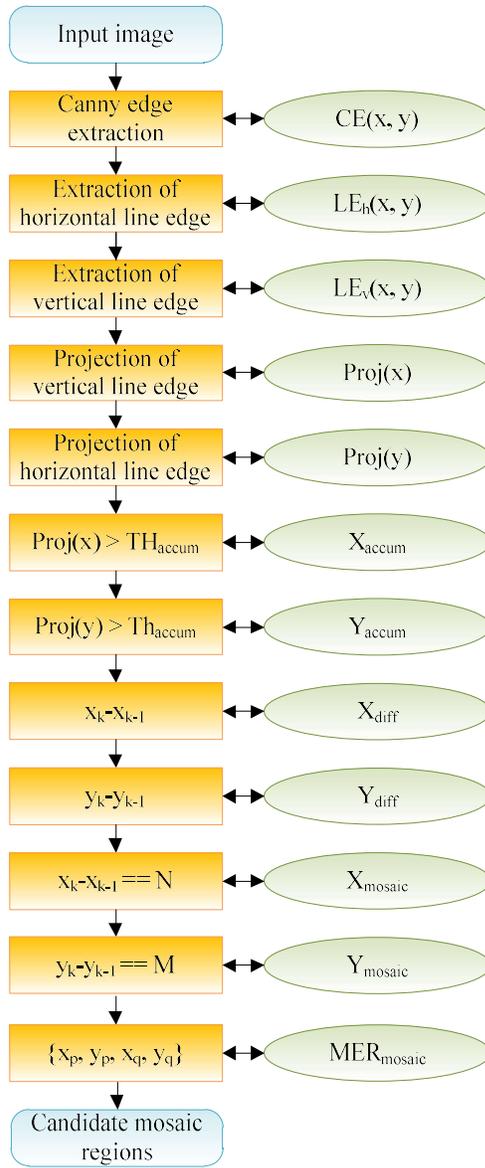
Fig. 5. Process flow diagram of detecting candidate mosaic areas.

4. Mosaic Detection through Filtering

In this paper, the geometrical features [45-50] are used to remove the candidate blocks determined to be non-mosaic blocks among the detected candidate mosaic blocks, and only actual mosaic regions are selected. The geometrical features F_{model} used for removing the candidate mosaic blocks are the size feature size(R_i), the aspect ratio feature aspect(R_i) and the compactness feature compact(R_i) of the candidate mosaic blocks, and such features are defined and utilized as shown in equations (10) to (13).

$$F_{model} = \{size(R_i),\, aspect(R_i),\, compact(R_i)\} \tag{10}$$

$$size(R_i) = \frac{Num(R_i)}{W \times H} \tag{11}$$

$$aspect(R_i) = \frac{MER_h(R_i)}{MER_w(R_i)} \tag{12}$$

$$compact(R_i) = \frac{Num(R_i)}{MER_w(R_i) \times MER_h(R_i)} \tag{13}$$

In equations (10) to (13), F_{model} represents a feature model for the mosaic region that contains three geometrical features. R_i refers to the i-th region among the candidate mosaic regions extracted in the previous phase. W and H represent the horizontal length and the vertical length of the input color image, respectively. Num(R_i) refers to the total number of pixels included within the candidate mosaic block R_i. Besides, $MER_w(R_i)$ and $MER_h(R_i)$ refer to the horizontal length and the vertical length, respectively, of the minimum enclosing rectangle included for candidate mosaic block R_i.

Initially, the candidate block's size feature size(R_i) is used to effectively remove the too small candidate mosaic blocks. Usually, a candidate mosaic region having a small size is extracted when an unexpected noise is inserted in an input color image or when a small-sized object having line edges exists in the color image [51-54]. In other words, an area that is too small is not considered a grid-type mosaic. Mosaic blocks are usually used to mask areas of the exposed body in harmful images. In particular, it produces mosaic blocks of larger size rather than precisely fitting to the exposed body region. Therefore, candidate regions of relatively small size are removed from the candidate group of mosaics.

The aspect ratio feature aspect(R_i) is used to effectively remove the non-mosaic blocks that too excessively lean toward one direction. Although grid-type mosaic blocks too excessively extended towards a particular direction may exist, such blocks are not frequently found in actual harmful or adult images. In other words, most of the exposed areas of the human body parts in the harmful image are located in the center of the input color image, and the shape of the corresponding area exists in a constant ratio in the horizontal direction and the vertical direction rather than in one direction.

The compact(R_i) is used to remove the non-mosaic blocks having low compactness within the block. Usually, since there is no hole inside the grid-type mosaic area, the density of the region is high. Therefore, in the proposed method, when the density feature of the candidate mosaic region is relatively low, the candidate region is considered as the non-mosaic region and removed from the candidate mosaic group.

In this paper, as shown in (14), (15) and (16), the geometrical features defined above are applied to the candidate mosaic blocks, and the blocks having a threshold lower than the predefined threshold [55-58] are determined to be non-mosaic blocks and are removed from the candidate blocks. In (14), (15) and (16), TH_{size}, TH_{aspect}, and $TH_{compact}$ refer to the predefined threshold size, threshold aspect ratio and threshold compactness of the candidate mosaic block, respectively. In this paper, such threshold values are artificially defined in advance through the repeated experiments.

$$IF\ (size(R_i) < TH_{size})\ THEN$$
$$\quad eliminate\ R_i$$
$$ELSE \tag{14}$$
$$\quad select\ R_i\ as\ a\ mosaic\ region$$

$$IF\ (acpect(R_i) < TH_{aspect})\ THEN$$
$$\quad eliminate\ R_i$$
$$ELSE \tag{15}$$
$$\quad select\ R_i\ as\ a\ mosaic\ region$$

$$IF\ (compact(R_i) < TH_{compact})\ THEN$$
$$\quad eliminate\ R_i$$
$$ELSE \tag{16}$$
$$\quad select\ R_i\ as\ a\ mosaic\ region$$

In this paper, the three features defined above are applied to the candidate mosaic blocks to remove all the blocks determined to be non-mosaic blocks, and the remaining blocks are finally determined to be the actual mosaic blocks.

Although it is possible that general mosaic blocks, for example, mosaic blocks used to blur out particular buildings or signs, can be removed due to geometrical features such as aspect ratio and size, since mosaic blocks that blur out body parts are set as the main detection targets in this paper, the geometrical features used in this paper are effectively operated. Figure 6 shows the overall flow of the approach for filtering candidate mosaic areas.

In the future, to further improve the performance of the proposed system, in addition to the features used in this paper, it is necessary to additionally define new features that can represent grid-type mosaics such as texture and smoothness features. There is also a need for an algorithm that accurately and quickly extracts the newly defined features from the input image. In particular, the proposed system should define the weights that indicate the importance of each feature, and it is also necessary to improve the performance of the proposed mosaic detection system by adaptively and automatically adjusting the weights according to the environment in which the image is taken.
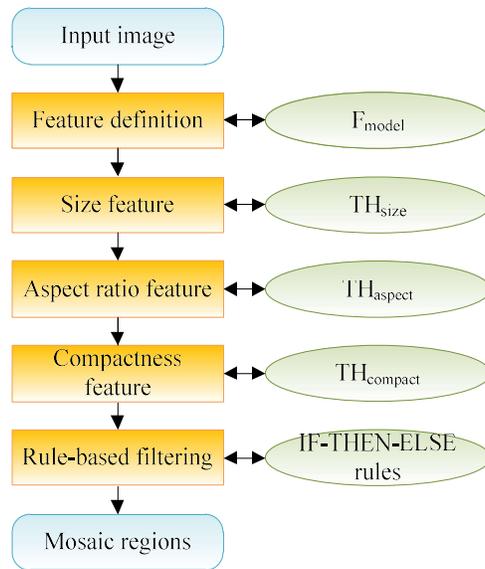
Fig. 6. Overall flow of the candidate filtering process.

5.   Experimental Results

In this paper, the personal computer used for experiments consisted of Intel Core(TM) i7 2.83 GHz CPU and 8GB main memory, and is equipped with Windows 7 operating system. Microsoft's Visual Studio and OpenCV open-source computer vision library are used to implement the proposed algorithm. To comparatively evaluate the performance of the algorithm proposed in this paper, various types of indoor and outdoor input color images including grid-type mosaic blocks are collected and utilized. Such images are captured in diverse natural indoor and outdoor environments where no particular restrictions are set. Figure 7 shows an example of the input color images with grid-type mosaic regions.
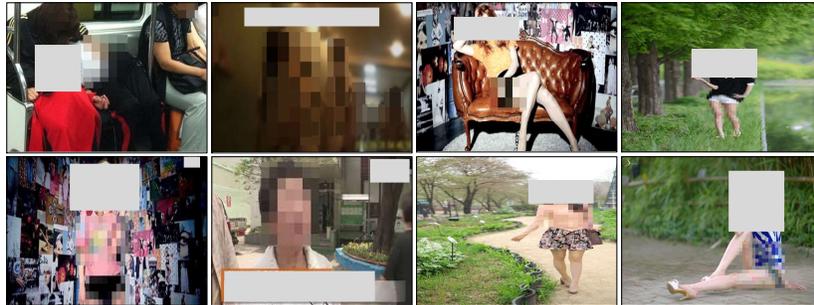
Fig. 7. Example of input images with mosaic.

Figure 8 (a) shows an input color image in which mosaic blocks exist, and Figure 8 (b) shows the result of the Canny edge features extracted from an input image. As shown in Figure 8 (a), there is a grid-type mosaic area near the center of the input image. In Figure 8 (b), it can be seen that the line-shaped edges are concentrated in the center of the image. As expected, Figure 8 shows that the Canny edge represents the object's appearance information relatively accurately.

(a) Input image   (b) Edge image
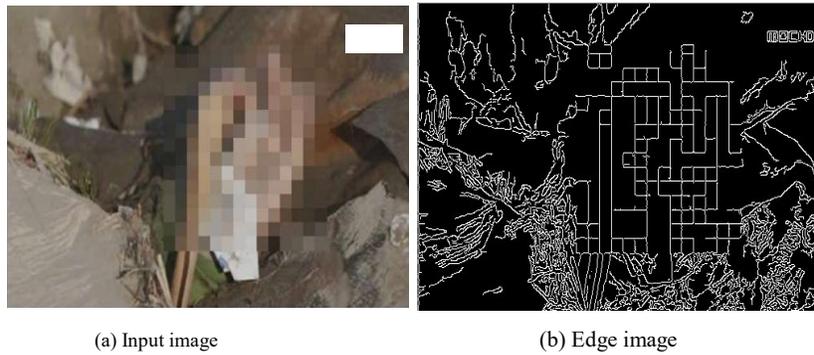
Fig. 8. Input image and edges.

Figure 9 (a) shows the histogram of horizontally projected line edges, and Figure 9 (b) shows the histogram of vertically projected line edges. As shown in Figure 9, it can be seen that the frequency of the histogram of the line edges is relatively high in the region where the grid-type mosaic blocks exist.
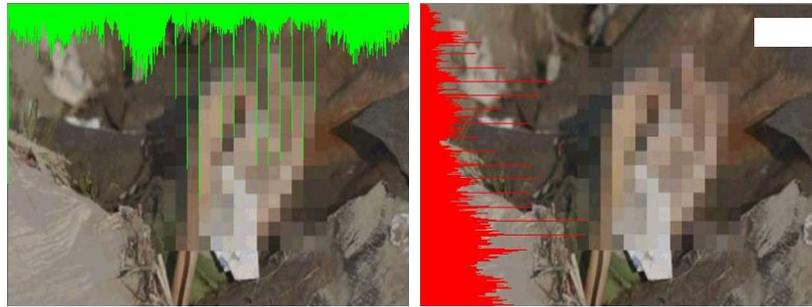
(a) Histogram of horizontally projected edges (b) Histogram of vertically projected edges

Fig. 9. Histogram of projected line edges.

Figure 10 shows the result of the actual mosaic blocks remaining after removing non-mosaic blocks from the candidate mosaic blocks through the application of geometrical features.
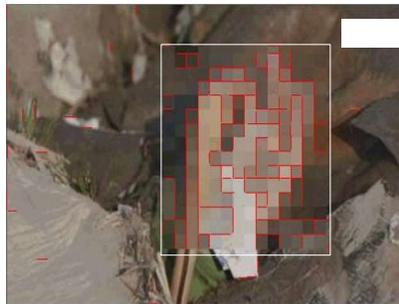
Fig. 10. Detected mosaic region.

In this paper, we used an accuracy measure defined as (17) and (18) to quantitatively evaluate the performance of the proposed mosaic detection algorithm [59-63]. In (17) and (18), N_{TP} indicates the number of mosaic regions accurately detected, N_{FP} represents the number of regions that are incorrectly detected as mosaic regions but not mosaic regions, and N_{FN} denotes the number of mosaic regions that are not detected. In (17) and (18), $R_{precision}$ represents a relative ratio of mosaic regions accurately detected among the entire mosaic regions detected from the input color image, and R_{recall} is a relative ratio of mosaic regions accurately detected among all mosaic regions present in the input image.

$$R_{precision} = \frac{N_{TP}}{N_{TP} + N_{FP}} \tag{17}$$

$$R_{recall} = \frac{N_{TP}}{N_{TP} + N_{FN}} \tag{18}$$

In this paper, the performance of the proposed method is compared and evaluated in terms of accuracy with that of the existing template matching method. Figures 11 and 12 show graphically the measurement results of the accuracy of the two mosaic detection algorithms obtained from (17) and (18). As shown in Figure 11 and Figure 12, the proposed feature projection-based method reduces false detection of the mosaic area, so it can be confirmed that it detects mosaic areas from an image more robustly.

As shown in Figures 11 and 12, it was confirmed that the proposed algorithm used the horizontal and vertical edge projections to detect grid-type mosaic blocks more accurately than the existing method. However, in the case where mosaic blocks are generated in the parts of the input image in which the image quality is degraded, it is possible that the detection accuracy of the proposed method may be decreased. In addition, in this paper, a line edge was used to decrease any error caused by a noise edge. Accordingly, although a noise line edge having a size below a particular size was not automatically considered, any error caused by a noise line edge having a size above that particular size was unavoidable. To solve this issue, it is necessary to either conduct pre-processing such as image smoothing to decrease the noise to the utmost extent possible or adaptively adjust the size of TH_{line}, the threshold determining the line edge, according to the noise inclusion level.
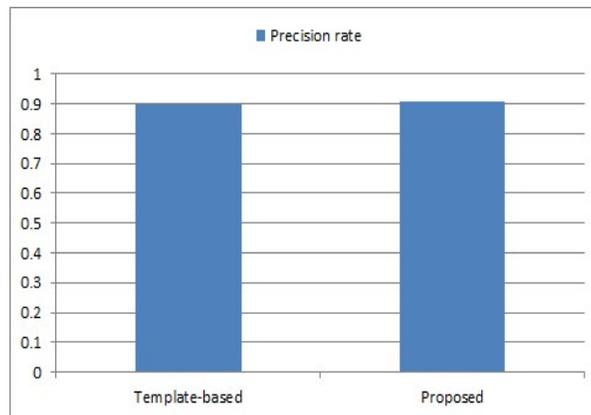
Fig. 11. Precision rates.
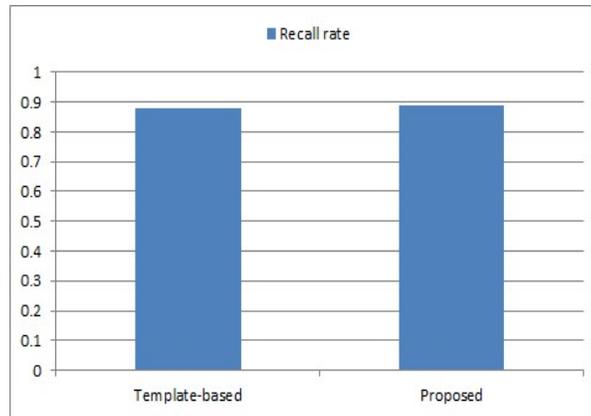
Fig. 10. Recall rates.

Besides, in this paper, as described above, harmful images were selected as the main targets for detecting mosaic regions. Namely, since the images used in the experiments were mostly images that explicitly reveal human body parts or show sexual intercourse performed between a male and a female, it is difficult to include such images as the contents of this paper. Therefore, we selected color images that do not contain these body parts and inserted the input and resulting images into the paper.

6. Conclusion

Recently, as low-priced high-quality cameras are developed and information communication technology makes rapid progress, the types of available video data are diversified. However, such video data sometimes include information not preferred to be exposed to other people such as the face, resident registration number and exposed body part of a person. Accordingly, in the process of uploading a photo or a video on websites or blogs, the face of a person or particular objects are frequently processed with mosaic blocks to protect the portrait rights of a person or to prevent other people from being disgusted. In recent years, there is an increasing need for a new research for effectively detecting a target object region including personal information in an image and blocking the detected object.

In this paper, a new algorithm that more robustly detects the grid-type mosaic blocks from an input image based on the horizontal and vertical line edge projections was proposed. As far as the proposed method is concerned, initially, the Canny edge was extracted from a color image, then the line edges consecutively connected in the horizontal and vertical directions were extracted. Subsequently, the edges in the horizontal and vertical directions were projected, the frequency of the projected edges was calculated, and the candidate mosaic blocks were detected. Subsequently, geometrical features such as size, aspect ratio, and compactness of the detected candidate blocks were used to filter the candidate blocks. Accordingly, non-mosaic

blocks were effectively removed from the candidate mosaic blocks, and the actual mosaic blocks were accurately extracted. In the paper, the experiment results showed that the proposed line edge projection-based method detected grid-type mosaic regions from various types of input color images more robustly than other existing methods.

The proposed grid-type mosaic detection algorithm based on line edge feature projection is expected to be very useful for various practical applications related to video contents such as personal information blocking, video data security, image restoration and post-processing, and harmful image detection. Besides, the proposed grid-type mosaic detection algorithm can be linked to the process of recognizing the detected mosaic region and the procedure of removing the detected mosaic.

In the future, various parameters such as the threshold used in the grid-type mosaic detection algorithm proposed in this paper will be repeatedly tested and adaptively adjusted to enhance the efficiency of the current mosaic detection algorithm. Besides, artificial intelligence-based learning algorithms such as the deep new learning network frequently used nowadays will be used to systematically learn the line edge-based features extracted in this paper and enhance the mosaic block detection accuracy. Also, still and dynamic color images captured in more diverse indoor and outdoor environments will be applied to the proposed algorithm to enhance the robustness of the proposed algorithm.

References

1. Ding, S., Qu, S., Xi, Y., Wan, S.: A Long Video Caption Generation Algorithm for Big Video Data Retrieval. Future Generation Computer Systems, Vol. 93, 583-595. (2019)
2. Rim, D., Hasan, M. K., Puech, F., Pal, C. J.: Learning from Weakly Labeled Faces and Video in the Wild. Pattern Recognition, Vol. 48, No. 3, 759-771. (2015)
3. Lee, J.-E.: Robust Influenza Analysis Algorithm Based on Image Processing under Varying Radiometric Conditions. Journal of the Korea Academia-Industrial Cooperation Society, Vol. 20, No. 7, 127-132. (2019)
4. Cao, Z., Huang, Q., Wu, C.Q.: Maximize Concurrent Data Flows in Multi-radio Multi-channel Wireless Mesh Networks. Computer Science and Information Systems, Vol. 17, No. 3, 759–777. (2020)
5   Sun, Z., Lv, Z., Hou, Y., Xu, C., Yan, B.: MR-DFM: A Multi-path Routing Algorithm Based on Data Fusion Mechanism in Sensor Networks. Computer Science and Information Systems, Vol. 16, No. 3, 867–890. (2019)
6. Yin, Y., Zhang, W., Xu, Y., Zhang, H., Mai, Z., Yu, L.: QoS Prediction for Mobile Edge Service Recommendation with Auto-encoder. IEEE Access, Vol. 7, pp. 62312-62324. (2019)
7. Yin, Y., Xu, W., Xu, Y., Li, H., Yu, L.: Collaborative QoS Prediction for Mobile Service with Data Filtering and SlopeOne Model. Mobile Information Systems, vol. 2017, 1-14. (2017)
8. Yin, Y., Chen, L., Xu, Y., Wan, J., Zhang, H., Mai, Z.: QoS Prediction for Service Recommendation with Deep Feature Learning in Edge Computing Environment. Mobile Networks and Applications, Vol. 2019, 1-11. (2019)

9. Yin, Y., Aihua, S., Min, G., Xu, Y., Shuoping, W.: QoS Prediction for Web Service Recommendation with Network Location-Aware Neighbor Selection. International Journal of Software Engineering and Knowledge Engineering, Vol. 26, No. 4, 611-632. (2016)

10. Yu, J., Kuang, Z., Zhang, B., Zhang, W., Lin, D.: Leveraging Content Sensitiveness and User Trustworthiness to Recommend Fine-Grained Privacy Settings for Social Image Sharing. IEEE Transactions on Information Forensics and Security, Vol. 13, No. 5, 1317-1332. (2018)

11. Liu, W., Gao, Y., Ma, H., Yu, S., Nie, J.: Online Multi-Objective Optimization for Live Video Forwarding across Video Data Centers. Journal of Visual Communication and Image Representation, Vol. 48, 502-513. (2017)

12. Kim, D.-J.: A Systematic Study of Computer-Based Driving Intervention Program for Elderly Drivers. Journal of the Korea Academia-Industrial Cooperation Society, Vol. 20, No. 4, 293-302. (2019)

13. Lee, C.-H., Choi, H.-I.: Algorithm for Improving the Position of Vanishing Point Using Multiple Images and Homography Matrix. Journal of the Korea Academia-Industrial Cooperation Society, Vol. 20, No. 1, 477-483. (2019)

14. Lee, J.-S., Oh, M.-K.: Distribution Analysis of Land Surface Temperature about Seoul Using Landsat 8 Satellite Images and AWS Data. Journal of the Korea Academia-Industrial Cooperation Society, Vol. 20, No. 1, 434-439. (2019)

15. Wu, D., Zhang, F., Wang, H., Wang, R.: Fundamental Relationship between Node Dynamic and Content Cooperative Transmission in Mobile Multimedia Communications. Computer Communications, Vol. 120, 71-79. (2018)

16. Yin, L., Cheng, Q., Wang, Z., Shao, Z.: Big Data for Pedestrian Volume: Exploring the Use of Google Street View Images for Pedestrian Counts. Applied Geography, Vol. 63, 337-345. (2015)

17. Flores, A., Belongie, S.: Removing Pedestrians from Google Street View Images. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. San Francisco, USA, 53-58. (2010)

18. Guo, D., Tang, J., Cui, Y., Ding, J., Zhao, C.: Saliency-based Content-Aware Lifestyle Image Mosaics. Journal of Visual Communication and Image Representation, Vol. 26, 192-199. (2015)

19. Jang, S.-W., Jung, M.: Detection of Harmful Content Using Multilevel Verification in Visual Sensor Data. Wireless Personal Communications, Vol. 86, No. 01, 109-124. (2016)

20. Cheng, F., Wang, S.-L., Wang, X.-Z., Liew, A. W.-C., Liu, G.-S.: A Global and Local Context Integration DCNN for Adult Image Classification. Pattern Recognition, Vol. 96, 1-12, (2019)

21. Jang, S.-W., Lee, S.-H.: Robust Blocking of Human Faces with Personal Information Using Artificial Deep Neural Computing. Sustainability, Vol. 12, No. 6, 1-20. (2020)

22. Park, Y.-J., Lee, J.-W., Choi, G.-S., Park, J.-J.: A Study on Grid Mosaic Detection for Identifying Image Harmfulness. In Proceedings of the Korea Society of Industrial Information Systems, 1-5. (2015)

23. Liu, J., Huang, L., Lin, J.: An Image Mosaic Block Detection Method Based on Fuzzy C-Means Clustering. In Proceedings of the IEEE International Conference on Computer Research and Development. Shanghai, China, Vol. 1, 237-240. (2011)

24. Wei, Z., Lin, J., Zhang, L., Song, S.: Mosaic Defect Detection Based on Macro Block Solid Edge Detection. Research Journal of Applied Sciences, Engineering and Technology, No. 5, Vol. 13, 3549-3553. (2013)

25. Sun, S.-F., Han, S.-H., Wang, G., Xu Y.-C., Lei, B.-J.: Mosaic Defect Detection in Digital Video. In Proceedings of the IEEE Chinese Conference on Pattern Recognition. Chongqing, China, 1-5. (2010)

26. Jang, S.-W., Jung, M.: Robust Detection of Mosaic Regions in Visual Image Data. Cluster Computing, Vol. 19, No. 4, 2285-2293. (2016)

27. Yu, J., Zhang, B., Kuang, Z., Lin, D., Fan, J.: iPrivacy: Image Privacy Protection by Identifying Sensitive Objects via Deep Multi-Task Learning. IEEE Transactions on Information Forensics and Security, Vol. 12, No. 5, 1005-1016. (2017)

28. Jia, G., Han, G., Rao, H., Shu, L.: Edge Computing-Based Intelligent Manhole Cover Management System for Smart Cities. IEEE Internet of Things Journal, Vol. 5, No. 3, 1648-1656. (2018)

29. Zhou, R., Khemmarat, S., Gao, L., Wan, J., Zhang, J.: How YouTube Videos Are Discovered and Its Impact on Video Views. Multimedia Tools and Applications, Vol. 75, No. 10, 6035-6058. (2016)

30. Bassel, A., Nordin M. J., Abdulkareem, M. B.: An Invisible Image Watermarking Based on Modified Particle Swarm Optimization (PSO) Algorithm. International Journal of Security and Its Applications, Vol. 12, No. 2, 1-8. (2018)

31. Bassel, A., Nordin M. J., Abdulkareem, M. B.: An Invisible Image Watermarking Based on Modified Particle Swarm Optimization (PSO) Algorithm. International Journal of Security and Its Applications, Vol. 12, No. 2, 1-8. (2018)

32. Lu, R., Liu, X., Wang, X., Pan, J., Sun K., Waynes, H.: The Design of FPGA-Based Digital Image Processing System and Research on Algorithms. International Journal of Future Generation Communication and Networking, Vol. 10, No. 2, 41-54. (2017)

33. Zhou, P., Zhou, Y., Wu, D., Jin, H.: Differentially Private Online Learning for Cloud-Based Video Recommendation with Multimedia Big Data in Social Networks. IEEE Transactions on Multimedia, Vol. 18, No. 6, 1217-1229. (2016)

34. Ye, F., Zhang, Z., Chakrabarty, K., Gu, X.: Board-Level Functional Fault Diagnosis Using Multikernel Support Vector Machines and Incremental Learning. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, Vol. 33, No. 2, 279-290. (2014)

35. Han, B., Davis, L. S.: Density-Based Multifeature Background Subtraction with Support Vector Machine. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 34, No. 5, 1017-1023. (2012)

36. Yan, B., Xiao, L., Zhang, H., Xu, D., Zhang, Y.: An Adaptive Template Matching-Based Single Object Tracking Algorithm with Parallel Acceleration. Journal of Visual Communication and Image Representation, Vol. 64, 1-13. (2019)

37. Wang, M., Cui, Q., Sun, Y., Wang, Q.: Photovoltaic Panel Extraction from Very High-Resolution Aerial Imagery Using Region-Line Primitive Association Analysis and Template Matching. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 141, 100-111. (2018)

38. Gaurav, K., Ghanekar, U.: Image Steganography Based on Canny Edge Detection, Dilation Operator and Hybrid Coding. Journal of Information Security and Applications, Vol. 41, 41-51. (2018)

39. Shahdoosti, H. R., Rahemi, Z.: Edge-Preserving Image Denoising Using a Deep Convolutional Neural Network. Signal Processing, Vol. 159, 20-32. (2019)

40. Zhang, W. C., Zhao, Y. L., Breckon, T. P., Chen, L.: Noise Robust Image Edge Detection Based upon the Automatic Anisotropic Gaussian Kernels. Pattern Recognition, Vol. 63, 193-205. (2017)

41. Xu, Q., Varadarajan, S., Chakrabarti, C., Karam, L. J.: A Distributed Canny Edge Detector: Algorithm and FPGA Implementation. IEEE Transactions on Image Processing, Vol. 23, No. 7, 2944-2960. (2014)

42. Lee, J., Tang, H., Park, J.: Energy Efficient Canny Edge Detector for Advanced Mobile Vision Applications. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 28, No. 4, 1037-1046. (2018)

43. Jiang, N., Yang, W., Duan, L., Xu, X., Liu, Q.: Acceleration of CT Reconstruction for Wheat Tiller Inspection Based on Adaptive Minimum Enclosing Rectangle. Computers and Electronics in Agriculture, Vol. 85, 123-133. (2012)

44. Kwak, E., Habib, A.: Automatic Representation and Reconstruction of DBM from LiDAR Data Using Recursive Minimum Bounding Rectangle. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 93, 171-191. (2014)

45. Lei, Y., Zheng, L., Huang, J.: Geometric Invariant Features in the Radon Transform Domain for Near-Duplicate Image Detection. Pattern Recognition, Vol. 47, No. 11, 3630-3640. (2014)

46. Gang, L., Li, X.-H., Zhou, J.-L., Gong X.-G.: Geometric Feature-based Facial Expression Recognition Using Multiclass Support Vector Machines. In Proceedings of the IEEE International Conference on Granular Computing. Nanchang, China, 318-321. (2019)

47. Wang, L., Xu, F., Hagiwara, I.: Investigation into Registration of Scanned 3D Image Based on Geometric Feature Identification. In Proceedings of the International Conference on Computer Supported Cooperative Work in Design, 648-653. (2009)

48. Zhang, L., Zhang, D., Sun, M.-M., Chen, F.-M.: Facial Beauty Analysis Based on Geometric Feature: Toward Attractiveness Assessment Application. Expert Systems with Applications, Vol. 82, 252-265. (2017)

49. Lang, H., Wu, S.: Ship Classification in Moderate-Resolution SAR Image by Naive Geometric Features-Combined Multiple Kernel Learning. IEEE Geoscience and Remote Sensing Letters, Vol. 14, No. 10, 1765-1769. (2017)

50. Ballihi, L., Amor, B. B., Daoudi, M., Srivastava, A., Aboutajdine, D.: Boosting 3-D-Geometric Features for Efficient Face Recognition and Gender Classification. IEEE Transactions on Information Forensics and Security, Vol. 7, No. 6, 1766-1779. (2012)

51. Jin, L., Zhang, W., Ma, G., Song, E.: Learning Deep CNNs for Impulse Noise Removal in Images. Journal of Visual Communication and Image Representation, Vol. 62, 193-205. (2019)

52. Giannatou, E., Papavieros, G., Constantoudis, V., Papageorgiou, H., Gogolides E.: Deep Learning Denoising of SEM Images Towards Noise-Reduced LER Measurements. Microelectronic Engineering, Vol. 216. 1-8. (2019)

53. Chen, R., Zhao, F., Yang, C., Li, Y., Huang T.: Robust Estimation for Image Noise Based on Eigenvalue Distributions of Large Sample Covariance Matrices. Journal of Visual Communication and Image Representation, Vol. 63, 1-10. (2019)

54. Xu, J., Zhang, P., Yang, Z., Gao, Z., Nie, K., Ma, J.: Image Lag Elimination Algorithm for Pulse-Sequence-Based High-Speed Image Sensor. IEEE Access, Vol. 7, 159575-159583. (2019)

55. Bohat, V. K., Arya, K. V.: A New Heuristic for Multilevel Thresholding of Images. Expert Systems with Applications, Vol. 117, 176-203. (2019)

56. Elaziz, M. A., Lu, S.: Many-Objectives Multilevel Thresholding Image Segmentation Using Knee Evolutionary Algorithm. Expert Systems with Applications, Vol. 125, 305-316. (2019)

57. Chakraborty, F., Nandi, D., Roy, P. K.: Oppositional Symbiotic Organisms Search Optimization for Multilevel Thresholding of Color Image. Applied Soft Computing, Vol. 82, 1-19. (2019)

58. Talukder, A., Alam, M. G. R., Tran, N. H., Niyato, D., Park, G. H., Hong, C. S.: Threshold Estimation Models for Linear Threshold-Based Influential User Mining in Social Networks. IEEE Access, Vol. 7, 105441-105461. (2019)

59. Zhang, X., Feng, X., Xiao, P., He, G., Zhu, L.: Segmentation Quality Evaluation Using Region-Based Precision and Recall Measures for Remote Sensing Images. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 102, 73-84. (2015)

60. Ziolko, B.: Fuzzy Precision and Recall Measures for Audio Signals Segmentation. Fuzzy Sets and Systems, 279, 101-111. (2015)

61. Bautista-Gomez, L., Benoit, A., Cavelan, A., Raina, S. K., Sun, H.: Coping with Recall and Precision of Soft Error Detectors. Journal of Parallel and Distributed Computing, Vol. 98, 8-24. (2016)

62. Hassler, E. E., Hale, D. P., Hale, J. E.: A Comparison of Automated Training-by-Example Selection Algorithms for Evidence-Based Software Engineering. Information and Software Technology, Vol. 98, 59-73. (2018)
63. Thara, D.K., PremaSudha, B.G., Fan, X.: Epileptic Seizure Detection and Prediction Using Stacked Bidirectional Long Short Term Memory. Pattern Recognition Letters, Vol. 128, 529-535. (2019)

Seok-Woo Jang received his B.S., M.S., Ph.D. degrees in Computer Science from Soongsil University, Seoul, Korea, in 1995, 1997, and 2000, respectively. From October 2003 to January 2009, he was a Senior Researcher with the Construction Information Research Department at Korea Institute of Construction Technology (KICT), Ilsan, Korea. Since March 2009, he has been a Professor in the Department of Digital Media, Anyang University, Korea. His primary research interests include robot vision, augmented reality, video indexing and retrieval, biometrics and pattern recognition.

Network Analysis of Social Awareness of Media Education for Primary School Students Studied through Big Data*

Su-Jeong Jeong[1] and Byung-Man Kim[2, †]

[1] Creativity & Personality Laboratory, Tongmyong University,
Pusan, 48570, Republic of Korea
jsjs@tu.ac.kr
[2] Dept. of Early childhood education, Kyungnam University,
Changwon, 51767, Republic of Korea
bmkim@kyungnam.ac.kr

Abstract. One of the core competencies of students who want to improve in future education in the 21st century is the ability to utilize media. Primary school students entering the school age can be regarded as the right time for media education because of their high adaptability and capacity for media. The purpose of this study is to examine the social debate about media education in Korean society, how media education is being conducted in this important primary school period. For this study, big data was collected in the last 5 years (2014.08.07-2019.08.07) from internet portal sites with keywords of "primary school media education" and "primary school media literacy". The data collected with Textom and Ucinet 6.0 was utilized as a data analysis solution. Semantic network analysis, CONCOR analysis, and content analysis were used as data analysis methodology. As the result of CONCOR analysis of 'Primary school Media Education' in this study, 'Direction of future education in the era of the 4th industrial revolution', 'Preparation for future education', 'Various factors related to expansion of future education program', 'Expansion to Four factors were derived, including 'Application to advanced classes' factor. In the 4th industrial revolution, primary school students are using media and digital devices in class. In particular, it was confirmed that not only curriculum for academic subjects such as English and Mathematics, but also new curriculum for new subjects such as coding, big data education are actively being conducted in the education field. Accordingly, it is revealed that it is the right time to provide future education that can have a sound digital identity so that media education can be achieved in a media-friendly local community and educational environment.

Keywords: Big Data Analysis, Network Analysis, Content Analysis, CONCOR Analysis, Primary School Student, Media Education

---

1. Introduction

The media environment is rapidly changing and developing with the development of high-tech information and communication technology in this era of the Fourth Industrial Revolution. Media literacy is becoming more important, and the need to develop the ability to critically accept media is increasing as the perception spreads that no one can escape the influence of the media in the knowledge information society. Through this, media education is designed to facilitate the use, sharing and fostering communication skills of the media as a preparation for media changes [22].

The development of communication technology has brought a change in media education. The dissemination of digital media has combined the functions of each media, the change to an interactive form of communication through the Internet has led to the need for a more active and creative producer image, and the educational environment was transformed by the development of information and communication technology and communication technology emphasizes the need for a learner-oriented active and participatory media education from childhood.

Just as we have to learn letters and grammar to read, write and understand text, media education is needed to facilitate understanding of grammar and skills of emerging media and content analysis. In particular, since the use of media medium by children continues to increase, the effectiveness of education has been constantly studied in South Korea and abroad.

Previous studies have reported that the media use of children is a major cause of negative development, but cannot be determined. However, it has been reported that it has a significant effect and reported the potential harmfulness of media.

This is the time to emphasize on the importance of media education especially for students in elementary school, when media use is rapidly increasing.

In the digital revolution and the fourth industrial revolution, the educational ecosystem is being recreated and rapidly changed. Moreover, this is an era that the educational ecosystem should be designed to customize learning materials and information for learners [14]. From this point of view, the media is coming to us with forms of smart running, digital revolution, etc., and it can be seen that rapid media changes and developments are affecting the way children think and behave in sociocultural changes.

The use of the Internet and TV by children in today's homes is very common, and the proportion of elementary school students owning cell phones has also increased recently, children's use of the Internet, TV, and mobile phones has become an aspect of their lifestyle. A meaningful prior study [11] reported that the Internet and TV could be tools for developing elementary school students' potential, such as information exchange, learning promotion, and leisure activities, while having ambivalence that can hinder elementary school students' healthy growth and emotional development, including game addiction, pornography, distribution of violence, and various cyber crimes. As such, the importance of media education cannot be overemphasized because the media has great influence on the growth and development of children depending on the direction of media utilization.

Given that media education should be conducted in various ways and channels, media education should be implemented in conjunction with home and society, not only through conventional school education. And the study of media education should

continue throughout one's life. Media education, especially in elementary school, is more important because it has a profound impact on the overall development of children.

The Davos Forum in 2012 presented 'big data research' as the most powerful tool for social awareness and resolution for social issues [7]. It is necessary to consider big data generated on the web to review social discourse and practical requirements related to media education for elementary school students. Big data encompasses the vast scale and various kinds of web-based data generated in a digital environment[7, 31], characterized by hyper-connectivity and super-intelligence in the era of the Fourth Industrial Revolution, allowing the rapid processing and analysis of vast amounts of Internet-connected information to derive meaningful results and implications, and to analyze and deduce complex meanings [15]. In the field of education, various studies have already been accumulated, and opportunities to connect and analyze them have increased. Big data and network analysis help to actively accommodate changes in the environment by forming new theories and reconstructing their own perspectives.

Although some studies on the perception and current status of media education studied so far have been conducted [6, 14, 19, 30, 39], most of the studies have been to develop programs for media education and to improve attitudes toward media education. Therefore, it can be seen that there is a lack of research on the general perception of media education including media education. Moreover, media education research for primary school students has not been conducted until now. Given the predictions of futurists that the digital learning ecosystem will occupy a large part of the educational space, it is necessary to examine the social perceptions of media education for primary school students through big data.

In this study, we want to look at the social perception of media education of Korean elementary school children created over the past five years through big data. We also want to analyze the network and contents of social awareness to derive the social discourse contained in media education. The study, which explores major issues through social awareness of elementary school children through big data, will reveal social interest and awareness of educational phenomena and suggest implications in more diverse aspects. The research questions of this study are as follows.

1. What is the social recognition about media education for primary school students shown in Big data?

2. What is the social recognition about media education for primary school students shown through content analysis?

2.   Related Works

As media discussions progress, media education is used as an extended concept not only to media use and understanding, but also to media abilities and communication skills, which are active dimensions. In other words, it includes the content delivered by the media and the human's total ability to be free from the media, and the ability of individuals to participate in the communication domain as a political component [32].

In fact, the expansion of this concept can facilitate the entry of media education into schools. This is because school education seeks direction through creative education in

the knowledge-based society as the times change. However, the reality is that it is difficult to grasp what kind of education is being conducted in connection with media education and to what extent substantial education is being implemented. This is because it is difficult to comprehensively grasp the current status of media education due to the variety of media education units, educational subjects, and program types at the school site [17].

Therefore, in order to improve the system of media education, it is necessary to take a closer look at what level of media education currently being conducted. Especially, the problems of media education are often mentioned, so it is necessary to check the status of media education. The preceding studies related to media education are as follows.

As a result of [25] analysis using an in-depth interview to analyze how internet-related media education is defined and recognized in the field of primary education, the media is not grasped from a simple tool point of view, but it is suggested that the primary school education should go toward internet related media to gain critical media literacy perspective [8].

As a result of a survey conducted on primary school students, 'A study has been reported that excessive use of media for more than 7 hours a day is reported. There is also an intensive study [12] insisting that multi-dimensional instruction is needed to form daily habits for media use in the early stage of primary school period. From this early primary school period, it was suggested that special attention should be paid to the use of media, and that the quality of media used by students should be considered.

In addition, looking at existing media education literature, focusing on the intellectual acquisition of media and production of content, or obtaining the right attitude to use media is approached only through the program contents and evaluation of media education. Or, a critical interpretation of the media education program includes studies that remain at the primary school level [1, 20, 28, 33]. As described above, most of the previous studies constitute and evaluate content production and programs, or derive the suggested research results through surveys and interviews so there's limitation go generalize the research result. In order to elicit generalized implications, this study intends to examine the Korean society's perception of 'primary school media education' quantitatively and qualitatively by combining big data analysis and content analysis.

In particular, as big data is referred to as 'crude oil of the 21st century', and has been noted as a key technology and promising industry in the future [4], 'big data' is a hot topic in all walks of life. Based on the news coverage of the media through big data analysis and the public's thoughts, we can examine what issues and values of media education are forming discourse in our society, and deduce how the direction of social perception is moving forward.

3.   Research Method

In this study, for the process of collecting and refining big data for media education, original data were collected using Textom, and then the first and second data refinement processes were implemented [24]. Programs such as Textom, Ucinet 6.0, and Netdraw were utilized as data analysis solutions to perform network analysis among keywords

related to media education. As data analysis methodology, semantic network analysis, CONCOR analysis, and content analysis were conducted.

Table 1. Research procedure

Data collection and 1st coding	Internet Portal Site		
	Google	Naver	Daum
	•Big data is collected from news, articles from online communities, blogs, etc. on the Internet portal site • In the first refinement, only nouns were extracted when data was collected, and special characters and symbols were not included		

↓

| Data refinement and 2nd coding | •Read the articles collected by Internet portal sites
•In the second data refinement, data that had been first refined was used, numbers, words in English, and pronouns in the data were deleted, and overlapped documents were removed.
(Contents not related to primary school media education or media literacy education are deleted)
•Media education classification by Internet portal site |

↓

| Network analysis | •Textom(text mining), Ucinet 6.0, Netdraw |

↓

| CONCOR analysis | •Classification of media education factors by cluster extracted using CONCOR analysis |

↓

| Article selection and Content analysis | •Selecting articles suitable for factors extracted from CONCOR analysis
•Review of articles extracted by factors |

↓

| Results | •Research results and implications |

3.1.   Data Collection & Data Refinement

Data collection includes all stages of data collection, data cleaning, data clustering, and visualization, from large amounts of relatively low-valued data to the process of obtaining insightful and advanced information [26].

The data collection was performed using a computerized automated method, and the collection method was collected using a public API (Opne Application Interfac: Opne API).

The purpose of this study is to find out what social perception of primary school media education is. For this, big data was collected from newspaper articles, articles from online communities, blog articles from Internet portal sites such as Google, Naver, and Daum. Total collected big data from 2014.08.07. to 2019.08.07. is 2,755KB.

In the first refinement, only nouns were extracted when data was collected, and special characters and symbols were not included. In the second data refinement, data that had been first refined was used, numbers, words in English, and pronouns in the data were deleted, and overlapped documents were removed.

3.2.   Semantic network analysis

By analyzing specific topics of unstructured text big data through semantic network analysis, it is possible to accurately and quickly grasp various viewpoints, needs, and emotions of our society. Big data is a continual record of what is happening, and when analyzing data from various sources, it can be used as a direction to grasp the ripple effects of policies or to seek new policies.

Data and information created by a large number of unspecified masses are converted into text to analyze the relationship among key keywords according to needs and purposes, making people's behaviors and psychological states predictable, and recognizing keywords as nodes in text. So it is possible to make a network through visualizing by connecting the relationship with a line.

3.3.   CONCOR Analysis

CONCOR analysis was performed to grasp the relationship among keywords. In particular, CONCOR analysis can be grouped into the same group among words. In other words, it is a technique used to identify characteristics of similar types or to observe differences between the group and other groups by analyzing the characteristics by dividing the target group with many changes into a certain group and analyzing the distance between the data. In addition, weights are assigned depending on the importance in context and appear as bold lines. Words with similarities among words are not arbitrarily manipulated by the researcher, and given data are defined by themselves through computers and grouped within the same cluster. Accordingly, it is possible to check the meaning according to which word clusters appear.

3.4.   Content Analysis

This study selects the content analysis method [3] to identify the social trends in media education of primary school students. This is because it is suitable for systematically analyzing the characteristics of messages from various types of text displayed on the

Internet portal sites. Regarding the content analysis method, [38] defines "content analysis is a systematic procedure designed to investigate the content of recorded information".

4.   Experimental Results

4.1.   The statistical significance test of whole network for media education for Primary School Students

In this study, the statistical significance of primary school students' media education was examined with 2,755KB of data refinement from big data searched in the last 5 years (2014.08.07.-2019.08.07.) for the purpose of identifying the social recognition with the key words, 'Media education for Primary school students' and ' Media literacy of Primary school students'.

In order to statistically test whether the estimated density of the network was accidental, the single-sample mean difference was tested using bootstrapping[5].

As a result of the study, the mean distribution of the mean network data was 24.31 and the standard error was 2.38. Looking at the Z-score value, the standard score, it was found to be statistically significant at ***$p<.001$ level with Z=9.9578 and P=0.0002.

In other words, in this study, it can be seen that it is appropriate to interpret the analysis results with the entire network for media education of primary school students.

4.2.   Frequency analysis of keywords related to media education for primary school students

In Primary School Media Education, the words 'Child', 'Education', 'Teacher', 'Object', 'Primary School', 'Class', 'Student', 'Program', 'Media', 'Digital' etc. are the key words and it was reveled that they are the new agenda in the new era.

In particular, it can be assumed that digital device-based education is being activated in education programs for children and adolescents as the future education of the 4th industrial revolution.
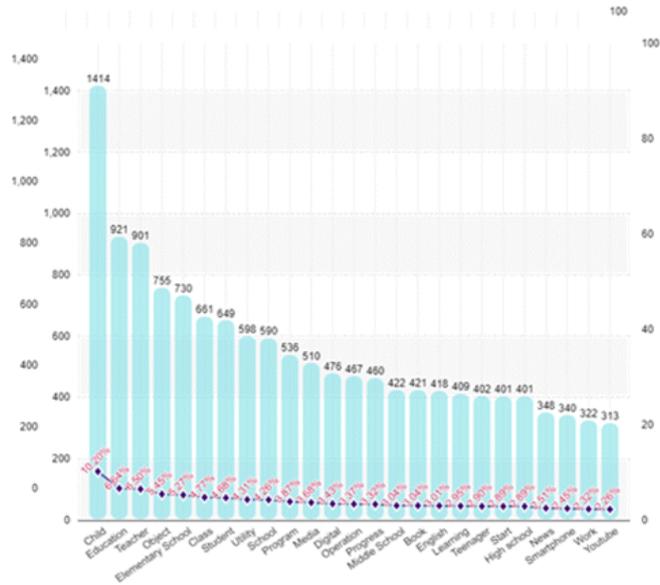
Fig. 1. Top 25 word bar chart of "Primary School Media Education"

4.3.   Network analysis of media education for Primary School Students

In order to examine how the centrality of the keywords in the network appears, the centrality analysis focused on the center of Dgree, the Closeness, and the Eigenvector. Centrality is a measure of the relative importance of vertices or nodes in a graph or Semantic network[23]. It can be seen that the words located at the center of the network represent a high value and constitute a core issue.

Table 2. Analysis result of 'Degree Centrality' using Ucinet

No	Word	Degree	No	Word	Degree
1	Child	3170	11	Digital	1677
2	Teacher	2779	12	High school	1562
3	Education	2508	13	Operation	1554
4	Object	2422	14	Media	1553
5	Utility	2048	15	Progress	1479
6	Primary School	1904	16	Learning	1312
7	Program	1870	17	English	1248
8	School	1804	18	Middle School	1243
9	Student	1751	19	Teenager	1214
10	Class	1680	20	Mathematics	1168

First, looking at <Table 2> showing the 'Degree Centrality' figures, 'Child', 'Education', 'Teacher', 'Destination', 'Primary School', 'Class' and 'Student' in terms of 'Degree Centrality', 'Program', 'Digital', and 'Media' are used in connection with many words. Based on this, it implies that when applying primary school media education to actual classrooms, it is necessary to provide a "media program" focusing on individual "children" education services.

Table 3. Analysis result of 'Closeness Centrality' using Ucinet

No	Word	Closeness	No	Word	Closeness
1	English tutoring	63	11	Coding	52
2	Mathematics	62	12	Study	52
3	Practice	57	13	Posts	52
4	Future	55	14	Multimedia	52
5	Training	54	15	Parents	52
6	Summer Vacation	53	16	Object	51
7	Development	53	17	Primary School	51
8	Game	53	18	Media	51
9	Smartphone	52	19	Teenager	51
10	Youtube	52	20	Textbook	51

Second, looking at <Table 3> showing the values of 'Closeness Centrality', it is composed of the words such as 'English', 'the future', 'Education', 'Mathematics', 'Primary School', 'Smartphone', 'Youtube', 'Digital', 'media', 'coding', 'game', and 'teenagers'. This suggests that media and digital education utilizing 'games' and 'codings' is being introduced in the education field for teenagers.

Table 4. Analysis result of 'Eigenventor Centrality' using Ucinet

No	Word	Eigenvector	No	Word	Eigenvector
1	Object	0.324	11	Media	0.119
2	Program	0.273	12	Education	0.118
3	Child	0.265	13	English	0.116
4	Operation	0.23	14	Smartphone	0.111
5	Progress	0.205	15	Learning	0.103
6	Utility	0.183	16	Start	0.101
7	Summer Vacation	0.174	17	Work	0.099
8	Teenager	0.157	18	Book	0.093
9	Digital	0.137	19	Youtube	0.091
10	Experience	0.121	20	News	0.09

Third, looking at <Table 4> showing the 'Eigenventor Centrality' figures, the 'target', 'program', 'child', 'operation', 'progress', 'utilization', 'summer vacation', 'experience', 'digital'. As indicated by keywords such as', 'media', 'Youtube', and 'news', it can be seen that when media education is applied to children, it is proposed to apply various media in various educational situations.

4.4.   CONCOR analysis on media education for Primary School Students

Fig. 2. CONCOR Analysis of "Primary School Media Education"

The table 2 above indicates the data after grouping and visualizing with CONCOR analysis. As you see in the table 2, it can be identified that which words have the highest correlation with the key word" Primary school media education".

The 4 representative clusters are as follow. They are named as factors for the direction of future education in the 4th Industrial Revolution era, preparation factors for future education, various factors related to expansion of future education programs, and application factors to advanced classes.

Table 5. "CONCOR Analysis of Primary School Media Education"

Clusters	Key word
1. Direction of future education in the era of the 4th industrial revolution	Future, education, thinking
2. Preparation for future education	Education, Textbook, Training, Need, Media, Video, Era, You-tube, Primary School, Digital, Gaming, Understanding, Multimedia
3. Various factors related to expansion of future education program	Child, Student, Target, Operation, Experience, Class, Coding, Start, Parents, Summer Vacation, Progress, Program, News, Middle schooler, Teenager, English, Utilization
4. Application to advanced classes	Teacher, School, High School, Middle School, English Tutoring, Mathematics

4.5.   Content analysis of big data of media education for primary school students: focusing on 4 clusters of Concor analysis

Direction of future education in the 4th industrial revolution era

The cases in the big data related to future education in the 4th industrial revolution era are as follow.

   [Case 1] …(omitted) The 4th industrial revolution affects our society and overall culture / not ending with changes of industries and economy. We often hear the new term "new normal era" these days. It means that the basic principle which drives our society has changed. Also, this means that the things that were previously taken for granted are no longer taken for granted, and a new naturalness is created. This explains that the 4th industrial revolution has become a "ism" not ending with simple technology and industrial changes. Just as Fordism created modern school system, it can be anticipated that the 4th industrial revolution will create new education system while creating 'new normal' in education.(*Siminsori , May 10th , 2018: The 4th industrial revolution, the direction of future education?*)

   [10], the renowned historian and a writer, predicted that there is high possibility that the knowledge they learn from school education would be useless, when our children are in their 40s. So we need to try to find implications with these changes in our education field.

   [Case 2] …(omitted) Naun primary school was supported the state-of-the art smart devices such as electronic boards, smart solution, wireless network, 360 gear cameras, tablets, laptops, and mobile devices and so on because the school was designated as Samsung Electronics's social contribution project. Currently, the school is equipped with smart e-class and it is conducting various classes such as creativity integration education, trouble shooting education, and creativity education by using the various devices that they were supported. The primary school is considered as the examplary case that improves school education environment and tries to provide education conditions to lead future education revolution going beyond the limitation as a small regional community. (*Naver blog, March 24th, 2019: " Naun Primary school in South Korea euipped with the cutting edge smart e-class , leading future education revolution "*)

   This is the time that each city needs efforts to prepare new education revolution that is suitable for the 4th industrial revolution era beyond the limits of the classes with using simple multimedia and regional space.

   [Case 3] … (Omitted) In South Korea, coding classes have been mandatory in 2018 for 17 hours per year as a regular curriculum for primary school students in grades 5-6, and they are keeping pace with the global trend in preparation for the 4th industrial revolution. Already in developed countries, education has been mandated to foster talent for the 4th industrial revolution including coding, and voices have been raised in teaching responsible usage and digital citizenship in media use (*Naver Blog, June 13th , 2019: . "Digital Citizen School"*)

Now that coding and big data have become important core education which is indispensable to our children, we have indicated that it is time to teach digital citizenship to children because we need to use digital media with systematic and correct content.

[Case 4] … (Omitted) Vice Chairman Kwak said, "At the point of going from Society 4.0 to Society 5.0, above all, for a rewarding future society, a society in which values are created, a society capable of exerting various abilities, a society where everyone can get a chance, and anyone can safely challenge and a society that goes along with nature need to be established and we need to build capacity and cultivate virtues." He continued, "As technology has created an intelligent information society for humans, above all, we should not neglect to develop social and emotional competence in education for the future life of children." In other words, people-centered education with the virtue of being considerate of others and yielding to others must be emphasized. It is also to nurture people who can create values that the future demands by empowering education that can create better values. (*Digital Today, August 5th, 2019: "People-Oriented Future Creator"*)

Since digital transformation will create positive and negative things that affect society, we will have a future where new science and technology can have a sound digital identity based on common sense of responsibility and true communication to develop human-centered public interest functions. Education should focus on it.

[Case 5] … (Omitted) On September 4th, at the Samsung Primary School in Seoul, The event '2017 Beautiful Internet World Weekly Internet Ethics Tour Lecture and Golden Bell' was held. 131 students in grades 5-6 attended and were able to learn desirable Internet usage habits. … (Omitted) Seoul Samsung Primary School said, "We expect the students to learn how to use the Internet properly and cultivate healthy usage habits and Internet ethics through this lecture on Internet ethics." We will continue to make efforts to promote for software education activation. (*Naver Blog, September 7th, 2017: Internet Ethics Easily Learned from Quizzes and Cases*)

These events can be a great opportunity for students to think about the seriousness of Internet ethics and cyber bullying.

Preparation for future education

The key words for Preparation for future education are appeared with the node 'Education', 'Textbook', 'Training', 'Need', 'Media', 'Video', 'Era',' You-tube', 'Primary School', 'Digital',' Gaming', 'Understanding' and 'Multimedia'

The cases in the big data related to preparation for future education are as follow.

[Case 6] … (Omitted) Nowadays, many people enjoy a game culture from infants to children, teenagers, and adults with smartphones that are easily accessible. There are many positive aspects, such as concentration, learning ability improvement, and challenging spirit, depending on how to use the game that used to be considered bad. The 2017 Game Literacy Teacher Job Training is a project sponsored by the Korea Contents Promotion Agency and sponsored by the Ministry of Culture, Sports and Tourism for the purpose of raising teachers' awareness of game culture and enhancing their capabilities to operate classes with each subject. (*Naver Blog, December 13, 2017: 2017 Game Literacy Teacher Job Training - Preparation for Future Education by utilizing Games*)

In the game, the collaboration between the head and limbs is vital, and it needs access to a complex text that stimulates various sensations such as visual and tactile senses. As there are many positive aspects in this game, if used well in future education, it can be an interesting medium to study.

[Case 7] KT (KT corporation) has improved the educational environment such as artificial intelligence software coding education, sports experience space using mixed reality (MR) technology, and content production support so that Daeseong-dong village primary school students can develop their dreams in the state-of-the-art education infrastructure…. (Omitted) Students can train physical fitness regardless of fine dust or outside weather. The 'MR Screen Sports' provided in the school auditorium can perform 25 kinds of sports activities such as soccer, basketball, and boxing, and it is also possible to conduct simultaneous classes with other schools through the network. In addition, it supports 5G smartphone and 360-degree shooting for one-person media content production and supports a neckband-type camera called 'FITT 360 (Fit 360)'. Students are planning to create a content containing the story of Daeseong-dong village, where is the village with difficulties communicating with other cities or any other communities, and share it on social media such as YouTube to inform the peace message of Korea, the only divided country in the world. (*Naver News, June 27, 2019: Improving educational environment such as AI software coding training, MR sports experience space*)

It can be seen that 5G is installed in the primary school located in the DMZ (Demilitarized Zone) to improve the educational environment so that dreams can be developed in the state-of-the-art education infrastructure.

[Case 8] 'Book' Live Science is in the form of learning cartoons, making it easy to build and understand primary science knowledge.(Omitted)… In addition, multimedia videos on various topics are introduced with fun, and services such as 3D and 2D animation science videos are also provided. (*Naver Blog, November 18, 2018: Learning cartoons- Live Science Internet of Things -Fun Primary Science Book*)

In science books for primary school students, it can be seen that parents with children have a preference for books in which media education materials are combined, and it is confirmed that they are encouraging children's development through games and future education through cartoons.

Various factors related to expansion of future education program

The key words are appeared with the node, Child, Student, Target, Operation, Experience, Class, Coding, Start, Parents, Summer Vacation, Progress, Program, News, Middle schooler, Teenager, English, Utilization etc.

The cases in the big data correlated with Various factors related to expansion of future education programs are below.

[Case 9]… (Omitted) There was a scratch challenge event by WCG on the theme of 'play of the future'. Scratch is a programming language for fostering children's creative thinking and systematic reasoning. It was created to learn coding. Unlike conventional text coding, a simple game or animation by connecting and coding the script like a block can be created. Korea also has a growing interest because coding is included in the

regular courses of primary school. (*Next News, July 21, 2019: What will the future play look like?*)

Coding, one of the future education programs, was involved in the regular curriculum, but it is being introduced to children as a fun way to learn so that they can naturally learn and enjoy future play.

[Case 10] …(Omitted) Seoul Arts Foundation Seoul Arts Education Center held a cultural arts education festival, 'Arts Vacation,' which can be participated in by any citizens, including children, teenagers, and families, on the summer vacation. The main program of 'Arts and Vacations' is an art education experience in the form of a play created by six artists residing at the Seoul Arts Education Center Art Play LAB (Lab). These are three integrated art programs created by fusion of various genres such as visual arts, sound, and theater. The detailed programs include 'Moving Doremipasolasido', which uses a media device that sounds when focusing on color to play a space composed of various colors. (*Daum news, July 26, 2019: With summer vacation culture… 'Art and Vacation' 'Circus Family Camp'*)

As such, various experiences of future education programs are provided by the local community, and in particular, as media education and arts are presented as integrated art programs, various media education directions can be suggested.

[Case 11] During the summer vacation, Sejong City announced that it would conduct a task-oriented English camp to improve English conversation skills for all primary and middle school students. The Primary English Education Support Center (primary school) provides an English camp with an interesting topic that utilizes media for 4 hours a day from 9 am to 12:10 am, from 22nd to 26th of July. 36 students in 2 classes in 4th grade and 54 students in 3 classes in 5th to 6th grade participate in the English Camp. The English Camp for Junior High School students operates 4 classes with 18 advanced classes, 2 intermediate classes, 30 classes, and 14 basic classes, and the number of students is divided equally by schools, and selected by drawing lots among applicants. (*next news, July 29, 2019: Sejong Office of Education runs a summer vacation English camp for primary and junior high school students*)

It can be seen that various courses by regions are connected to media education to construct and operate an experiential program that considers students' interests. In addition, it can be seen that it is a popular program that many students want to experience, as the number of people is divided evenly by school and selected by lottery. As such, it is desperate to provide quality experience-based media education programs that students want to participate in.

Application to advanced classes

The key word is appeared with the node, Teacher, School, High School, Middle School, English Tutoring, Mathematics, etc. The cases in the big data related to Application to advanced classes are below.

[Case 12] The Seoul Metropolitan Government prepared an experiential program for teenagers to spend a summer vacation making' the hyper-connected DNA'. The sectors are diverse, such as IT science, art and culture, service, career and career, international, ecology and environment, history and society, sports, and others. In particular, during this summer vacation, the programs such as the 'Media Literacy' education program that

can help have the right perspective to read the times and information, the 'IT Science', 'Technology Convergence' program to learn and learn advanced technologies, and the 'Camp and Volunteer Activities' program to nurture creative personality are added…(Omitted)... In the media field, "Shooting the Start-up" will be opened for young people interested in starting and creating video contents. (*next news, July 26, 2019: If you hesitate, there's no room for you. the deadline will be pouring out for the summer vacation for youth experience!*)

As a vacation experience program for teenagers, they were providing programs to experience various fields such as IT science, arts, service, and sports, and to develop personal competencies. Especially, in the hyper-connected age, 'media literacy' education programs are increasing, so many teenagers can check the supply and demand for media education programs.

[Case 13] Now, the world with smartphones and the Internet has arrived. According to a recent survey by the Ministry of Gender Equality and Family, 200,000 Korean teenagers are at risk of Internet and smartphone addiction. The problem of various media addictions such as TV, internet, and games is now a matter of society as a whole. Is it the best way to ban viewing unconditionally from indiscriminate exposure of various media? In the new semester, we are trying to find a way to protect children from media addiction and educate them properly in a variety of media floods such as TV, Internet, ...(Omitted)...Today is the time that media must be actively utilized. (*Daum Cafe, August 23, 2016: Youth, how to protect yourself from media addiction?*)

Media in the present era is a mean of dialogue and communication between people, and it is an era in which media must be actively utilized. Since media and smartphones that are frequently used today are value-neutral, their importance may vary depending on who uses them and how. Therefore, through media and digital media education, media education should be conducted for all ages so as not to fall behind in the era of the fourth industrial revolution.

[Case 14] On July 26, the Sejong Special Self-governing City Office of Education announced that it developed an educational content distribution platform (smart-eye) in July, last year and applied it to the classrooms of front-line schools. Currently, a total of 5.8 million commercial contents are registered such as EBS video lectures and teaching and learning materials, for student learning, teacher teaching, and other related contents. These vast educational contents are being actively utilized in the classroom by displaying them on a smart electronic blackboard in accordance with the textbook units and chronologies of the corresponding grade at the click of the mouse with the teachers. The City Office of Education plans to add self-directed learning functions, where students self-diagnose online learning and develop correction strategies for each type. ...(Omitted)...In addition, one-click smart classes are also expected to be upgraded. The city office of education uses in-depth analysis of the textbooks of middle and high schools to utilize the smart eye system anywhere in Sejong City's first-line schools to overcome the limitations of not being able to share teaching and learning materials, because textbook publishers are different for each middle school and high school, unlike primary schools. Improvement measures were devised to prevent restrictions. In addition, education utilizing information technology from Sejong City, which is leading Korea's smart education, is also spreading across the country. (*Daum, April 3, 2015: Upgrading educational content distribution platform*)

In addition to large-scale online learning, we also prepare a tailored correction strategy for each type that provides multimedia content, so it is expected that the concentration of classes will be greatly improved by providing customized classes for individual students. Furthermore, it can be seen that online learning is not only limited to primary schools, but is also spreading and expanding to middle and high schools.

5.   Result & Discussion

The purpose of this study was to confirm the social awareness of media education of Korean elementary school children created over the last five years through big data. It was also intended to derive social discourse contained in media education by conducting network and content analysis on social awareness. Specifically, big data consisting of Internet articles and blog posts related to social trends in media education for elementary school students was identified through text mining analysis.

Afterwards, scientific and objectivity was secured through analysis of semantic network of collected data, and content analysis was conducted by directly reviewing data after pre-processing and classifying it into four aspects of media education for elementary school students. The study, which explores major issues through social awareness of elementary school children through big data, suggested implications in more diverse aspects by revealing the social interest and perception of educational phenomena through a mixed research method.

The conclusions and significance of this study are as follows. First, as a result of analyzing big data to analyze the social perception and trends of primary media education, meaningful information about primary media education was extracted. First of all, the most frequently appeared keywords were "children, education, teachers, subjects, elementary schools, classes and students." In particular, as a result of semantic network analysis, "children, education, teachers, subjects, elementary schools, classes, and English tutoring" were the key keywords. Keywords with high proximity centrality appeared as "Destination, Elementary School, Media, Operation, Teenagers, Smartphones, English, and Coding," confirming the influence of smartphones. Keywords with high mediated center were "children, education, teachers, classes, students, utilization, programs, digital, and English."

These research results show that elementary school students are using smartphones or digital devices a lot, especially media education for English education or coding education [29, 36], so it is time to understand media as an educational and cultural environment and to further require critical media literacy education.

Applying various media to education in the field of education has now become a natural phenomenon, and the importance of media education is expected to continue to be emphasized more in the future. Utilizing these important factors of media education, such as immediate accessibility, knowledge scalability, and collaborative interaction, students will increase their ability to learn new knowledge by collecting, editing, and generating various information. Education using media has become more common in recent years, and considering the situation in which students are actively using the media as a customized self-directed learning method as well as elementary school class sites

[37], the ability to use the media is expected to draw attention as a more important core ability in the future [27].

Next, words such as education, elementary school, smartphone, YouTube, digital, media, and information appeared frequently. Through this, media and digital devices are actively used to educate elementary school students, and media utilization has become more common, especially as smartphones have become popular among elementary school students [21, 35]. In particular, it can be seen that various methods are being worked on at educational sites to acquire diverse information or to enhance problem-solving skills in order to enhance learning ability using media [34]. Also, words such as teachers, schools, middle schools, high schools, English extracurricular activities, and math education were frequently mentioned. In other words, it can be inferred that teachers are using media education to educate middle and high school students in English or math. Finally, words such as "future," "lecture," and "think" appeared frequently.

According to the study of [40], future media education is required considering the developmental characteristics of primary school students. In particular, it is revealed that the subjects of class need to pay close attention to how to solve the obstacles in class according to the level of development of the students. This means that in order to provide high-quality lectures to future primary school students, it is necessary to deeply understand the level of development of students and to provide lectures appropriate to the needs of students.

In addition, when media education is conducted in elementary school classrooms, the focus should be on interaction between teachers and students rather than focusing on media utilization that becomes educational media. According to a study by [40], since interactive communication positively affects students' learning attitude, it can be seen that the interaction between teachers and students is important when using media education mediums in class. In particular, it is suggested that teachers give careful guidance on the spot so that students can improve their ability to solve problems through deep thinking.

Second, in this study, CONCOR analysis was conducted to derive groups with appropriate levels of similarity, and four clusters were created. It was named 'Direction of future education in the era of the 4th industrial revolution', 'Preparation for future education', 'Various factors related to expansion of future education program', and 'Application to advanced classes' factor.

Presenting the direction of future education in the era of the Fourth Industrial Revolution, support is needed not only within the school but also within the community. According to the research results of [17], which analyzed the experience of participating in after-school media education programs in local communities, it is vital that all main players of school education such as teachers, students and parents need to have close communication and cooperation to operate and vitalize media education. In other words, content should open the world and strengthen the ability to live together through education that draws empathy through emotion.

However, on the other hand, just as the encounter with game addiction or cybercrime hinders children's healthy growth, how the media is used has a great influence on the psychosocial development of children and adolescents [14, 16]. Therefore, Ethical education for Internet ethics is needed to solve social problems such as cyberbullying.

In order to prepare for future education, it was confirmed that various programs such as future education using play, AI, MR content education, and multimedia media platform were requested to be expanded and presented. These research results show that elementary school students are using smartphones or digital devices a lot, especially media education for English education or coding education, so it is time to understand media as an educational and cultural environment and to further require critical media literacy education. In the study of [2], the cutting-edge educational cultural environment was effective in enhancing learners' academic performance. In addition, it was reported that learners' spontaneous learning motivation was improved by introducing and activating various learning strategies and core educational technologies in the educational field. Moreover, it is now natural to apply various media to education in the field of education, and the importance of media education will continue to be emphasized more in the future.

Using these important factors of media education, such as immediate accessibility, knowledge scalability, and collaborative interaction, students will increase their ability to learn new knowledge by collecting, editing, and generating various information. Education using media has become more common in recent years, and even considering the situation in which students are actively using the media as a customized self-directed learning method as well as elementary school class sites, the ability to use the media is expected to draw attention as a more important core ability in the future.

6.   Conclusion

This study is meaningful in that it suggested basic information and materials to prepare a plan for revitalizing media education for primary school students by analyzing social discussions of media education through big data collected various channels. Moreover, to analyze the social discussion of media education, a mixed methodology using both quantitative and qualitative research methods was applied. The use of hybrid methodology enabled to have in-depth insight and understanding of current status [9, 18].

In Korea, there have been some prior researches conducted in relation to the perception and discussion of media education using big data [6, 14, 20, 30, 39]. However, it can be seen that there is a lack of research that examines the general perception of media education, including media education, as the main focus is on the development of programs for media education and research on improving attitudes. Moreover, media education research for primary school students has not been conducted so far.

And it is difficult to find a study that has both quantitative and qualitative studies. Therefore, it is different from previous studies in that it uses big data to conduct quantitative study for recognizing social perception of media education and this study is isolated since it finds the context in the qualitative studies, which was not found in the quantitative studies.

Furthermore, the fact that this study shows concrete clusters of social perception of media education for primary school students through Concor analysis and it presents educational implications make this study more isolated from previous studies.

Presenting implications and suggestions based on these conclusions is as follows. First, when media education is conducted at primary school sites, the focus should be on interaction between teachers and students, rather than focusing only on the media utilization that becomes the educational medium. In particular, it is suggested that teachers give careful guidance on the spot so that students can improve their ability to solve problems through deep thinking.

Second, from the time of elementary school, we should continue to strengthen our educational efforts on the use of smart media and the correct attitude and ethical norms of activities in the world of smart media, while increasing our ability to control smart media rather than being subordinate to smart media, so that we can have a more positive attitude and belief.

Finally, from the time of elementary school, we should continue to strengthen our educational efforts on the use of smart media and the correct attitude and ethical norms of activities in the world of smart media, while increasing our ability to control smart media instead of being subordinate to smart media, so that we can have a more positive attitude and belief.

References

1. An, J. I., Jeon, K.L.: Media Education Program Materials and Curriculum Directions. Korea Press Foundation, Seoul. (2004)
2. Bae, Y. N., Ahn, M.L.: A Case Study on the Introduction of Digital Badge for Future Education Learning Environment. The Korean Association of Computer Education, Vol.22, No.1, 37-40. (2018)
3. Berelson, B.: Content analysis in communication research, Hafner publishing company, New York (1971)
4. Beyer, M. A., Laney, D.: The Importance of "Big Data": A Definition. Stamford, CT. (2012)
5. Choi, H. J., Choi, Y. C.: A Study on Children`s Creativity and Character based on Big Data. Korean Society of Children's Literature & Education, Vol.17, No. 4, 601-627. (2016)
6. Choi, M. S., Kang, B. H.: Recognition of Infants and Guardians under Smart Media Education Environments. Journal of the Korea Entertainment Industry Association, Vol.8, No.3, 335-345. (2014)
7. Chung, Y. C.: Big data. Communication Books, Seoul. (2013)
8. Eun, M. S.: The Relationship of Smart Media Literacy`s Factors for Primary School Student on Subject Attitude and Achievement. Journal of Korean Association for Educational Information and Media, Vol. 21, No. 2, 215-243. (2015)
9. Green, J. C.: Qualitative paradigm evaluation. In N. K. Denzin. & Y. S. Lincoln(Eds.), Handbook of qualitative research. Sage, CA. (1994)
10. Harari, Y. N.: Sapiens. Gimmyoun, Gyeonggi. (2015)
11. Jeon, S. K., Kim, Y. J.: A Study on the Development of the Goals and Content System of Media Education for the Parents of Elementary School Students. Journal of Korean Practical Arts Education, Vol. 21, No.2, 265-285. (2008)
12. Jeong, S. J.: The Factors and Development Outcomes of Latent Types in Children's Time Use. Doctor´s thesis, Ewha Womans University. (2018)

13. Jeong, S. J.: A Semantic Network Analysis of Research Trends on Media Education of Elementary School Students. ICIA. (2020). Unpublished.
14. Kang, S. J., Lee, Y, S.: A study on social perception of childrens smart media education based on big data. The Journal of Korea Open Association for Early Childhood Education, Vol. 22, No. 4, 45-72. (2017)
15. Kim, B. M.: Network Analysis of the Social Discourse of Integration of ECEC, Based on Big Data. Journal of Educational Innovation Research, Vol. 29, No. 2, 17-39. (2019)
16. Kim, E. G., Kim, S. B.: A Study to frame up media education text for Primary-school children in a home life section. Korean Publishing Science Society, Vol.34, No.1, 93-127. (2008)
17. Kim, E. G.: A study for media education practice plan through after-school program in local society. Korea Regional Communication Research Association, Vol. 12, No. 2. 200-239. (2012)
18. Kim, M.O., Kim, J. H., Chung, I. J.: The Mixed Methods Research on the Effect of Mentoring Service Activities Targeted at Low-income Youth on Corporate Employees. Korean Journal of Social Welfare Studies, Vol.46, No.3, 95-123. (2015)
19. Kim, Y. E.: The study of media education in changing media environment. Chung-Ang University Doctor´s thesis. (2001)
20. Kim, Y. R., Bae, S. M., Bark, K. S.: On the Actual Condition of the Smart Phone Use of the Infants and Young Children Perceived by Mothers. Journal of Early Childhood Education & Educare Welfare, Vol.20, No.3, 337-374. (2016)
21. Kim, Y. H.: KISDI STAT REPORT: Analysis of mobile phone ownership and usage behavior of children and adolescents. Korea Information Society Development Institute, Seoul. (2018)
22. Kwon, S. H., Seo, Y. K.: The Theory and Practice of Media Education from an Educational Engineering Perspective, Seoul Hanwul. (2005)
23. Leal, J. P.: Visualization of path patterns in semantic graphs. Computer Science and Information Systems, Vol. 17, No. 1, 229-252. (2020), https://doi.org/10.2298/CSIS180717038L
24. Lee, S. S.: Network Analysis Methodology. Nonhyung, Seoul. (2013)
25. Lee, J. S.: Defining the Internet for Internet Media Education In Elementary School. Media, Gender & Culture, No. 15, 195-238. (2010)
26. Lee, M. J.: Big Data Analytics and Utilization of Public Data. Communications of the Korean Institute of Information Scientists and Engineer. Vol. 30, No. 6, 33-39. (2012)
27. Lee, S. B., Son, Y. G.: Brand Development for the Enhancement of Media Literacy Education. Korea Communications Commission, Seoul. (2017)
28. Lee, S. H., Oh, S. N.: The Effect of Media Utilization in Early Childhood on the Adaptation of Elementary School: Social competence mediated effects. Korean Joural of Children's Media, Vol. 18, No. 2. 89-108. (2019)
29. Lee, Y. J.: A suggestion for using media literacy skills in screen English. STEM journal, Vol. 20, No1, 45-67. (2019)
30. Lim, E. M.:. Re-examine the value of educational technology for the digital generation. The Korean Society for Early Childhood Education & care, (_), 59-77. (2012)
31. Mark, A. B., Douglas, L.: The Importance of 'Big Data': A Definition. Gartner. (2012)
32. Moon, H. S.: Medien padagogik. Korean Broadcasting Institute, Seoul. (2004)
33. Na, K. A., Lee, S. S.: A Study on Media Educational Programs for Youth and the Effects. Korean Journal of Journalism & Communication Studies. Vol. 54, No. 3. (2010)
34. Noh, E. H., Shin, H.J., Lee, J.J., Jeong, H. S.: A Study on the Current Status of Digital Literacy Education in Elementary and Secondary Curriculum and Improvement Plan. Korea Institute for Curriculum and Evaluation, Seoul. (2018)
35. NYPI.: Research on the use of social media to activate youth culture. National Youth Policy Institute, Seoul. (2015)

36. Seo, S. Y.: A study on the development of a Korean classroom class model based on 'MBL: Media Based Learning'. Ministry of Education, Seoul. (2016)
37. Son, M. S.: Problems and improvement plans of self-directed learning for elementary school students. The 32nd Seoul Education Research Paper, Seoul. (2010)
38. Walizer, M.H., Wienir, P.L.: Research methods and analysis: Searching for relationships. Harper & Row, New York. (1978)
39. Yoo, K. J., Kim, M.K., Lee, J. S., Han, M. O.: An analysis on early childhood teacher's awareness on digital equipment, smart equipment and smart e-book. The Journal of Korea Open Association for Early Childhood Education, Vol.18, No.3, 43-70. (2013)
40. Yoon, S. K., Kim, M. J., Choi, J. H.: Effects of Digital Textbook's Interactivity on the Learning Attitude: With a focus on the Tablet PC-based Digital Textbooks of Social Studies. The Korea Contents Society, Vol.14, No. 2, 205-222. (2014)

Su-Jeong Jeong received a Ph.D. in social welfare from Ewha Womans University. She is currently a research professor of creativity & personality laboratory at Tongmyong University. Her main research interests are children and youth welfare, family welfare, social exclusion and media education. Recently, she is actively conducting network analysis research based on big data analysis and complexity theory on areas of interest.

Byung-Man Kim received a Ph.D. in Early Childhood Education from Pusan National University. He is currently a professor of early childhood education at Kyungnam University. His main interests are early childhood education and child care policies, teacher education, creativity/personality and happiness education. Recently, he is actively conducting network analysis research based on big data analysis and complexity theory on areas of interest.

Machine Learning Based Distributed Big Data Analysis Framework for Next Generation Web in IoT

Sushil Kumar Singh[1], Jeonghun Cha[1], Tae Woo Kim[1], and Jong Hyuk Park[1, *]

[1]Department of Computer Science and Engineering,
Seoul National University of Science and Technology,
(SeoulTech) Seoul 01811, Korea
{sushil.sngh001007, ckwjdgns, tang_kim, jhpark1}@seoultech.ac.kr

Abstract. For the advancement of the Internet of Things (IoT) and Next Generation Web, various applications have emerged to process structured or unstructured data. Latency, accuracy, load balancing, centralization, and others are issues on the cloud layer of transferring the IoT data. Machine learning is an emerging technology for big data analytics in IoT applications. Traditional data analyzing and processing techniques have several limitations, such as centralization and load managing in a massive amount of data. This paper introduces a Machine Learning Based Distributed Big Data Analysis Framework for Next Generation Web in IoT. We are utilizing feature extraction and data scaling at the edge layer paradigm for processing the data. Extreme Learning Machine (ELM) is adopting in the cloud layer for classification and big data analysis in IoT. The experimental evaluation demonstrates that the proposed distributed framework has a more reliable performance than the traditional framework.

Keywords: machine learning, big data analysis, extreme learning machine, IoT, security, and privacy.

1.   Introduction

With the fast-growing development of the digital world as next-generation web, IoT (Internet of Things) is adopted in several applications such as smart services, smart communication, smart community, public safety, and many more. Every aspect of our lives combines many things and next-generation web in IoT communication mediums, such as sensor devices, Bluetooth, Wi-Fi, GPRS, etc. [1]. Internet is a fascinating medium for communication, and it is offered in IoT to connect all objects with their automatic features. IoT has the most significant role with advancement applications in future revolutions with next-generation web, and utilization is continuously increased over the coming times. Next-generation web in IoT means to provide distinct requirements, such as smart devices, accuracy, efficient analysis, low energy, and others. According to Keenan et al.'s report [2], the global IoT data management market value at around $27.13 billion in 2017. It would reach approximately $94.47 billion in 2024 and above 19.51 percent between 2018 to 2024 at a CAGR. Nowadays, both Big

---

* Corresponding author

data and IoT technology are growing for the data science field with great attention. Big data is collecting the enormous amount of structured, unstructured, or semi-structured raw data that is more complex to managing and analyzing with many traditional tools. As per the study of Verma et al. [3], around 4.4 trillion GB data will be produced by 2020 using IoT devices in smart applications. IoT is serving as the primary role when any enterprises have a vast number of data for analysis purpose in reinforcements. Big data analytics is a rapidly advancing field for managing and analyzing IoT data. [4, 5]. It is connected to smart devices, which helps to take the initiative to improve decision making. Due to the popularity of online media, including WhatsApp, Instagram, Snapchat, Linkedin, and an expanding number of IoT devices, significant data analysis issues in IoT have been raised in smart technological fields. Big data analytics's main task is to extract useful patterns from the massive amount of IoT data that can be used in decision and prediction making responsibilities. However, many researchers are used various technologies such as edge computing, predictive analytics, Apache Spark, Apache Flink, and so on, which have some challenges according to the next generation web in IoT [6]. The machine learning approach analyzes and extracts accurate data from raw structured, unstructured, and semi-structured data to mitigate these challenges by IoT devices [7].

The existing mechanism to big data analysis in IoT applications for next-generation web on a centralized cloud is not adequately satisfied for specific requirements such as resource management, latency, scalability, accuracy, communication bandwidth. However, with the consecutive growth in data-driven applications in IoT and generated a massive information. In recent years, various machine learning paradigms have been discussed to promote valuable data interpretation for IoT applications [8, 9]. Many operations adapt to data control among various communication devices, and it provides intelligence processing, analysis in IoT. Cloud computing is employed for delivering high performance to the IoT server [10]. Traditional machine learning also has significant issues such as low precision, low rate, low latency, and less computational in the cloud layer. An extreme learning machine (ELM) is utilized, which provides excellent performance to address these issues. It increased the learning speed of feedforward neural networks with a hidden layer of transferring data in the cloud network layer to analyze the IoT data.

Feature extraction and scaling are utilizing for processing the IoT data at the edge layer with address the load balancing, data computation issues on the cloud layer. However, the parts of the device layers with data are moving at the cloud, which reduces the intermediate data computation and processing at the edge layer. [11]. On the different side, edge computing serves as a backbone in the IoT and provides computation power and desired latency to smart applications' IoT devices. To mitigate standard limitations in IoT applications such as high computation, load balancing, network traffic and storage by feature extraction, and scaling at the edge layer to data processing. The device layer delivered massive IoT data, and it is collected from various complex and noisy environments [12]. The edge layer is investigating as a type of feature extraction and scaling-based intelligent computing that could overcome the cloud layer's limitation. Due to data transfer with low network enforcement, the centralized cloud computing layer is becoming inefficient for analyzing and processing a massive amount of data collected from IoT devices. Edge nodes provide efficient storage, computing, and networking services with essential data in IoT applications at the edge layer [13]. The cloud layer has a centralized database with an advanced

analysis of IoT data using extreme learning techniques and transferring these data to other IoT devices. It provides the leading the creation of current data related smart applications. For identifying the object in the collected video data, AlexNet (Convolutional Neural Network) tools are deployed, where we train the machine intelligence network with the help of the Kaggle open dataset. Then we detect the correct image [14].

The existing research studies to big data analysis in IoT applications on a centralized cloud is not adequately satisfied and have some challenges or requirements such as resource management, computational cost, scalability, accuracy, and latency. This paper addresses and discusses the challenges of accuracy, privacy, load balancing, resource limitation, and centralization using the proposed machine learning-based distributed big data analysis framework for the next-generation web in IoT. In this framework, data storing is utilized in the device layer, the edge layer has a data processing part, and big data analysis is completed in the cloud layer. The primary goal of our study is to provide decentralized and secure big data analysis for the next-generation web in IoT by the proposed framework.

To summarize, the main contributions of this paper are as follows:

- Propose Machine Learning-Based Distributed Big Data Analysis Framework for Next Generation Web in IoT, which provides the precise necessities of advanced applications, including accuracy, performance analysis, load balancing, resource limitation, energy consumption, and schedule.
- We deploy the PCA algorithm for feature extraction, K-means algorithm for scaling for processing data in the edge layer, and ELM algorithm for classification and analysis of big data at the cloud layer.
- Evaluate the proposed framework's performance by comparing it with different traditional machine learning classifiers with the NSL-KDD dataset and evaluating performance for various IoT applications such as attack detection and object detection.
- Finally, we graphically analyze the big data with accuracy, testing time, training time, and precision using the ELM algorithm for IoT applications and compare our proposed work with the existing research.

The remains of the paper, we present several existing methods or techniques to big data analysis for the next-generation web in Section 2; Section 3 proposes a Machine learning-based distributed big data analysis framework that introduces an edge computing paradigm for processing the data and ELM based big data analysis in cloud layer for analysis and classification of IoT data and provides principal component analysis algorithm for feature extraction, K-means algorithm for scaling on edge layer and ELM algorithm for classification and analysis of data in cloud layer; in Section 4, we graphical analyze of the proposed framework on the KDDTest+, KDDTest-21 dataset. Finally, we conclude in Section 5.

2.   Seminal Contribution

This section shows the existing research for next-generation IoT with addressing issues such as accuracy, security, latency, energy consumption, centralization, and uses of specific own proposed framework. Li et al. [14] discussed a novel moving strategy to

optimize IoT applications' production with deep learning. It tested the fulfillment of executing various deep learning tasks in the edge computing paradigm for IoT. For identifying objects in the collected data, the AlexNet convolutional neural network is used with six layers. The convolution layer is the first five layers, and the other three are a fully connected layer and completed the feature extraction task. However, it is used only for video data and did not provide a real-world edge computing paradigm. Peng Li et al. [15] proposed a deep convolution computational model (DCCM) and learned hierarchical features of big data in IoT. This model is utilized to extend CNN and improve efficiency by using vector space to tensor space. Chhowa et al. [16] provided a smart proposal for health monitoring in IoT to big data analysis and gave accurate healthcare data on IoT-based system. They focused on a deep learning based IoT system for health monitoring devices and provided efficient results to the various doctors in the IoT environment. It ensures the proper knowledge about the critical patient. However, the number of mammogram devices increases, then the delay is also raised to diagnose disease. Mishra et al. [17] provided a framework related to big data analysis in IoT applications with cognitively oriented infrastructure. It provided implementation architecture for adequate data supervision and information search in manufacturing IoT applications. Zhang et al. [18] described a double projection model with deep computation (DPDCM) to feature big data learning in IoT. Raw input data is separated into two subspaces in the hidden layers to understand the interacted big data feature in IoT applications.

Hosseini et al. [19] utilized a dimensionality reduction technique to improved classification accuracy, reduced communication bandwidth, and computation time of data. Also, they proposed a cloud computing solution for interpretation of big EEG data. However, a large number of training samples problem and heterogeneous data of multidomain propagation are not entirely resolved. Vinay et al. [20] discussed the novel FR approach-based framework on ELM to perform appearance tagging for friendly networks operation on extensive data using machine learning. It is only used for face recognition and has more centralized data, communication bandwidth on cloud problems.Ying et al. [21] proposed an integrated framework to enhance smart city applications' performance by enabling effective orchestration of networking and computing supports. Liu et al. [22] developed a deep learning-based visual food recognition algorithm to achieve the best accuracy of massive data analysis in IoT. It designed edge computing-based paradigms to overcome traditional mobile cloud computing's inherent problems in the food recognition system for IoT. Liangzhi Li et al. [11] adopted state-of-the-art edge computing arrangement to address the crowdsensing problem. It provided distribute deep learning principles to extract characteristics from taken IoT data. It reduced communication costs and increased safety protection of data in IoT. However, it has a compatibility problem for all cloud environments. Jeong et al. [23] present a paradigm to address intrusion detection for various research areas such as image segmentation, security distribution networks, fingerprint matching, human tracking, image watermarking, and big data analysis in IoT.

We are categorized related work in some subsections such as security architecture, technological aspects, methodology, dataset labeling requirement, and conventional machine learning classifier. We are providing Table 6 in Section 4 and compare the proposed framework with existing research studies. Some research used centralized security architecture and smart city technological aspects, but we use distributed security architecture and next-generation web in IoT technical aspects. In summary,

existing studies have used cloud and edge layer frameworks to big data interpretation in advanced applications. However, such a framework and architecture have some limitations on the centralized cloud, such as a massive amount of data, resource utilization, low accuracy and latency, security, and privacy, etc. Moreover, existing strategies use conventional methods and architecture, which requires more computational power and efficiency. It is essential to compose and develop a new framework for big data interpretation in IoT that considers all instant and future difficulties. Thus, we provide a machine learning-based edge computing framework where feature extraction and scaling concepts are used to process IoT data on the edge layer. Also, we are giving ELM based cloud structure for analyzing structured and unstructured big data in IoT efficiently and quickly on the cloud layer.

3.   Proposed Framework for Big Data Analysis in IoT

Based on the limitations such as low efficiency, low latency, centralization, computational cost, resource management of existing studies, we propose a machine learning-based distributed big data analysis framework for the next-generation web in IoT. We focused on processing and analyzing a lot of data on the cloud to develop our framework with ELM based big data analysis in the cloud layer for IoT. We discussed feature extraction and scaling in the edge layer to find specific data with the clustering concept on training data and testing data. We introduce the PCA algorithm for feature extraction, K-means algorithm for scaling, and Naive Bayes algorithm for classification in the edge layer. The proposed framework has various advantages for the next-generation web in IoT, which are the following:

- Provides the distinct requirements of IoT applications such as accurate data, better performance analysis, load balancing, maximum resources.
- With the next-generation web, improves the transparency and connectivity in IoT applications.
- Provide comfortable environments and appropriate high compatible or accurate big data analysis for IoT infrastructure.
- With the proposed framework, improve reliability and efficient operations in IoT applications.

3.1.   Proposed Framework Design

IoT structure is a core technology for connecting smart devices and humans to the internet with a wired or wireless medium. It is known as the Internet of Everything (IoE). It deals with a massive amount of IoT data at the device layer. It produced big data from sensing devices and various IoT applications such as smart services, smart communication, smart community, and many more. The device layer is mainly used for data acquisition and recorded from sensors and IoT applications. Due to increasingly continuous various data sources, accuracy, latency, and trust become a challenge in big data analytics. In this situation, big data analysis in IoT is a very critical issue in the cloud layer. Therefore, this data representation generated different types of big data challenges to extract useful data from unstructured and semi-structured data on the

device layer. To mitigate these problems, we utilize state-of-the-art machine intelligence-based edge computing layer. High connectivity, processing, scaling, and feature extraction capabilities for edge device nodes at the edge layer with machine-intelligence are the primary role of next-generation web in IoT applications and provide efficient operations with accurate big data analysis. Next-generation web connectivity is provided by advanced technologies between the device layer and edge layer.
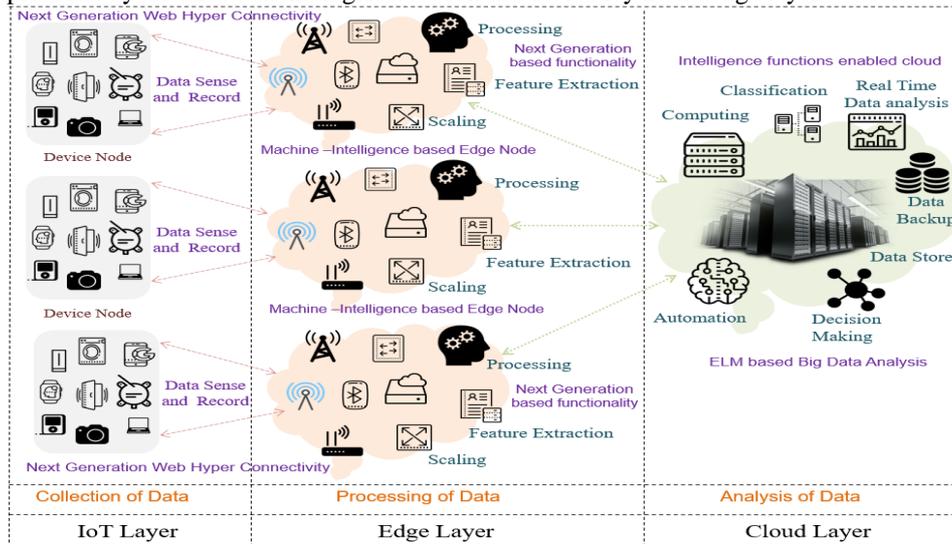

Fig. 1. Architectural overview of the proposed framework.

The edge layer has two next generation-based functionalities, including Feature extraction and scaling for processing of IoT structured and unstructured data. Used data and unused data are categorized to provide data. Feature subset and hidden layer applied to it and found training, testing data set. It used the PCA algorithm for feature extraction, the K-means algorithm for scaling. Every machine intelligence-based edge node has base stations, networking devices, and machine learning, which provide computation power to the physical layer's IoT devices. All edge nodes transfer the extracted data to the cloud layer with a base station and networking device such as a router. Thus, load balancing and energy efficiency issues resolve in the edge and cloud layer using machine-intelligence-based feature extraction and scaling. However, the cloud layer has one data center, so accuracy, speed, computational storage is very low. Therefore, the ELM algorithm is used to classify and analyze data in the cloud layer's proposed framework. It examines the data and improved performance, accuracy, latency, and efficiency of IoT data. The overview architecture for the proposed framework is as shown in Fig. 1.

Table 1. Abbreviation table

Abbreviation	Description	Abbreviation	Description
Z	Dimensional linear Subspace	$\{S_1, S_2, ... S_k\}$	Cluster data center sets.
$\{a_1, a_2 ..., a_M\}$	Labeled or Unlabelled data sets	K	Number of clusters or groups
$\{b_1, b_2 ..., b_M\}$	Projected data sets	α_i	Output weight of a hidden node
w_i	Weight vector or eigenvector value	$H(p)$	ELM hidden layer output
X	Input vector Matrix	M	Training sample
Y	Output vector Matrix	H	Matrix hidden layer output
β_{mk}	Binary indicator variable sets.	T	Training data set a target matrix

3.2.   Functional Components of the Proposed Framework

This subsection presents the main functional component of the proposed framework. It is divided into three parts, including data acquisition or collection, data processing, and data analysis. Data acquisition is used in the IoT layer, data processing is completed at the edge layer, and data analysis is utilized in the cloud layer. PCA algorithm used for feature extraction, K-means algorithm utilized for scaling data processing at the edge layer, and the cloud layer have an ELM algorithm for analyzing massive data in IoT applications. Machine-Intelligence based distributed big data analysis flow is shown in Fig. 2. (a). The principal component analysis (PCA) algorithm is utilized for feature extraction, the K-means algorithm is used for scaling, and the ELM algorithm for classification and analysis of data. ELM-based classification and data analysis on the cloud layer are shown in Fig. 2. (b). The abbreviation table is shown in Table 1.

Data Collection: Data collection is an essential function for the proposed framework, and it is used in the device layer. Various IoT applications such as smart services, smart communication, smart community, and others generated a massive amount of IoT devices. IoT devices are connected to various sensors and detect the data in several forms, including video, audio, multimedia, etc. It can measure and monitor the data in real-time. Various data types are stored in a device layer such as automation, location, streaming, status data, and others. Internal sensors collect IoT data from consumer devices such as smart appliances, smart televisions, wearable health meters, and commercial devices such as traffic monitoring systems, weather forecasting, and commercial security system.
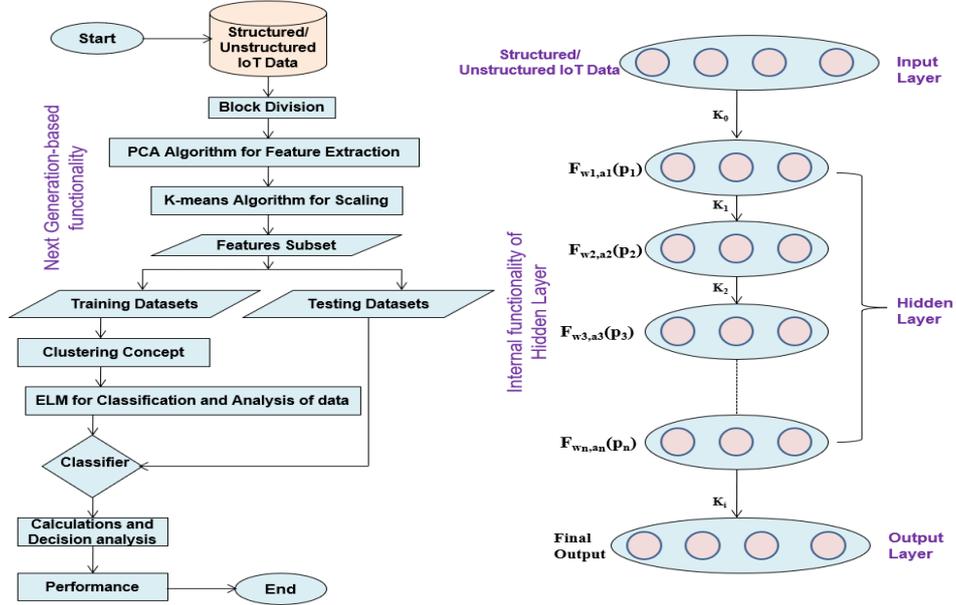
Fig. 2. **(a)** Flow of Machine Learning-based Distributed Big Data Analysis Framework
(b) ELM-based Classification and Analysis of Data on Cloud Layer

Data Processing: Data processing is the second functional component of the proposed architecture used in the edge layer. The IoT sensor devices must generate an extensive quantity of data that must be processed before the IoT information can be utilized. However, these data come from IoT devices in various types of formats. To mitigate this problem, feature extraction and scaling concepts are used to process IoT data at the edge layer. For data processing, PCA and K-means algorithms are utilized. a' is average value.

Algorithm 1. Principal Component Analysis

Input Dimensional linear subspace is Z and the input vector of a labeled or unlabelled data set is $\{a_1, a_2, \ldots \ldots a_M\}$.

Output: Projected data set $\{b_1, b_2, \ldots \ldots b_M\}$ and weight vectors $\{w_j\}$ which have essential subspace form.

Process:

1: $a' = \frac{1}{M} \Sigma_m\ a_m$ /* M training samples and a_m input vector

2: $S = \frac{1}{M} \Sigma_m\ (a_m - a'_m)(a_m - a'_m)^T$ /*$a'_m = \Sigma_{j=1\ to\ R} b_{mj}\ w_j + a'$

   Eigenvector value is $\{w_j\}$ of S.

3: **for** all m = 1 to M

4:    **do**

5:       **for** all j= 1 to Z

6:          **do**

7:             $b_{mj} = (a_m - a'_m)^T w_j$

8:          **end**

9:    **end**

Feature Extraction

It is one of the primary next-generation web function for machine learning to identify strong and weak relevant features and labels for IoT applications. It is used mainly for load balancing, and improving the communication bandwidth, time delay in IoT applications. Due to the continuously increasing massive amount of data, it is characterized into three groups: big data, data usage, data quality. Big data is also characterized as volume, variety, velocity, and veracity of IoT data. Completeness, noiseless, semantics are the part of data usage, and accuracy, redundancy, efficiency, loading balancing are the part of data quality. Therefore, feature extraction methods are using on the edge layer. IoT devices have many types of sensors, analyzed, monitoring data. Processing system components and communication protocols in IoT is divided into three parts: device to device, the device to the server, and server to server. While many feature extraction algorithms are available, but we are using the PCA algorithm. This algorithm's main objective is to extract the relevant features from various IoT structured and unstructured data sets, which is stored in the device layer of the proposed model. Feature extraction function discriminates the essential and useful features by eliminating redundant features and noise from IoT applications and provide the best-predicted output features of IoT data. The extracted relevant and non-relevant features can help us to identify new useful information that is used in machine learning.

From a machine learning perspective to IoT data, Feature extraction is an essential concept for processing the data on the edge layer. Firstly, IoT data is divided into variable or fixed blocks, then feature extraction and scaling concepts are used. Training sets have various inputs as samples, which are used by machine learning. Supervised learning utilized input vector and related output vector (labels); these all are samples. Vectors (labels) are not required for unsupervised learning. Reinforcement learning is used to understand to expropriate steps to be taken for a specific condition. Determine the main groups (cluster) between comparable sample clusters that comprehend as clustering. Original input samples transferred into a new variable sample or space are called feature extraction and improves the result.

PCA is the most straightforward algorithm for feature extraction of data set in IoT applications, and it is based on actual eigenvalues of IoT data. The PCA algorithm's main objective is to overcome the overfitting problem and decrease the dimensionality of IoT data sets. It may be large or less, while retaining the variation present in the dataset, up to the maximum extent [24]. According to principal component analysis (PCA), the principal subspace has the orthogonal data points. It is the property of PCA and has the maximum projected variance of data. The main task defined as the finding the complete Z orthogonal data set, it is based on M-dimensional vectors $\{w_j\}$. Parallelly find the corresponding linear projections data points $\{Y_{n_j}\}$ and minimized the reconstruction error. Average of all data points is a'.

Assume that the given training set have M training samples dented as (a_i, b_i) where i = 1,2,3……..M, a_i is training N-dimensional input vector, b_i is corresponding desired P-dimensional output vector. $X = (a_1, a_2, \ldots \ldots a_M)^T$ is input vector matrix and $Y = (b_1, b_2, \ldots \ldots b_M)^T$ is corresponding output vector matrix.

$$B = \frac{1}{M} \Sigma_m (a'_m - a_m)^2 \qquad\qquad (1)$$

$$a'_m = \Sigma_{j=1 \text{ to } R} b_{mj} w_j + a' \qquad\qquad (2)$$

According to algorithm 1, provide the concept of principal component analysis, and overcome the overfitting problem, extract the features of IoT data in processing. PCA has various versions; it is based on data size. It may be a structured dataset or unstructured dataset. Due to calculation of $\{w_1, w_2, \ldots \ldots w_z\}$, algorithm have different run times values.

For preprocessing, the principal component analysis algorithm is an essential method in machine learning. In preprocessing for PCA, find the features and labels in the processing of IoT data. Various practical applications are involved for PCA such as data compression, data visualization, face recognition, image rendering, and so on.

Data Scaling

Data Scaling is another next-generation web task for processing the IoT data at the edge layer. Scaling is the task of dividing the data set value into several specific groups. The data sets a value in the same groups that are more similar to other data sets value in the same group than those in other groups. The main objective of scaling is to segregate data set groups with the same properties. K-means algorithm is used for scaling the IoT data where identifies a similar type of data set values in a group [25]. It gives the groups (K) which is related to each other.

For machine learning, the K-means algorithm is one of the simplest and popular for clustering. It does not have labels or results in data processing, so it is called unsupervised learning [10, 26]. The K-means algorithm's primary idea is to group (cluster) related data sets values and recognize underlying designs. Based on algorithm, find a set of K group clusters $\{S_1, S_2, \ldots . S_k\}$. It reduces the distance between data features values and the most adjacent data hub. To denote data points value to the group hubs, then apply binary pointer variables βmk ∈ {No, Yes} or {0, 1}. We formulate dilemma as regards with the equation:

$$Minimize \Sigma_{m=1 \text{ to } M} \Sigma_{k=1 \text{ to } K} \beta mk (a_m - s_k)^2 \qquad (3)$$

Where,

$$\beta mk (a_m - s_k)^2 = 1 \text{ for data point } a_m \text{ belong to cluster, otherwise } \beta mk$$
$$= 0 \text{ and } m = 1, 2, 3, \ldots \ldots \ldots \ldots . M$$

Minimization has two sections: 1) Distance minimize concerning β_{mk} and s_k stable; 2) Distance minimize concerning s_k and β_{mk} is stable [10].

Genetic clustering methods can predict the movement of points is known as k-means. It is used for various IoT applications, but this method has some considerations, such as low efficiency and communication bandwidth compared to Euclidian distance. For classifying the IoT data, it is a successful machine learning approach. It may lead to unsuitable clusters in some cases. With the help of a 1-nearest neighbor classifier, find new data value in the existing clusters. Algorithm 2 describes how to get the optimal group datacenters S_k and the authorization of the data points β_{mk} [10].

Algorithm 2. K-Means [10]

Input: The number of same data points groups is K and unlabelled data sets are $\{a_1, a_2, ... a_M\}$.
Output: $\{s_k\}$ is cluster data center and $\{\beta mk\}$ is assign data points randomly initiate with $\{s_k\}$
Process:
1: Repeat
2:    **for** $\forall(m = 1$ to M$)$
3:        **do**   /* m=1,2,3………………...M
4:          **for** $\forall$ (k = 1 to K)
5:            **do**
6:              **if** $k = \arg\min_i (s_i - x_i)^2$
7:                **then** $\beta mk = 1$ /* βmk is set of binary assigned data points
8:              **else** $\beta mk = 0$
9:            **end**
10: **end**
11: **for** $\forall$ (k = 1 to K)
12:   **do**
13:     $s_k = \Sigma_{m=1\ to\ M}\ a_m\ \beta mk / \Sigma_{m=1\ to\ M}\ \beta mk$ /* s_k is set of K cluster center
14:   **end**
          while confluence adjust $\{\beta mk\}$ or $\{s_k\}$
15: **end procedure**
16: Return $\{a_1, a_2, ... a_M\}$

Data Classification and Analysis

Data classification and analysis are the final next-generation web function of the proposed framework and used with the ELM algorithm's help in the cloud layer. The ELM algorithm is used to analyze and classify the IoT data to address centralization and data handling issues with feedforward neural networks [27]. The relationships between the input and the hidden layer are randomly allocated and continue uninterrupted during the ELM algorithm's training method [28]. It includes two steps in the learning phase: 1) creating the hidden layer output model and 2) find the output combinations. Then, the output combinations are tuned by reducing the cost function using a linear system. By this system, the computational weight of the ELM is continuously decreased in IoT applications. The low computational weight or complexity are used for evaluation result in machine learning and utilized in high dimensional and large-data applications. To mitigate energy consumption and the massive amount of data in IoT, ELM Algorithm is utilized in the proposed framework. ELM has resolved various classification problems because it gives more excellent efficiency for handling the massive amount of data. ELM classifier adopts the hidden connections for classification, and the hidden connection output used as a sigmoid activation formula $q(y) = 1/(1 + e^{-y})$ to evaluate the output value [29].

Suppose that the output formula of the ith hidden node is $h_i(p) = J(a_i, b_i, p)$, where (a_i, b_i) is the hidden connection parameter in the given single hidden layer of ELM. It is the basic method of the ELM algorithm for hidden layer feedforward neural interfaces. The ELM is an algorithm, and it is used for addressed single hidden layer neural interface problems and provide several hidden layers. With the use of the ELM algorithm, we can easily analysis of big data in various IoT applications.

The output function of ELM with Z hidden nodes

$f_Z(p) = \Sigma_{i=1\ to\ Z}\ \alpha_i h_i(p)$, Where α_i is output weight of ith hidden node

$h(p) = [J(h_i(p), ……….. h_Z(p))]$ is the ELM hidden layer output.

If M given the training samples, then the ELM hidden layer output formula is given as:

$$H = \begin{bmatrix} h(p_1) \\ h(p_2) \\ \vdots \\ \vdots \\ h(p_M) \end{bmatrix} = \begin{pmatrix} J(a_1, \ b_1, p_1) & \cdots & J(a_Z, \ b_Z, p_1) \\ \vdots & \ddots & \vdots \\ J(a_1, \ b_1, p_M) & \cdots & J(a_Z, b_Z, p_M) \end{pmatrix} \tag{4}$$

Training data set targetmatrix

$$T = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ \vdots \\ t_M \end{bmatrix} \tag{5}$$

ELM is a regulation neural network. However, hidden layer mapping formed by both random hide nodes and its objective function is as follows:

$$Minimize: (\alpha)_r^{\tau 1} \ + \ C(H\alpha - T)_g^{\tau 2} \tag{6}$$

$$\text{Where } \tau_1 > 0, \tau_2 > 0 \ and \ r, g = 0, \frac{1}{2}, 1, 2, \dots\dots\dots\dots \infty$$

We can use a different combination of $\tau_1, \tau_2, r, g,$ and find different results in various learning algorithms for regression, classification, compression, clustering, and, others. With ELM, computes the hidden layer output formula using training and classification.

Input: Training set F= {a$_1$, a$_2$.......a$_M$} with class variable Q = {Q$_1$, Q$_2$,.............Q$_M$}, representation of hidden connections Z and anonymous testing examples Fu = {d$_1$, d$_2$..............d$_i$}

Training:
- Assign the input weight {w$_1$, w$_2$,................w_M} and G= {b$_1$, b$_2$.........b_M }^{T}
- Compute the hidden layer output formula H= f(w. F + G)
- Compute the output weight H$^+$. Q.

Classification:
- Compute hidden layer output formula of new instances Hu = f (w. Fu + G).
- Find the class label of new examples of Fu: Qu = Hu. h.

We solve the single hidden layer neural network and classify IoT data problems using the ELM algorithm at the cloud layer of the proposed framework. It provides efficient performance for handling the massive amount of data and easily analyzing the big data in IoT applications.

4.   Experimental Evaluation

This section discusses the experimental evaluation part to evaluate the proposed framework's adequate performance with high accuracy and low latency. We employed the NSL-KDD dataset consisting of sample events associated with five big data analysis classes, as presented in Table 2. We used KDDTest+, KDDTest-21, and KDDTrain+ dataset and disposed of it in a real-time infrastructure. The details and results of our algorithm are described in the subsection, and significant data analysis methods with machine learning and IoT Applications shown in Table 3 [30].

Table 2. Big Data Analysis Categories of IoT Applications.

Real_Time	Offline	Database Level	Manufacturing Level	Large Level
GreenPlum	Skribe	MongoDB	Data Analysis Plan	MapReduce
HANA	Kafka	TB-Level Data	Distributed File	Scala
Parallel Processing	TumeTunnel		TB-Level Data	
Memory Based	Chukwa			

Evaluation Methodology

Weka [31] tool is used to estimate the experimental analysis of our proposed framework for big data analysis. This tool is mainly used for data tunneling, evaluating the analysis model using various machine learning models. Every data occurrence in the dataset has 41 input qualities, and these are categorized into four types: essential attributes, content attributes, traffic attributes, and host-based attributes. Traffic attributes are time-based, which are extracted from traffic by utilizing different types of windows. 10-fold cross-validation technique, we can apply [31,32] across the assembled NSL-KDD dataset and use two classes normal and anomaly as shown in the confusion matrix Table 4. 10 equal size subsets data are offered with the help of a fold cross-validation approach. In this division, training use nine subset data, and testing use one data set.This process is returned but has one situation that 10 datasets hold as the testing set exactly one time. For ELM classifier, the sigmoidal formula is offered as a hidden outcome operation. We used 50 simulations for the ELM algorithm with some parameter numbers on training data. A big data analysis report is evaluated with a confusion matrix. We practiced a productive workstation with Processor E5-1620 v3 (30 MB, 3.70Ghz processor rate), and the bandwidth specifications have been promoted from 1GE-100 GE.

Table 3. Big Data Analysis Methods and IoT Applications

Methods Applications	Classification	Clustering	Association Rule	Prediction	Time Series	Proposed ELM Method
Social Networking	-	√	-	√	√	√
Bioinformatics	-	√	√	-	-	√
Healthcare	-	√	√	-	-	√
Transportation	-	√	√	-	-	√
Market Analysis	-	√	√	√	-	√
Disaster Management	-	-	-	√	√	√
Speech Recognition	√	-	-	-	√	√
e-governance	√	√	√	-	√	√
Industry Management	√	√	√	-	-	√
Human Genetic	-	√	-	-	-	√
Medical Imagine	√	√	-	-	√	√

Table 4. Confusion Matrix [33]

Actual	Classified	
	Normal	**Anomaly**
Normal	T_p	F_n
Anomaly	F_p	T_n

Where, T_p is quantity of the normal profiles correctly classified as normal profiles as true positive,

F_p is number of anomaly profiles incorrectly classified as normally once as false-positive,

F_n is indicates the quantity of normally profiles incorrectly classified as anomaly once as false-negative,

T_n is the number of anomaly profiles correctly classified as anomaly once as true negative.

Confusion matrix, which is used for evaluating the performance of the proposed framework for big data analysis and used the accuracy, false positive rate, precision, recall, F-measure, MCC, and AUC formulas or equation from 7 to 12 [33].

a)   Accuracy (True Positive Rate_ACC):

$$Accuracy = \frac{Tp+Tn}{Tp+Fn+Tn+Fp} \tag{7}$$

b)   False Positive Rate (FPR):

$$FPR = \frac{Fn+Fp}{Tp+Fn+Tn+Fp} \tag{8}$$

c)   Precision:

$$Precision = \frac{Tp}{Tp+Fp} \tag{9}$$

d)   Recall (Detection_Rate):

$$Detection_{Rate} = \frac{Tp}{Tp+Fn} \tag{10}$$

e)   F-Measure:

$$F-Measure = \frac{2*Precision*Detection\ Rate}{Precision+Detection\ Rate} \tag{11}$$

f)   MCC (Mathew Correction Coefficient):

$$MCC = \frac{Tp*Tn-Fp*Fn}{\sqrt{(Tp+Fn)(Tp+Fp)(Tn+Fp)(Tn+Fn)}} \tag{12}$$

g) The area under the receiver operation curve (AUC) shows the true and false positive sample rate.

4.1.   Proposed Framework Evaluation

This subsection shows of our framework's evaluation. We applied KDDTest+ and KDDTest-21 datasets and following different machine learning classifiers: Naive Bayes

(NB) [34, 35], Logistic Regression (LR) [36, 37], Jrip (JR) [38], J48 Decision Tree (J48) [39], LMT Decision Tree (LMT), Random Forest (RF), Support Vector Machine (SMO) [40-42], K-Nearest Neighbors (IBK) [43, 44]. All classifier machine learning methods are notified in Table 5. We used the classification method with the ELM algorithm for classification, K-means algorithm for scaling, and PCA algorithm for feature extraction (proposed in subsection 3.2). The 10-fold cross-validation methods displayed accuracy, FPR, precision, detection rate, F-measure, and others over the NSL-KDD dataset With the NSL-KDD data set, we trained 8 different machine learning classifiers in section 4, then utilized 10-fold cross-validation methods to estimate the results. Fig. 3 and Fig. 5 summarize all classifiers' performance results regarding standard evaluation standards as beforehand reported.

According to Fig. 3 and Fig. 4 for the KDDTest+, KDDTest-21 dataset, all machine learning classifiers have a perfect classification capability to execute. According to the proposed framework's performance on the KDDTest+ dataset, it observed accuracy, FPR, precision value, F-measure is the highest of NB classifier compared to others.NB algorithm has the accuracy 98.8% of the proposed framework on the KDDTest+ dataset and 97.8% on the KDDTest-21 dataset. The latency time is 0.01 sec of the K-Nearest Neighbors algorithm of the KDDTest+ dataset's proposed framework. It is smaller than other; 0.08 sec is the NB algorithm's latency time of the proposed framework on the KDDTest-21 dataset.

There is a slightly different of around 0.1 in detection_rate for the classifier NB, J48, and LMT. Furthermore, the RF and NB classifier obtained a similar value of MCC and AUC. Time taken time for making the model is less for IBK classifier compares to others. According to the proposed framework's performance on the KDDTest-21 dataset, the NB classifier's accuracy is similar to RF classifier and FPR, precision, detection_rate, F-measure, and MCC value is highest of NB classifier compares than others. Time taken time for making the model is less for the NB classifier compares to others. FPR, Precision, detection_rate, F-Measure and MCC value of NB algorithm for proposed framework on KDDTest+ dataset is 15.8%, 98.6%, 98.5%, 98.7%, and 97.4%, it is also greater than another algorithm. Similarly, FPR, Precision, detection_rate, F-Measure, and MCC value of the NB algorithm for the proposed framework on KDDTest+ dataset are 31.5%, 95.6%, 94.0%, 93.9%, and 92.7%; it is also greater than another algorithm.
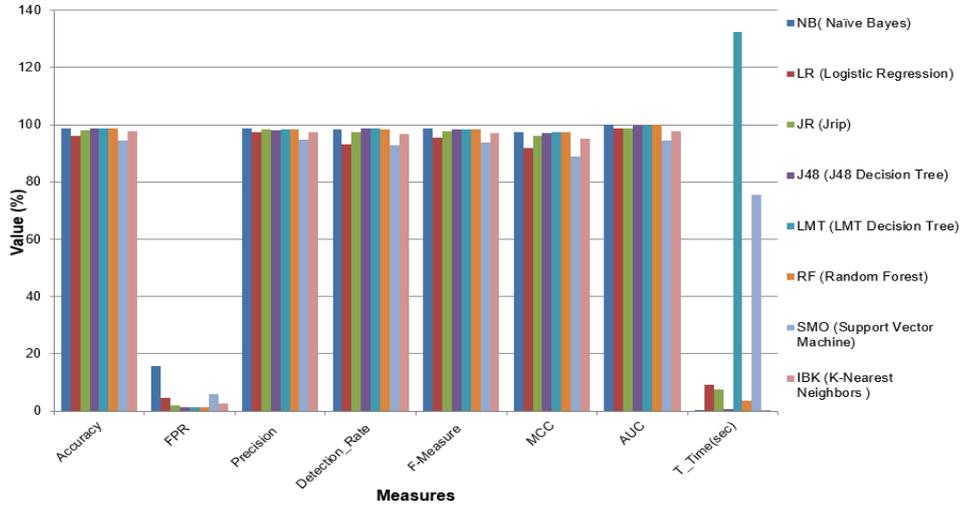
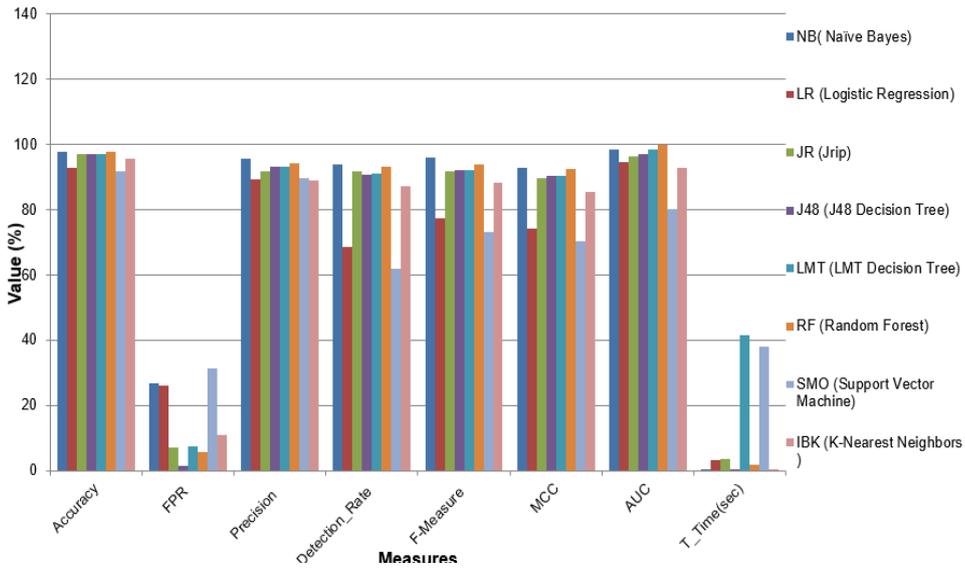**Fig. 3.** Proposed framework's evaluation performance on KDDTest+.

Fig. 4. Proposed framework's evaluation performance on KDDTest-21.

Table 5. Proposed framework's execution for Several IoT applications

Application	Evaluation measure		
	Dataset	Accuracy (%)	Latency(sec)
Attack Detection	KDDTest+	86.53	0.011
	KDDTest-21	75.77	0.013
Object Detection	MNIST	84.2	0.057
	MS-COCO	78.32	0.048

The performance of the proposed framework of IoT applications, such as attack detection and object detection, is shown in Table 5. Attack detection evaluated on KDDTest+ and KDDTest-21 data set. Accuracy is 86.53% on KDDTest+ and 75.77% on KDDTest-21 data set. Latency time is 0.011sec on KDDTest+ dataset and 0.013sec on KDDTest-21 data set. Object detection was evaluated on MNIST and MS-COCO datasets. Accuracy is 84.2% on MNIST and 78.32% on the MS-COCO data set. Latency time is 0.057sec on the MNIST dataset and 0.048sec on the MS-COCO data set. Comparison with existing research is shown in Table 6 with methodology, security architecture, technology aspects, dataset labeling requirements, and conventional machine learning classifiers.

Table 6. Comparison with existing research studies

Research Work	Year	Security Architecture	Technology Aspects	Methodology	Requirement of dataset labeling	Conventional Machine Learning Classifier
Li et al. [11]	2018	Centralized	IoT	The offloading approach with edge computing is used to optimize the performance of IoT applications	High	No
Peng Li et al. [12]	2018	Cluster-based architecture	IoT	A deep convolutional computational model for big data features learning using tensor representation model	High	Yes
Chhowa et al. [13]	2019	Centralized	IoT	Deep Machine learning for big data analysis	Low	No
Mishra et al. [14]	2015	Centralized	IoT	Cognitive Oriented IoT big data framework	Low	Yes
Zhang et al. [15]	2018	Centralized with BGV encryption	IoT	Double projection deep computational model for bigdata feature learning	High	Yes
Hosseini et al. [16]	2016	Centralized	Epileptic Seizure Prediction	Deep learning with cloud for Epileptic Seizure Prediction	Low	No
Vinay et al. [17]	2015	Centralized	Social Network	For face identification, a cloud-based big data analytics framework is used	Low	Yes
Ying et al. [18]	2017	Centralized	Smart City	Deep reinforcement learning approach with SDN and mobile edge computing	High	No
Liangzhi Li et al. [8]	2018	Centralized	IoT	Distributed deep model for mobile crowdsensing	High	Yes
Singh et al. [39]	2020	Distributed	Smart City	Provide IoT oriented infrastructure for the smart city based on deep learning and blockchain	Medium	No
Proposed work	2020	Distributed	Next Generation Web in IoT	Machine learning-based distributed big data analysis framework for Next Generation Web in IoT	Less	No

However, the proposed work encourages a distributed framework for big data analysis. It is suitable for advanced applications. It needs less labeled data to promote big data interpretation with precision and time detection in advanced applications. It provides higher execution over another significant data analysis time detection, and accuracy. Table 3 shows big data analysis methods and IoT applications via machine learning methods such as classification, clustering, association rule, prediction, and time series [45-47]. Proposed ELM methods are used in maximum IoT applications such as social networking, bioinformatics, smart energy, smart home, e-government, and others

compare to other machine learning methods [48-52]. ELM based framework on the cloud layer provides excellent performance at a high data rate.

5. Conclusion

This paper proposed a machine learning-based distributed big data analysis framework for Next Generation Web in IoT. It relies on the edge layer with feature extraction and scaling and cloud layer with the ELM algorithm to facilitate real-time big data analysis and classification in IoT. The PCA algorithm used for feature extraction, the K-means algorithm for scaling, and the ELM algorithm for classification. Analysis of big IoT data and provides a distributive capability wherein extract, scale, and classified IoT data features at the edge layer and improved data fulfillment. Furthermore, the recommended approaches solve big data analysis issues by combining the feature extraction and scaling algorithms. The experimental evaluation on the KDDTest+ and KDDTest-21 dataset determines that the framework realized excellent performance 0.06sec time taken for making the model, 98.8% accuracy rate for KDDTest+ dataset, and 0.08sec time taken for making the model, 97.8% accuracy rate for the KDDTest-21 dataset. Comparing the proposed machine learning-based big data analysis framework approach with the conventional machine learning classifier shows that it can manage label data issues and attain better appearance to the other classifier. The ELM-based method was used in the cloud layer and improved the proposed framework's performance and big data analysis issues easily using a distributed cloud. Finally, we compare the proposed framework with existing research according to methodology, security architecture, technology aspects, dataset labeling requirements, and conventional machine learning classifiers.

We have two ways according to our proposed framework for the future. First, we expect to explore other improvements in our frameworks' advanced applications, such as using machine learning models to eradicate manual feature extraction and scaling for big data analysis. Second, we intend to construct our NSL-KDD dataset to developing a common framework across big data analysis.

References

1. Zantalis, F., Koulouras, G., Karabetsos, S., Kandris, D.: A Review of Machine Learning and IoT in Smart Transportation. Future Internet, Vol. 11, No. 4, 94. (2019)
2. Keenan, M: The Future of Data with the Rise of the IoT. [Online]. Available:https://www.rfidjournal.com/articles/view?17954(Accessed in October 2018)
3. Verma, A.: Internet of Things and Big Data- Better Together. [Online]. Available: https://www.whizlabs.com/blog/iot-and-big-data/ (Accessed on August 2018)
4. Li, Y., Gao, H., Xu, Y.: Special Section on Big Data and Service Computing. Intelligent Automation and Soft Computing, Vol. 25, No. 3, 511-512. (2019)

5. Zhang, Y., Wang, Y. G., Bai, Y. P., Li, Y. Z., Lv, Z. Y., Ding, H. W.: A New Rockburst Experiment Data Compression Storage Algorithm Based on Big Data Technology. Intelligent Automation and Soft Computing, Vol. 25, No. 3, 561-572.(2019)

6. Liu, M., Cheng, L., Qian, K., Wang, J., Liu, Y.: Indoor acoustic localization: a survey. Human-centric Computing and Information Sciences, Vol. 10, No. 1, 2. (2020)

7. Singh, S.K., Rathore, S., Park, J. H.: BlockIoTIntelligence: A Blockchain-enabled Intelligent IoT Architecture with Artificial Intelligence, Future Generation Computer Systems, Vol. 110, 721-743. (2019)

8. Abeshu, A., &Chilamkurti, N.: Deep learning: the frontier for distributed attack detection in fog-to-things computing. IEEE Communications Magazine, Vol. 56, No. 2, 169-175. (2018)

9. Park, J. H., Salim, M. M., Jo, J. H., Sicato, J. C. S., Rathore, S., & Park, J. H.: CIoT-Net: a scalable cognitive IoT based smart city network architecture. Human-centric Computing and Information Sciences, Vol. 9, No.1, 29. (2019)

10. Singh, S. K., Salim, M.M., Cha, J., Pan, Y., Park, J. H.: Machine Learning-Based Network Sub-Slicing Framework in a Sustainable 5G Environment. Sustainability, Vol. 12, 6250. (2020)

11. Li, L., Ota, K., Dong, M.: Human in the Loop: Distributed Deep Model for Mobile Crowdsensing. IEEE Internet of Things Journal, Vol. 5, No. 6, 4957-4964. (2018)

12. Mohammadi, M., Al-Fuqaha, A., Sorour, S., Guizani, M.: Deep learning for IoT big data and streaming analytics: A survey. IEEE Communications Surveys & Tutorials, Vol. 20, No.4, 2923-2960. (2018)

13. Hu, L., Ni, Q.: IoT-driven automated object detection algorithm for urban surveillance systems in smart cities. IEEE Internet of Things Journal, Vol. 5, No. 2, 747-754. (2017)

14. Li, H., Ota, K., & Dong, M.: Learning IoT in edge: Deep learning for the Internet of Things with edge computing. IEEE Network, Vol. 32, No. 1, 96-101. (2018)

15. Li, P., Chen, Z., Yang, L. T., Zhang, Q., Deen, M. J.: Deep convolutional computation model for feature learning on big data in Internet of Things. IEEE Transactions on Industrial Informatics, Vol. 14, No. 2, 790-798. (2017)

16. Chhowa, T. T., Rahman, M. A., Paul, A. K., Ahmmed, R.: A Narrative Analysis on Deep Learning in IoT based Medical Big Data Analysis with Future Perspectives. In 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE) 1-6. (2019)

17. Mishra, N., Lin, C. C., Chang, H. T.: A cognitive adopted framework for IoT big-data management and knowledge discovery perspective. International Journal of Distributed Sensor Networks, Vol. 11, No. 10, 718390. (2015).

18. Zhang, Q., Yang, L. T., Chen, Z., Li, P., Deen, M. J.: Privacy-preserving double-projection deep computation model with crowdsourcing on cloud for big data feature learning. IEEE Internet of Things Journal, Vol. 5, No. 4, 2896-2903. (2017)

19. Hossain, B., Morooka, T., Okuno, M., Nii, M., Yoshiya, S. and Kobashi, S.: Surgical Outcome Prediction in Total Knee Arthroplasty using Machine Learning. Intelligent Automation and Soft Computing, Vol. 25, No. 1, 105-115. (2019)

20. Vinay, A., Shekhar, V. S., Rituparna, J., Aggrawal, T., Murthy, K. B., & Natarajan, S.: Cloud-based big data analytics framework for face recognition in social networks using machine learning. Procedia Computer Science, Vol. 50, 623-630. (2015)

21. He, Y., Yu, F. R., Zhao, N., Leung, V. C., Yin, H.: Software-defined networks with mobile edge computing and caching for smart cities: A big data deep reinforcement learning approach. IEEE Communications Magazine, Vol. 55, No. 12, 31-37. (2017)

22. Liu, C., Cao, Y., Luo, Y., Chen, G., Vokkarane, V., Yunsheng, M., Hou, P.: A new deep learning-based food recognition system for dietary assessment on an edge computing service infrastructure. IEEE Transactions on Services Computing, Vol. 11, No. 2, 249-261. (2017)

23. Jeong, Y. S., Park, J. H.: Advanced Big Data Analysis, Artificial Intelligence & Communication Systems. Journal of Information Processing Systems, Vol. 15, No. 1. (2019)

24. Mahdavinejad, M. S., Rezvan, M., Barekatain, M., Adibi, P., Barnaghi, P., Sheth, A. P.: Machine learning for Internet of Things data analysis: A survey. Digital Communications and Networks, Vol. 4, No. 3, 161-175. (2018)

25. Tang, Z., Liu, K., Xiao, J., Yang, L., Xiao, Z.: A parallel k-means clustering algorithm based on redundance elimination and extreme points optimization employing MapReduce. Concurrency and Computation: Practice and Experience, 29(20), e4109. (2017)

26. Sun, W., Du, H., Nie, S., He, X.: Traffic Sign Recognition Method Integrating Multi-Layer Features and Kernel Extreme Learning Machine Classifier, Computers, Materials & Continua, Vol. 60, No. 1, 47-161. (2019)

27. Cao, J., & Lin, Z.: Extreme learning machines on high dimensional and large data applications: a survey. Mathematical Problems in Engineering, (2015). https://doi.org/10.1155/2015/103796

28. Duan, M., Li, K., Liao, X., Li, K.: A parallel multiclassification algorithm for big data using an extreme learning machine. IEEE transactions on neural networks and learning systems, Vol. 29, No. 6, 2337-2351. (2017)

29. Akusok, A., Björk, K. M., Miche, Y., Lendasse, A.: High-performance extreme learning machines: a complete toolbox for big data applications. IEEE Access, Vol 3, 1011-1025. (2015)

30. Marjani, M., Nasaruddin, F., Gani, A., Karim, A., Hashem, I. A. T., Siddiqa, A., Yaqoob, I.: Big IoT data analytics: architecture, opportunities, and open research challenges. IEEE Access, Vol. 5, 5247-5261. (2017)

31. Rathore, S., Loia, V., & Park, J. H.: SpamSpotter: An efficient spammer detection framework based on an intelligent decision support system on Facebook. Applied Soft Computing, Vol. 67, 920-932. (2018)

32. Luo, M., Wang, K., Cai, Z., Liu, A., Li, Y., Cheang, C.F.: Using Imbalanced Triangle Synthetic Data for Machine Learning Anomaly Detection, Computers, Materials & Continua, Vol. 58, No. 1, 15-26, (2019)

33. Rathore, S., Park, J. H.: Semi-supervised learning based distributed attack detection framework for IoT. Applied Soft Computing, Vol. 72, 79-89. (2018)

34. Zhuang, X., Zhou, S.: The Prediction of Self-Healing Capacity of Bacteria-Based Concrete Using Machine Learning Approaches, Computers, Materials & Continua, Vol. 59, No. 1, 57-77. (2019)

35. Ma, T., Pang, S., Zhang, W., Hao, S.: Virtual Machine Based on Genetic Algorithm Used in Time and Power Oriented Cloud Computing Task Scheduling. Intelligent Automation and Soft Computing, Vol. 25, No. 3, 605-613. (2019)

36. Lin, J., Yin, J., Cai, Z., Liu, Q., Li, K., Leung, V.: A secure and practical mechanism of outsourcing extreme learning machine in cloud computing. IEEE Intelligent Systems, Vol. 28, No. 6, 35-38. (2013)

37. Peng, F., Zhou, D. L., Long, M., Sun, X. M.: Discrimination of natural images and computer-generated graphics based on multi-fractal and regression analysis. AEU-International Journal of Electronics and Communications, Vol. 71, 72-81. (2017)

38. Chernick, M. R., Murthy, V. K., &Nealy, C. D.: Application of bootstrap and other resampling techniques: Evaluation of classifier performance. Pattern Recognition Letters, Vol. 3, No. 3, 167-178. (1985)

39. Singh, S. K., Jeong, Y. S., & Park, J. H.: A Deep Learning-based IoT-oriented Infrastructure for Secure Smart City. Sustainable Cities and Society, Vol. 60, 102252. (2020)

40. Kuang, F., Zhang, S., Jin, Z., & Xu, W. A novel SVM by combining kernel principal component analysis and improved chaotic particle swarm optimization for intrusion detection. Soft Computing, Vol. 19, No. 5, 1187-1199. (2015)

41. Duan, L., Han, D., & Tian, Q.: Design of intrusion detection system based on improved ABC_elite and BP neural networks. Computer Science and Information Systems, Vol. 16, No. 3, 773-795. (2019)

42. Zhang, L. B., Peng, F., Qin, L., & Long, M.: Face spoofing detection based on color texture Markov feature and support vector machine recursive feature elimination. Journal of Visual Communication and Image Representation, Vol. 51, 56-69. (2018)

43. Tanwar, S., Tyagi, S., Kumar, S.: The role of internet of things and smart grid for the development of a smart city. In Intelligent Communication and Computational Technologies 23-33. (2018)

44. Qi, E., Deng, M.: R&D investment enhance the financial performance of company driven by big data computing and analysis. Computer Systems Science and Engineering, Vol. 34, No. 4, 237-248. (2019)

45. Kumari, A., Tanwar, S., Tyagi, S., Kumar, N., Parizi, R. M., & Choo, K. K. R.: Fog data analytics: A taxonomy and process model. Journal of Network and Computer Applications, Vol. 128, 90-104. (2019).

46. Wang, T., Wen, L., Zhou, Y., & Zhang, J.: A energy balanced routing scheme in wireless sensor networks based on non-uniform layered clustering. International Journal of Sensor Networks, Vol. 27, No. 4, 239-249. (2018)

47. Oh, B.D., Song, H.J., Kim, J.D., Park, C.Y., Kim, Y.S.: Predicting Concentration of PM10 Using Optimal Parameters of Deep Neural Network. Intelligent Automation and Soft Computing, Vol. 25, No. 2, 343-350. (2019)

48. Azzaoui, A. E., Singh, S. K., Pan, Y., & Park, J. H.: Block5GIntell: Blockchain for AI-Enabled 5G Networks. IEEE Access. (2020) https://doi.org/10.1109/ACCESS.2020.3014356

49. Yang, H., Yi, J., Zhao, J., Dong, Z.: Extreme learning machine based genetic algorithm and its application in power system economic dispatch. Neurocomputing, Vol. 102, 154-162. (2013)

50. Cha, J., Singh, S.K., Pan, Y., Park, J. H.: Blockchain-Based Cyber Threat Intelligence System Architecture for Sustainable Computing. Sustainability, Vol. 12, 6401. (2020)

51. Sicato, J. C. S., Singh, S. K., Rathore, S., Park, J. H.: A Comprehensive Analyses of Intrusion Detection System for IoT Environment. Journal of Information Processing Systems, Vol. 16, No. 4, 975-990, 2020. (2020)

52. Zhang, T., Hou, M., Zhou, T., Liu, Z., Cheng, W., & Cheng, Y.: Land-use classification via ensemble dropout information discriminative extreme learning machine based on deep convolution feature. Computer Science and Information Systems, Vol. 1, No. 00, 10-10. (2020)

Sushil Kumar Singh received his M.Tech. Degree in Computer Science and Engineering from Uttarakhand Technical University, Dehradun, India. Currently, he is pursuing his Ph.D. degree under the supervision of Prof. Jong Hyuk Park at the Ubiquitous Computing Security (UCS) Laboratory, Seoul National University of Science and Technology, Seoul, South Korea.His current research interests include Blockchain, Artificial Intelligence, Big Data, and the Internet of Things.

Jeonghun Cha received the B.S. degree in computer science from DongyangMirae University, Seoul, South Korea. He is currently pursuing the master's degree in computer science and engineering with the Ubiquitous Computing Security (UCS) Laboratory, Seoul National University of Science and Technology, Seoul, South Korea, under the supervision of Prof. Jong Hyuk Park. His current research interests include Information security, Cyber Threat Intelligence, and Internet-of-Things (IoT) security.

Tae Woo Kim received the B.S. degree in computer science from Kumoh National Institute of Technology, Gumi, South Korea. He is currently pursuing the master's degree in computer science and engineering with the Ubiquitous Computing Security

618   Sushil Kumar Singh et al.

(UCS) Laboratory, Seoul National University of Science and Technology, Seoul, South Korea, under the supervision of Prof. Jong Hyuk Park. His current research interests include Cloud security, Software Defined Network, and Internet-of-Things (IoT) security.

Dr. Jong Hyuk Park received Ph.D. degrees in Korea University, Korea and Waseda University, Japan. He is now a professor at the Department of Computer Science and Engineering, Seoul National University of Science and Technology, Korea. Dr. Park has published about 300 research papers in international journals and conferences. He is editor-in-chief of Human-centric Computing and Information Sciences (HCIS) by Springer, The Journal of Information Processing Systems (JIPS) by KIPS. His research interests include IoT, Information Security, Smart City, Blockchain, etc. Contact him at jhpark1@seoultech.ac.kr.