Computer Science and Information Systems 21(3):989–1012 https://doi.org/10.2298/CSIS231225025Y

News Recommendation Model Based on Encoder Graph Neural Network and Bat Optimization in Online Social Multimedia Art Education

Jing Yu[1], Lu Zhao[1,⋆], Shoulin Yin[2], and Mirjana Ivanovic[3]

[1] Lu Xun Academy of Fine Arts
Shenyang 110004, China
ohyou@sina.com
zzll_201@foxmail.com

[2] Software College, Shenyang Normal University
Shenyang 110034, China
yslin@hit.edu.cn

[3] Faculty of Sciences, University of Novi Sad
21000 Novi Sad, Serbia
mira@dmi.uns.ac.rs

Abstract. At present, the existing news recommendation system fails to fully consider the semantic information of news, meanwhile, the uneven popularity of news will also cause the phenomenon of long tail. Therefore, we propose a novel news recommendation model based on encoder graph neural network and Bat optimization in online social networks. Firstly, Bat optimization algorithm is used to improve the effect of news clustering. Secondly, the concept of metadata is introduced into the graph neural network, and the ontology of learning resources based on knowledge points is established to realize the correlation between news resources. Finally, the model combining Convolutional Neural Network (CNN) and attention network is used to learn the representation of news, and Gate Recurrent Unit (GRU) is used to learn the short-term preferences of users from their recent reading history. We carry out experiments on real news datasets, and compared with other advanced methods, the proposed model has better evaluation indexes.

Keywords: news recommendation system, encoder graph neural network, Bat optimization, online social networks, GRU.

1. Introduction

With the rapid development of network environment, information technology has been widely used, and people's news reading habits have gradually shifted from the traditional media such as newspaper and TV to the Internet. However, people's ability to read news information is often lower than the speed of news production, and the news information environment is becoming increasingly dense, eventually leading to "information overload". Therefore, how to efficiently, quickly and effectively recommend news text resources of interest to users has become increasingly important [1,2].

⋆ Corresponding author

In recent years, personalized news recommendation has gradually become the hot spot of Web technology, and the recommendation system has a large number of applications in real-time information, news, microblog, movie, music, blog, e-commerce and so on [3]. Through the recommendation system, problems such as information overload can be quickly, efficiently and accurately dealt with, users' interest rating results on news content and user behavior characteristics can be analyzed, and users' interest model can be automatically constructed by using historical data. There is no need to deliberately conduct interest surveys on users, which greatly and effectively reduces the burden on users and greatly improves the popularity of the recommendation system [4].

Personalized recommendation function is also very common in the mobile Internet industry, today's headlines, Meituan public point rating network have personalized recommendations [5]. Although people have carried out in-depth technical research and development of recommendation algorithms for many times, the existing recommendation algorithms still have limitations: (1) News has strong timeliness. News platforms automatically generate a lot of new news every day, and old news will quickly disappear. This causes a large degree of cold start problem, resulting in many user behavior dependent recommendation methods such as collaborative filtering can not be used; (2) News articles have a rich body, which contains important information that is not used, and cannot be directly represented by ID and other features simply and intuitively; (3) Accurately model user interest. Users' interest is diverse and changes over time, which requires a lot of mining and modeling based on user feedback behavior. But there is often no explicit user feedback on news platforms, and even implicit feedback is sparse. Therefore, personalized news recommendation has become a research hotspot of many scientific research institutions at home and abroad, and has received high attention from domestic and foreign academic conferences in the fields of information retrieval, data mining and artificial intelligence [6,7].

In order to alleviate information overload and meet users' reading needs, many news platforms, such as Google News [8], apply recommendation system technology to recommend news and push personalized article content to users, which can significantly improve users' article click rate and reading satisfaction, and improve user experience. The potential value brought by news recommendation is multifaceted. In addition to intuitively increasing the user click rate and reading probability, the recommendation system can continuously meet the needs of users' preferences, increase user stickiness, and cultivate user loyalty for media platforms. Compared with other recommendation systems [9], news recommendation is not only affected by problems such as cold start and sparse data, but also faces unique problems in this field, such as rapid iteration of news content, unstructured data generated by news text, and uneven popularity of news. Therefore, researchers need to apply the right medicine and propose effective solutions to the actual problems in the field of news recommendation [10].

To solve the above problems, this paper divides the user's interest into multiple dimensions for learning when dealing with the user's long-term reading history, combines the user's own attribute information, and uses the attention mechanism to distinguish the influence of each dimension, so as to extract the user's interest in different dimensions in the reading history over a long period of time as the user's long-term preference. In addition, this paper uses the Convolutional Neural Network (CNN) [11] and the attention network [12] to learn the news representation model, and uses Gate Recurrent Unit

(GRU) [13] to learn the short-term preferences of users from the news reading sequence of users in the recent period, and integrates the long-term and short-term preferences to form a complete user representation. Therefore, this paper proposes a news recommendation model based on encoder graph neural network and Bat optimization. To a certain extent, the imbalance between the head and the tail can be alleviated, and the unpopular news can be accurately recommended to the required users as far as possible. Our main contributions are as follows:

1. Firstly, Bat optimization algorithm is used to improve the effect of news clustering.
2. Secondly, the concept of metadata is introduced into the graph neural network, and the ontology of learning resources based on knowledge points is established to realize the correlation between news resources.
3. Finally, the model combining Convolutional Neural Network (CNN) and attention network is used to learn the representation of news, and Gate Recurrent Unit (GRU) is used to learn the short-term preferences of users from their recent reading history.

2. Related Works

By analyzing and processing the original news data and user behavior data, the personalized news recommendation system combines different news recommendation methods to model news and users, fully extracts the features of news content, and mining user preferences to generate news and user embedded representation. Its architecture is shown in Figure 1. When the user enters the personalized news recommendation system, the recommendation engine will select the news that meets the user's needs and preferences from the candidate news set according to the user's reading history, location, preference and other factors, and sort the candidate news according to the prediction model to generate a recommendation list and display it to the user. For example, if the user has read football news before, the recommendation system may recommend the latest World Cup news to the user. If the user is located in Beijing, then the recommendation system may recommend news about the Beijing area. In addition, the user interface will display news on different topics for each user, collect user feedback and update the recommendations accordingly, thus enabling personalized news recommendations.

Although PNR technology has made remarkable progress, it still needs to further improve the level of personalized recommendation, including more comprehensive mining of news semantics, more granular extraction of user preferences and the construction of more efficient personalized news recommendation models. With the vigorous development of mobile Internet technology, personalized news recommendation based on mobile terminal has become the mainstream trend. Personalized mobile news recommendation system can provide news information to users anytime and anywhere, with good interactivity, and bring more convenient and comfortable experience for users to obtain news information in real time. However, the small screen size of mobile devices, the unstable network quality and the variability of usage scenarios may affect the effect and efficiency of personalized news recommendation system, which is still an important issue to be solved in the future personalized mobile news recommendation research.

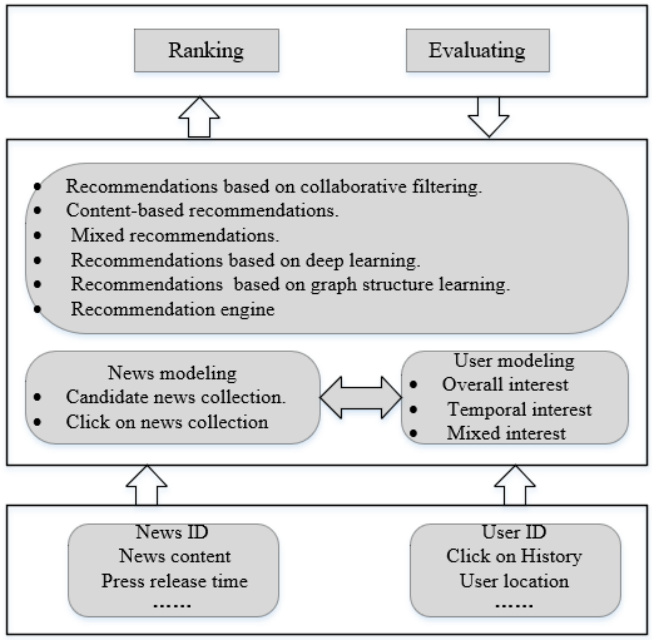

Fig. 1. Personalized news recommendation system architecture

2.1. Traditional News Recommendation Methods

Early news recommendation usually relied on the correlation between news and semantic correlation [14]. Both of these methods are to model news and calculate the similarity between news. However, neither of these two methods can accurately describe the user's interest degree, they only represent the similarity between news and lack the expression of user's interest degree. Therefore, how to mine and express users' interest in news reading in personalized news recommendation task is a very important issue.

Sequential recommendation attempts to predict a users next behavior by exploiting their historical behavior-sequences [15], which has been widely adopted in modern online information systems, such as news, video, advertisements, etc. Differ from traditional recommendation tasks that model user preferences in a static fashion, sequential recommendation is capable of capturing users evolved and dynamic preferences. For example, a user may prefer to watch soccer news only during the period of World Cup, which can be regarded as a kind of short-term preference.

Collaborative filtering (CF) [16] calculates the behavior similarity between the target user and the neighbor, predicts the target user's possible score on a certain item based on the nearest neighbor, and finally pushes the Top-N score to the target user. However, because the news recommendation platform produces a large number of new news every day, the news recommendation system has a serious cold start problem. However, the collaborative filtering method is difficult to effectively solve the cold start problem of the system, so it is not suitable for news recommendation. For this reason, many news recommendations are based on news content, such as Koning et al. [17] combine the semantics of TFIDF and domain ontology to make recommendations. Yang et al. [18] introduced a method of using word embedding to build user interest model and realize personalized news recommendation. Both of these methods only encode news from a single perspective and cannot fully and accurately express the semantic information of news.

2.2. News Recommendation Based on Deep Learning Methods

In 2017, Yahoo News used a deep learning network as a feature extractor to study news content, so as to find out users' preferences, which effectively improved the accuracy of the recommendation system compared with traditional methods [19]. However, this method only vectored news from the perspective of news words. Huang et al. [20] proposed a new deep hybrid recommendation model (DMFL), it vectored news from multiple perspectives, and finally summed and averaged the vectors from each perspective to obtain the feature vectors of users, which could not accurately model users.

In order to realize news recommendation, Google proposed Wide and Deep model [21], which could train both linear model and deep network, but had low learning efficiency in real scenes. Tiwari et al. [22] proposed personalized news recommendation for users based on their location and interest. This method mainly vectored users and could not accurately consider news semantic content. Liang et al. [23] proposed that knowledge perception neural network learned news representations from news article titles, and then learned user representations from news representations according to the similarities between candidate news and each browsed news. This method only quantified news from the perspective of news headlines, and could not accurately model news content.

Cold-start Recommendation. Making recommendations for new users or items with limited interaction samples is a challenging problem in recommender systems, also named cold-start problem. According to the number of interaction samples, the cold-start problem can be divided into two phases: cold-start phase (zero sample) and warm-up phase (a few samples).

For the cold-start phase, it is common to use auxiliary information, e.g., user attributes, item attributes, knowledge graph, samples in auxiliary domain.

The warm-up phase is a dynamic process that gradually improves the recommendation performance with the increase of samples. DropoutNet, MetaEmb and MeLU are able to solve the cold-start problem in warm-up phase, and they have been introduced above. MetaHIN and MAMO share similar idea with MeLU. These methods can be categorized into three groups: (1) MeLU, MAMO, MetaHIN try to personalize the parameters of the deep model. (2) MetaEmb exploits a good pre-trained embedding. (3) DropoutNet learns a more robust item embedding. Note that the proposed MWUF is completely different from these methods, and does not belong to the three groups. MWUF learns to warm up cold items by using meta networks to predict scaling and shifting functions which can transform the cold ID embeddings to fit the model better.

Multimodal-based news recommendation. Most of the existing news representation methods usually only learn the news representation from the news text, and ignore the visual information in the news (such as pictures, animations, etc.). In fact, users click on news not only because they are interested in the headlines, but also because they are likely to be attracted by multimodal features such as images, audio, and video. Therefore, integrating visual and text information to learn multimodal features is particularly important for news modeling and predicting news click rates. Guo et al. [24] proposed a news recommendation method based on deep reinforcement learning by integrating multimodal features to learn news representations and representing users' interests to multi-modal information. Wu et al. [25] adopted a pre-trained visual language model to encode news text and the region of interest images extracted from news images, and proposed a multimodal news recommendation method. Xun et al. [26] adopted the method of visual semantic modeling to capture the visual impression information perceived by users when browsing news, so as to further understand the process of users reading news. Experiments showed that news recommendation with multi-modal features could describe news content more comprehensively and improve the effect and accuracy of news recommendation.

Reference [27] proposed a method to predict news click rate through deep knowledge perception network and improved news representation learning by using knowledge graph. Reference [28] proposed a deep reinforcement learning framework, which could dynamically complete the modeling of news and users, and improve the diversity of news recommendation while ensuring the accuracy of recommendation. Reference [29] proposed a deep neural network model combined with attention mechanism to complete the task of news recommendation. The model included three extractors, which learned news representation, sequence information features and user interests respectively, effectively improving the accuracy of news matching. References [30,31] respectively proposed a neural network news recommendation model based on personalized attention mechanism and a neural network news recommendation method based on attention multiple views. The former focused on mining the impact of news on users at the word and document levels by applying attention mechanism, while the latter focused on mining different types of news information by integrating multi-view news recommendation methods for better news recommendation. It can be seen that most of the mainstream methods in the field of news recommendation are based on neural network and deep learning related technologies. In the future, neural network and deep learning will be important directions in this field.

3. Proposed News Recommendation Model

Firstly, Bat optimization algorithm is used to improve the effect of news clustering. Secondly, the concept of metadata is introduced into the graph neural network, and the ontology of learning resources based on knowledge points is established to realize the correlation between news resources. Finally, the model combining Convolutional Neural Network (CNN) and attention network is used to learn the representation of news, and Gate Recurrent Unit (GRU) is used to learn the short-term preferences of users from their recent reading history. The proposed scheme process is shown in figure 2.

3.1. Bat Optimization-based News Clustering

In the process of data modeling, project type characteristics are set in this paper, which can accurately feedback news preferences and provide accurate recommendation content. For personalized recommendation system, there are many types of characteristics in each project. This results in fewer project types than project numbers. In addition, different

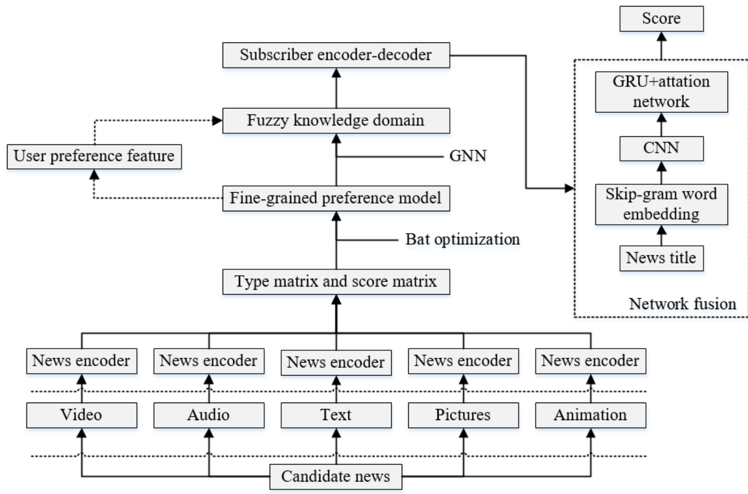

Fig. 2. Proposed news recommendation model framework

users will have different news interests and preferences for each project type. According to the above characteristics, this paper chooses the method combining item-scoring matrix and item-type matrix to design the ration-proportion-item-preference (RPIP) algorithm. Based on this, the user's preferences for various items are predicted, and the fine-grained user-project type preference matrix is constructed.

Rating proportion (RP) represents the proportion of the total rating R_u of a user u to the total rating of an item of type $R_{u,i}$. The calculation process of the score ratio is shown in equation (1).

$$Q_{u,i} = \frac{\sum_{i \in I_{u,e}} R_{u,i}}{R_u}. \tag{1}$$

Where $Q_{u,i}$ is the score ratio. i is a self-scoring variable.

Item preference (IP) represents the proportion of the item number $C(I)$ of type e in the total number $C(I_e)$, which can effectively avoid the difference in user preference caused by popular items of type e. The matrix $M_{i,e}$ is compiled by calculating the items in equation (2). To avoid the problem that the number of items of type e is 0, this article uses $C(I_e) + 1$ as the denominator.

$$M_{i,e} = ln \frac{C(I)}{C(I_e) + 1}. \tag{2}$$

User's preference for project type P_u is:

$$P_u = Q_{u,i} \times M_{i,e}. \tag{3}$$

Formula (3) is used to calculate the degree of user preference for project type, and a fine-grained user-project type preference matrix is constructed. In this paper, the real evaluation data are selected in the process of analysis and calculation. For the case that the evaluation data is empty, part of the data is represented as 0.

The processing process of bat-optimized user fuzzy clustering algorithm is as follows: Firstly, the optimal initial clustering center is determined by bat-optimized algorithm; Then, fuzzy C-means (FCM) clustering is implemented for users. When bat optimization algorithm [32] is used, each bat is represented as a cluster center matrix . Bat optimization clustering is shown in **Algorithm 1**.

Algorithm 1 Bat optimization clustering

1: **Input:** Fine-grained user-item type preference matrix, bat population size h, cluster number, maximum number of iterations T.
2: **Output:** Membership matrix of user cluster containing W and c user clusters.
3: The randomly generated initial bat population C, and the initialized bat individual C_i, velocity V_i, position x_i, pulse emissivity r_i, loudness A_i, and frequency f_i.
4: Calculating membership.
5: The individual fitness values of all bats in the population are calculated and sorted, and the bat individuals with the best fitness values are selected.
6: The individual position and speed parameters of bats are modified.
7: Generating a random number r_0 and traversing individual bats within the population. When the conditions are met $r_0 < r_i$ to generate x_{new} and the fitness F_{fit} is calculated.
8: The random number r_1 is generated and the individual bats in the population are traversed. If condition $r_1 < A_i$ is true, receive x_{new} of step 5 and modify loudness and pulse emissivity.
9: When the number of iterations is less than T or the convergence condition of cluster center is not satisfied, the iteration is carried out again in step 2. On the contrary, bat individual x_{best} under the condition of optimal fitness value is output, and it is regarded as the optimal initial cluster center and cluster division is implemented to generate user cluster membership matrix W and c user clusters.

3.2. News Resource Ontology Construction Based on Improved Graph Neural Network

At present, most online education platforms provide users with services that integrate high-quality online courses and other learning resources. However, in the face of massive learning resources, it is difficult for users to quickly find learning materials suitable for themselves, and even problems such as ”cognitive load” and ”information trek” may occur. Therefore, it is an urgent need of online education platform to intelligently analyze the rules of learning behavior and cognitive characteristics of users according to their learning process, and to provide personalized learning resource recommendation.

Information management. The management module of the NEWS ontology provides a vocabulary which allows the definition of metadata regarding the news item life-cycle management. Aspects like authorship information, rights, etc. are covered by this module. An application of such information is, for example, finding adequate journalists to cover a certain event (e.g. a soccer match, a political meeting, etc.). This is simply implemented by searching the NEWS system datawarehouse to find the authors of former news items reporting on similar events. Another example: management metadata could be used to define the urgency of a certain news item. This is helpful in deciding the order in which news items are sent to clients or are processed by certain procedures in the agency's production workflow (automatic annotation, for instance).

In order to realize effective recommendation of news resources, it is far from enough to consider users' preferences only. In the current educational environment, learning resources show an obvious trend of complexity and diversification both in structure and form, among which the forms of expression can be divided into courseware, cases, documents, indexes, online courses, test questions, test papers, assignments, texts, etc. In this context, the ontology construction based on knowledge points is realized by improving the graph neural network. The basic principle of this process is to fully reflect the data sharing value of news resources at all levels and types. The disordered and non-structured news resources are divided into text data and media data. Among them, the text materials include courseware, cases, test papers, workbooks of auxiliary materials, etc. Media materials include video, animation, audio, text, pictures, etc. For some resources whose classification is difficult to be defined directly, the improved graph neural network is used to restructure them, and the ontology representation construction method based on metadata is adopted to solve the fuzzy problem of knowledge domain definition in graph neural network.

To this end, the news resource model is first constructed, which is as follows:

$$\gamma = \frac{sim(x_1, x_2, \cdots, x_i)}{p}. \tag{4}$$

Where γ is the learning resource model. x_i is composed of knowledge points. $sim(x_1, x_2, \cdots, x_i)$ is the degree of similarity between the knowledge points. p indicates the proportion of metadata in the resource. This model is mainly used for the management and retrieval of knowledge points in learning resources. When the knowledge points are used to describe the resource ontology, the correlation between the ontology and knowledge points is established through metadata. Using metadata attributes to define the hierarchy of different categories, a graph neural network including parent classes (data, animation, course, etc.) and sub-classes (text, graphics, video, image, audio, etc.) is constructed. At this time, the hierarchical relationship of news resources can reflect the difference in the composition of knowledge points. In order to associate different resources, the relationship between knowledge points in the resource structure of improved graph neural network is designed into four types: parent-child, reference, dependence and parallel. Among them, the father-son relationship mainly reflects the part and integrity of resources, the dependence relationship mainly reflects the overlap of resources, and the parallel relationship mainly reflects the independence of knowledge points between resources.

3.3. Title Encoder

The title encoder vectorizes news headlines. The title encoder includes a preprocessing layer, a word vector training layer, a CNN layer and an attention layer.

The preprocessing layer performs word segmentation of the news titles, and the Jieba word segmentation tool is used to perform preprocessing operations such as word segmentation and deactivation of the titles. The word after the news title participle is expressed as $[s_1^t, s_2^t, \cdots, s_M^t]$, and M is the number of title words.

The CNN layer extracts the features between words by passing the word vectors of the training layer through the CNN network. The width of the convolution kernel is consistent with the word embedding dimension, and the number of window words for each convolution operation is denoted as h, that is, the convolution kernel $\varpi \in R^{h\times d}$. The convolution result of sliding the i-word window of the news title is:

$$N_i^t = ReLU(\varpi w_{i:i+h-1}) + b. \tag{5}$$

Where $ReLU$ is a nonlinear activation function. $w_{i:i+h-1}$ is the number of words taken for each convolution operation. $b \in R$ is offset. The title words are expressed as: $[N_1^t, N_2^t, \cdots, N_P^t]$.

The attention layer uses the attention mechanism to assign weight to each word. α_i^t represents the attention weight of the i-th word in the news title, and its calculation formula is:

$$a_i^t = \frac{q_t^T \tanh(V_t \times c_i^t + v_t)}{||q_t^T||_{L2}||q_t^T \tanh(V_t \times c_i^t + v_t)||_{L2}}. \tag{6}$$

$$\alpha_i^t = \frac{exp(a_i^t)}{\sum_{j=1}^{M} exp(a_i^t)}. \tag{7}$$

Where V_t represents the weight, v_t represents the offset item, c_i^t represents the vector of the $i-th$ word in the news title, and q_t represents the query vector. The news title vector is represented as:

$$r^t = \sum_{i=1}^{M} \alpha_i^t c_i^t. \tag{8}$$

3.4. Text Encoder

Text encoder vectorizes the news text. The text encoder includes a preprocessing layer, a word vector training layer, a CNN layer and an attention layer.

The preprocessing layer divides the news text into words, the Jieba word segmentation tool is used to divide the text into words, the words of the news text are represented as $[s_1^b, s_2^b, \cdots, s_P^b]$, P is the number of words in the text. The word vector training layer is to obtain the word vector of the news text from the word segmentation of the preprocessing layer through the BERT model, and the word vector of the text is represented as $[c_1^b, c_2^b, \cdots, c_P^b]$.

The CNN layer extracts the features between words by passing the word vectors of the training layer through the CNN network. The width of the convolution kernel is consistent with the word embedding dimension, and the number of window words for each convolution operation is denoted as h, that is, the convolution kernel $\varpi \in R^{h\times d}$. The convolution result of sliding the i-word window of the news text is:

$$N_i^t = ReLU(\varpi w_{i:i+h-1}) + b. \tag{9}$$

Where $ReLU$ is a nonlinear activation function. $w_{i:i+h-1}$ is the number of words taken for each convolution operation. $b \in R$ is offset. The title words are expressed as: $[N_1^t, N_2^t, \cdots, N_P^t]$.

The attention layer uses the attention mechanism to assign weight to each word. α_i^{tb} represents the attention weight of the $i-th$ word in the news title, and its calculation formula is:

$$a_i^b = \frac{q_b^T \tanh(V_b \times c_i^b + v_b)}{||q_b^T||_{L2}||q_b^T \tanh(V_b \times c_i^b + v_b)||_{L2}}. \tag{10}$$

$$\alpha_i^b = \frac{exp(a_i^b)}{\sum_{j=1}^{M} exp(a_i^b)}. \tag{11}$$

Where V_b represents the weight, v_b represents the offset item, c_i^b represents the vector of the $i-th$ word in the news title, and q_b represents the query vector. The news title vector is represented as:

$$r^b = \sum_{i=1}^{P} \alpha_i^b c_i^b. \tag{12}$$

3.5. Decoder

Since RNN has ability to learn current word from the previous words in the sequence and it can consider the encoder's representation and its internal state at the same time, we use RNN as the decoder. We denote x_d^i as the word embedding of the $i-th$ word in the scientific papers and apply gated recurrent unit (GRU) to solve gradient exploding or vanishing problem. The attention mechanism leans a weighted interpolation c_i based on the encoder's representation, i.e.,

$$c_i = \sum_j \alpha_{ij} s_j. \tag{13}$$

where α_{ij} is the output value derived from the softmax function and the $i-th$ word must be aligned with the $j-th$ output.

3.6. User Long-term Preference Fusion

The short-term interests of users are dynamic and changeable, and we usually learn users' short-term preferences from their recent news reading history. To capture sequential information in a user's reading history, we use a GRU network to learn a user's short-term interest representation. The news representation vector obtained by our news modeling is d, and the representation vector of the user's short-term reading history sequence can be expressed as $d_1, d_2, \cdots, d_N$. N is the length of the short-term sequence, then the GRU network calculates the expression of the user's short-term interest as shown in equations (14-18).

$$r_i = sigmoid(W_r[h_{t-1}, d_t]). \tag{14}$$

$$z_i = sigmoid(W_z[h_{t-1}, d_t]). \tag{15}$$

$$\tilde{h}_t = \tanh(W_{\tilde{h}}[r_1 \circ h_{t-1}, d_t]). \tag{16}$$

$$h_t = (1 - z_t) \circ h_{t-1} + z_t \circ \tilde{h}_t. \tag{17}$$

$$u_{short} = h_N. \tag{18}$$

Where $\circ$ indicates the hadamard product. W_r, W_z, and $W_{\tilde{h}}$ are parameters to be learned in the GRU network. The hidden state of the last output is recorded as the user's short-term interest indicating u_{short}.

3.7. Model Training

For the prediction of ratings between users and news, we use a simple dot product to efficiently calculate the matching score between users and news, i.e. $score = u^T d$.

We use negative sampling technique to train the model. For each positive sample, that is, the news clicked by the user, we randomly select K news that has not been clicked as negative samples, so as to transform the original complex prediction problem into $K+1$ binary classification problem, so as to simplify the training process and improve the computational efficiency. Under this condition, the posterior probability of each positive sample being clicked can be expressed as equation (19).

$$positive = \frac{exp(u^T d_i^P)}{exp(u^T d_i^P) + \sum_{j=1}^{K} exp(u^T d_{i,j}^N)}. \tag{19}$$

Where P represents the number of positive samples. d_i^P represents the $i-th$ positive sample. $d_{i,j}^N$ represents the $j-th$ negative sample of the $i-th$ positive sample, and the sum of the negative likelihood logarithm of the click-through rate of all positive samples is taken as the loss function of the model, namely,

$$Loss = \sum_{i=1}^{P} \log(\frac{exp(u^T d_i^P)}{exp(u^T d_i^P) + \sum_{j=1}^{K} exp(u^T d_{i,j}^N)}). \tag{20}$$

4. Experiment Results and Analysis

4.1. Datasets

This paper conducts experiments on the Chinese data set Baidu news, the English dataset MIND (Microsoft news data set) Adressa data set and Digg data set.

Baidu news is to crawl the browsing history of some users in September 2018 on Baidu News website, including users, news titles, news text, and browsing time stamps. We build the test set using data from September 25 to September 30, and the training set using data from September 1 to September 24.

MIND [33] is a large scale news recommendation English data set collected by Microsoft News from its anonymous user logs from October 12 to November 22, 2019, which contains users, news headlines, news text, news summaries, news categories, and news entities. This article builds a test set using data from November 15 to November 22, and a training set using data from October 12 to November 14.

Adressa data set [34] consists of newslogs collected from the Adresseavisen website over a period of 3 months, in both full and small datasets. The full version includes 3,083,438 users, 48,486 articles and 27,223,576 clicks in 10 weeks. The minor version has 561,733 users, 11,207 articles, and 2,286,835 clicks in one week.

The Digg dataset [35] consists of 3553 news items collected by the Information Science Institute of the University of Southern California on the Digg website in June 2009, including the digg_votes table and the digg_friends table. The digg_votes table contains 139,409 users and 3,018,197 votes; The digg_friends table contains 1731,658 link relationships between 71367 users.

4.2. Evaluation Index

In this paper, T is set as the test set to recommend news for user u. F_u is the recommended result. Suppose that the news set that user u is interested in in the test data set is denoted as T_u, and the performance indicators in this paper take Precision (P), Recall (R) and F1 index as references.

Precision refers to the proportion of content in the news list recommended by the news recommendation system that conforms to user preferences. The higher the accuracy rate, the more accurate the news recommendation is. The calculation formula is as follows:

$$P = \frac{\sum_u |F_u \cap T_u|}{\sum_u |F_u|}. \tag{21}$$

Recall refers to the proportion of news recommended by the recommendation system in the process of users interacting with many news, that is, the share of news recommended by the system in the news viewed by users, and its calculation formula is as follows.

$$R = \frac{\sum_u |F_u \cap T_u|}{\sum_u |T_u|}. \tag{22}$$

The measurement standard of the recommendation system is that the higher the accuracy rate and recall rate, the better. However, because these two indicators are contradictory to some extent, F1 is the evaluation index that combines the two. Therefore, this paper uses F1 values to measure the performance of the recommendation system. The higher the F1 value, the better the comprehensive performance. The calculation formula is as follows:

$$F1 = \frac{P + R}{PR}. \tag{23}$$

4.3. Influence of News Recommendation List N on P, R and F1 Values

The test set news count for the Chinese data set is 3654 click-throughs from 853 users, the test set for the English data set is 6325 click-throughs from 920056 users, test set of

Adressa is 4892 click-throughs from 10057 users and the test set of Digg is 7884 click-throughs from 50478 users.

The different length of the recommendation list will directly affect the P, R, and F1 values of the result. In order to verify the effectiveness of Attention-BTE under different length of recommendation list, in the test set, the range of recommendation list N is taken as 10,20,30,40,50 in this experiment. However, the value of N cannot be infinite, because it cannot be separated from the limit of the test set. Figures 3-6 (tables 1-4) respectively show the comparison of P, R and F1 values of Chinese, English, Adressa, Digg data sets under different recommendation list lengths N.

Table 1. Comparison of the number of different news recommendations in Chinese data set

Number of news recommendations	P	R	F1
10	0.38	0.40	0.20
15	0.40	0.40	0.21
20	0.42	0.40	0.22
25	0.51	0.56	0.28
30	0.60	0.72	0.34
35	0.56	0.60	0.29
40	0.46	0.48	0.24
45	0.43	0.44	0.22
50	0.40	0.39	0.19

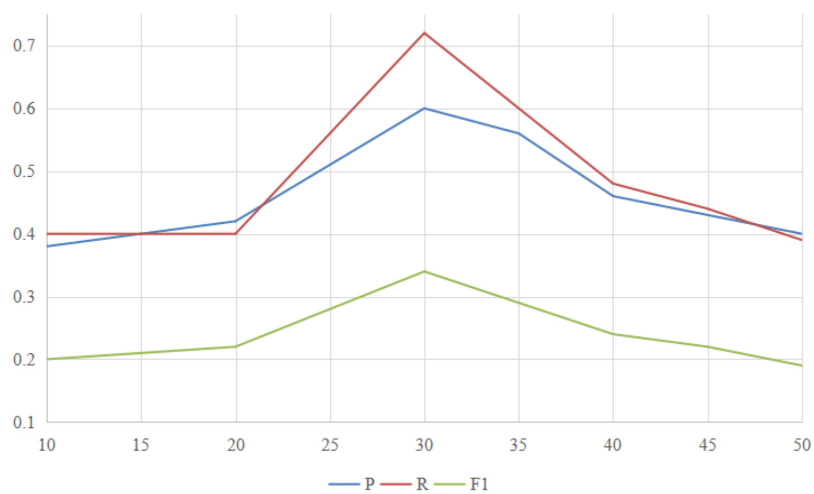

Fig. 3. Comparison of the number of different news recommendations in Chinese data sets

Table 2. Comparison of the number of different news recommendations in English data set

Number of news recommendations	P	R	F1
10	0.50	0.45	0.24
15	0.55	0.57	0.28
20	0.59	0.69	0.32
25	0.67	0.77	0.36
30	0.75	0.84	0.40
35	0.66	0.77	0.36
40	0.63	0.70	0.33
45	0.61	0.67	0.32
50	0.59	0.64	0.31

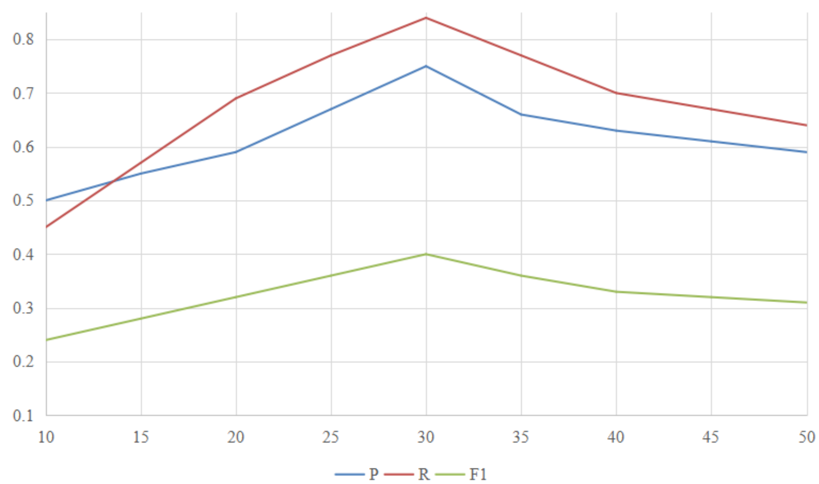

Fig. 4. Comparison of the number of different news recommendations in English data sets

Table 3. Comparison of the number of different news recommendations in Adressa data set

Number of news recommendations	P	R	F1
10	0.61	0.56	0.35
15	0.66	0.68	0.39
20	0.70	0.80	0.43
25	0.78	0.88	0.47
30	0.86	0.95	0.51
35	0.77	0.88	0.47
40	0.74	0.81	0.44
45	0.72	0.78	0.43
50	0.70	0.75	0.42

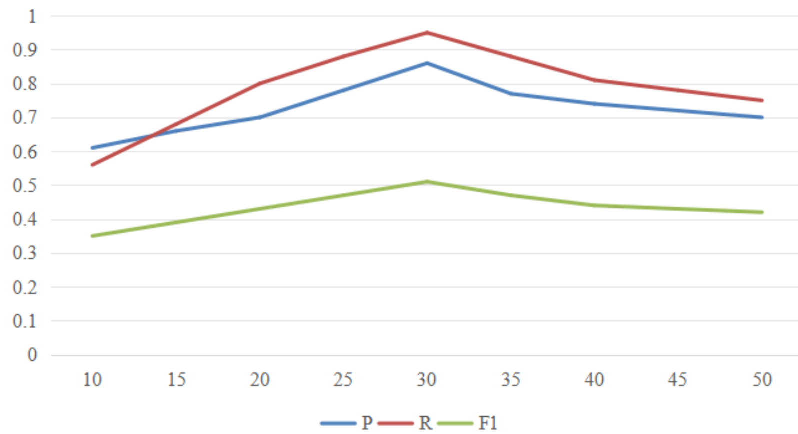

Fig. 5. Comparison of the number of different news recommendations in Adressa data sets

Table 4. Comparison of the number of different news recommendations in Digg data set

Number of news recommendations	P	R	F1
10	0.49	0.51	0.31
15	0.51	0.51	0.32
20	0.53	0.51	0.33
25	0.62	0.67	0.39
30	0.71	0.83	0.45
35	0.67	0.71	0.40
40	0.57	0.59	0.35
45	0.54	0.55	0.33
50	0.51	0.50	0.30

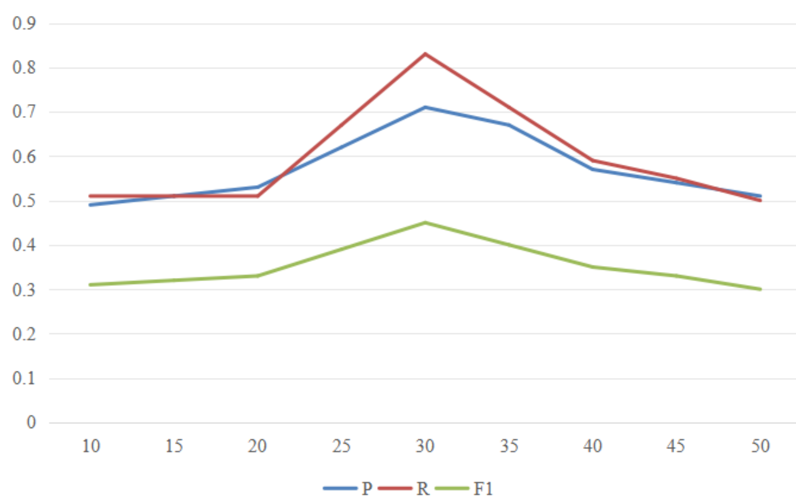

Fig. 6. Comparison of the number of different news recommendations in Digg data sets

It can be concluded from Figures 3-6 that when 30 news articles are recommended to users, P, R and F1 can achieve the maximum value, and the news recommendation effect is the best. As the list of recommendations grows, P, R, and F1 gradually increase. This is because the length of the recommendation list increases to include more items that interact with the user. However, the values of P, R and F1 show a decreasing trend when the number of recommendations is 40 and 50, which is due to the over-fitting phenomenon caused by the increase in the number of recommendations. That is, too small a number of recommendations cannot increase the number of projects that interact with users, and in the case of a small number of test sets, too large a number of recommendations is susceptible to the impact of quantitative over-fitting.

4.4. Comparison With Other State-of-the-art Methods

When the number of recommendation lists is the same, this new method is compared with news recommendation algorithms including KEHB [36], Feedrec [37], MACR [38], GACF [39]. The recommended precision, recall rate and F1 values in Chinese, English, Adressa, Digg data sets are shown in Tables 5-8 respectively. Meanwhile, pictorial diagrams of them are shown in figures 7-10. This article has the following observations.

Table 5. Comparison results of different models on Chinese data sets

Method	P	R	F1
KEHB	0.48	0.67	0.28
Feedrec	0.54	0.68	0.30
MACR	0.56	0.68	0.31
GACF	0.58	0.69	0.32
Proposed	0.61	0.72	0.33

Table 6. Comparison results of different models on English data sets

Method	P	R	F1
KEHB	0.62	0.75	0.29
Feedrec	0.62	0.76	0.29
MACR	0.69	0.80	0.32
GACF	0.72	0.82	0.33
Proposed	0.74	0.84	0.36

First, since the number of users and the number of news in English data set is larger than that in Chinese data set, the accuracy of Feedrec in Chinese data set is lower than that of KEHB, while the opposite is true in English data set. This may be due to the small number of users and news in the Chinese data set, and the serious problem of sparse user-item scoring matrix encountered in Feedrec, so the size of the data set is also the key to affecting the accuracy of news recommendation.

Table 7. Comparison results of different models on Adressa data sets

Method	P	R	F1
KEHB	0.51	0.64	0.24
Feedrec	0.51	0.65	0.23
MACR	0.58	0.79	0.31
GACF	0.69	0.81	0.32
Proposed	0.75	0.82	0.35

Table 8. Comparison results of different models on Digg data sets

Method	P	R	F1
KEHB	0.59	0.78	0.39
Feedrec	0.65	0.79	0.41
MACR	0.67	0.79	0.42
GACF	0.69	0.80	0.43
Proposed	0.72	0.83	0.44

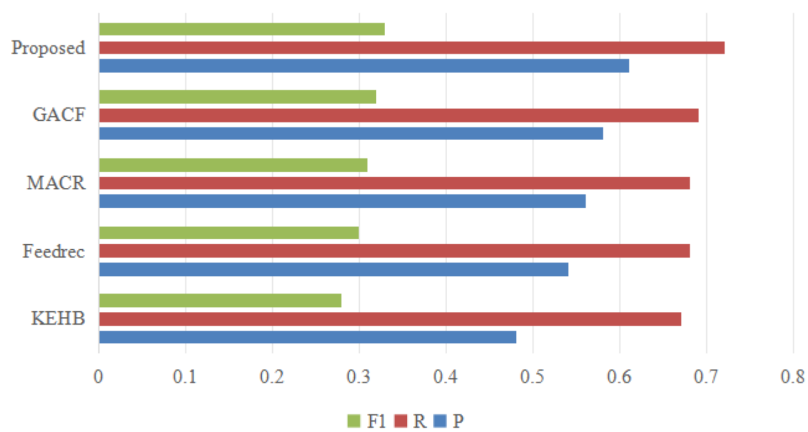

Fig. 7. Visualization display of table 5

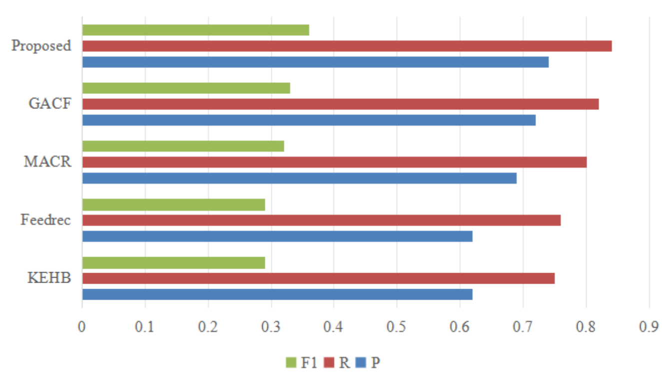

Fig. 8. Visualization display of table 6

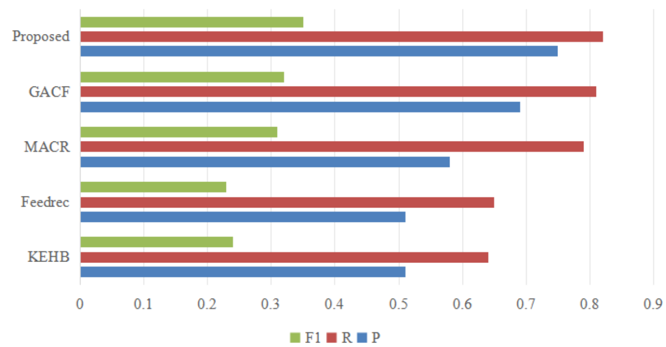

Fig. 9. Visualization display of table 7

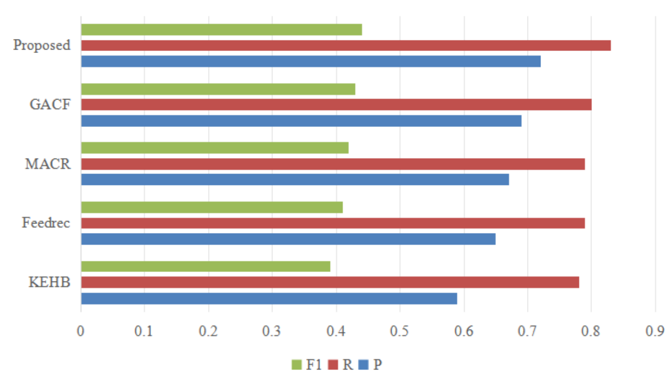

Fig. 10. Visualization display of table 8

Second, the method that combines headlines, body and categories (e.g. MACR) is superior to the method that only uses news headlines (e.g. GACF), which may be that rich content information helps to learn more accurate news presentation and proves the effectiveness of integrating multiple news information.

Third, the model in this paper is superior to KEHB, Feedrec, MACR and GACF, because the model makes full use of the theme, text and event of the news, and takes the event as a perspective of news information to form an informative news representation, and the F1 value is increased by about 6%.

4.5. Ablation Experiment

In this section, we conduct ablation experiments to show the effectiveness of proposed method. The results are shown in tables 9-12. From the results, we can see that the proposed method has the better effect.

Table 9. Ablation comparison results on Chinese data sets

Method	P	R	F1
CNN	0.49	0.66	0.29
CNN+GNN	0.57	0.68	0.31
CNN+GRU	0.58	0.69	0.32
Proposed	0.61	0.72	0.33

Table 10. Ablation comparison results on English data sets

Method	P	R	F1
CNN	0.66	0.78	0.30
CNN+GNN	0.68	0.79	0.31
CNN+GRU	0.71	0.82	0.34
Proposed	0.74	0.84	0.36

Table 11. Ablation comparison results on Adressa data sets

Method	P	R	F1
CNN	0.67	0.74	0.29
CNN+GNN	0.68	0.75	0.31
CNN+GRU	0.72	0.79	0.34
Proposed	0.75	0.82	0.35

Table 12. Ablation comparison results on Digg data sets

Method	P	R	F1
CNN	0.66	0.79	0.39
CNN+GNN	0.68	0.80	0.42
CNN+GRU	0.71	0.81	0.42
Proposed	0.72	0.83	0.44

5. Conclusions

Research on news recommendation has important practical significance and application needs. The accuracy of news recommendation lies in whether the content information of news is fully considered. This paper proposes a novel news recommendation model based on encoder graph neural network and Bat optimization in online social networks, news headlines, text and events are integrated through attention mechanism to get news vectorization. First of all, because words are the most basic unit of news headlines, text and events, and the same word has different meanings in different news, starting with the most basic words, each word is assigned different weights, which can fully consider news semantic information. Secondly, it can accurately vectorize users. Since each user is interested in news from different angles, the attention mechanism can accurately describe the user's perspective of interest in news, so as to accurately vectorize the user. Finally, the attention mechanism is used to assign weights to the headlines, text and events of the news, and the ratio of each Angle to the vectorization accuracy of the news is fully considered. The validity of this method is verified by experiments with real experimental data.

In exploring the diversity of news recommendation, the existing research usually only focuses on the diversity of recommended content, often ignoring the balance between diversity and accuracy, which is likely to lead to a large difference between the recommended news and the actual preferences of users. Thus, the user's satisfaction with the recommendation system is reduced, and even the user information is lost. Therefore, it is of great significance to design a more comprehensive diversified news recommendation model for improving the quality of online news service. In the follow-up research, the vectorization of events can be calculated using the advanced neural network method to further improve the accuracy of news recommendation.

Acknowledgments. The author is grateful for the anonymous review by the review experts.

References

1. Wu Y. The current situation and innovation of news communication under the environment of financial media[J]. Journal of Environmental and Public Health, 2022, 2022.
2. Li J. Research on readers' Reading Promotion in the digital age[C]//2022 International Conference on Artificial Intelligence in Everything (AIE). IEEE, 2022: 511-516.
3. Lian J. Research on Tourism Marketing Strategy Innovation in New Media Era[J]. Frontiers in Business, Economics and Management, 2023, 8(1): 103-106.
4. Pido N W T, Sujitno F D C. The effect of digital technology on students' reading Behavior[J]. Journal of English Teaching and Linguistic Issues (JETLI), 2022, 1(2): 81-93.

5. Jisi A, Yin S. A new feature fusion network for student behavior recognition in education[J]. Journal of Applied Science and Engineering, 2021, 24(2): 133-140.
6. Zhang P, Zheng J, Lin H, et al. Vehicle Trajectory Data Mining for Artificial Intelligence and Real-Time Traffic Information Extraction[J]. IEEE Transactions on Intelligent Transportation Systems, 2023.
7. Bhadrashetty A, Patil S. Movie Success and Rating Prediction Using Data Mining[J]. Journal of Scientific Research and Technology, 2024: 1-4.
8. Le H, Maragh R, Ekdale B, et al. Measuring political personalization of Google news search[C]//The World Wide Web Conference. 2019: 2957-2963.
9. Ko H, Lee S, Park Y, et al. A survey of recommendation systems: recommendation models, techniques, and application fields[J]. Electronics, 2022, 11(1): 141.
10. Das D, Sahoo L, Datta S. A survey on recommendation system[J]. International Journal of Computer Applications, 2017, 160(7).
11. Yin S, Wang L, Shafiq M, et al. G2Grad-CAMRL: An Object Detection and Interpretation Model Based on Gradient-weighted Class Activation Mapping and Reinforcement Learning in Remote Sensing Images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 16, pp. 3583-3598, 2023
12. Niu R, Sun X, Tian Y, et al. Hybrid multiple attention network for semantic segmentation in aerial images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-18.
13. Ren L, Cheng X, Wang X, et al. Multi-scale dense gate recurrent unit networks for bearing remaining useful life prediction[J]. Future generation computer systems, 2019, 94: 601-609.
14. Zhu L, Song J, Zhu X, et al. Adversarial learning-based semantic correlation representation for cross-modal retrieval[J]. IEEE MultiMedia, 2020, 27(4): 79-90.
15. Chang J, Gao C, Zheng Y, et al. Sequential recommendation with graph neural networks[C]//Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval. 2021: 378-387.
16. Fkih F. Similarity measures for Collaborative Filtering-based Recommender Systems: Review and experimental comparison[J]. Journal of King Saud University-Computer and Information Sciences, 2022, 34(9): 7645-7669.
17. de Koning E, Hogenboom F, Frasincar F. News recommendation with CF-IDF+[C]//Advanced Information Systems Engineering: 30th International Conference, CAiSE 2018, Tallinn, Estonia, June 11-15, 2018, Proceedings 30. Springer International Publishing, 2018: 170-184.
18. Yang Z, Cui L, Wang X, et al. MIAR: Interest-Activated News Recommendation by Fusing Multichannel Information[J]. IEEE Transactions on Computational Social Systems, 2022.
19. Liu L, Yan R J, Maruvanchery V, et al. Transfer learning on convolutional activation feature as applied to a building quality assessment robot[J]. International Journal of Advanced Robotic Systems, 2017, 14(3): 1729881417712620.
20. Huang Z, Yu C, Ni J, et al. An efficient hybrid recommendation model with deep neural networks[J]. IEEE Access, 2019, 7: 137900-137912.
21. Cheng H T, Koc L, Harmsen J, et al. Wide & deep learning for recommender systems[C]//Proceedings of the 1st workshop on deep learning for recommender systems. 2016: 7-10.
22. Tiwari S, Pangtey M S, Kumar S. Location aware personalized news recommender system based on Twitter popularity[C]//Computational Science and Its ApplicationsCICCSA 2018: 18th International Conference, Melbourne, VIC, Australia, July 2C5, 2018, Proceedings, Part IV 18. Springer International Publishing, 2018: 650-658.
23. Liang Y, Meng F, Zhang Y, et al. Infusing multi-source knowledge with heterogeneous graph neural network for emotional conversation generation[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(15): 13343-13352.
24. Guo N, Fu Z, Zhao Q. Multimodal News Recommendation Based on Deep Reinforcement Learning[C]//2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP). IEEE, 2022: 279-284.

25. Wu C, Wu F, Qi T, et al. Mm-rec: Visiolinguistic model empowered multimodal news recommendation[C]//Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2022: 2560-2564.
26. Xun J, Zhang S, Zhao Z, et al. Why do we click: visual impression-aware news recommendation[C]//Proceedings of the 29th ACM International Conference on Multimedia. 2021: 3881-3890.
27. Yang C, Xiao Y, Zhang Y, et al. Heterogeneous network representation learning: A unified framework with survey and benchmark[J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 34(10): 4854-4873.
28. Liu D, Shan L, Wang L, et al. P3oi-melsh: Privacy protection target point of interest recommendation algorithm based on multi-exploring locality sensitive hashing[J]. Frontiers in Neurorobotics, 2021, 15: 660304.
29. Zhang M, Wu S, Gao M, et al. Personalized graph neural networks with attention mechanism for session-aware recommendation[J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 34(8): 3946-3957.
30. Ahmed M, Chakraborty P, Choudhury T. Bangla document categorization using deep RNN model with attention mechanism[C]//Cyber Intelligence and Information Retrieval: Proceedings of CIIR 2021. Springer Singapore, 2022: 137-147.
31. Peng D, Yuan W, Liu C. HARSAM: A hybrid model for recommendation supported by self-attention mechanism[J]. IEEE Access, 2019, 7: 12620-12629.
32. Yang X S, Hossein Gandomi A. Bat algorithm: a novel approach for global engineering optimization[J]. Engineering computations, 2012, 29(5): 464-483.
33. Yunanda G, Nurjanah D, Meliana S. Recommendation system from microsoft news data using TF-IDF and cosine similarity methods[J]. Building of Informatics, Technology and Science (BITS), 2022, 4(1): 277-284.
34. Z. Ou, Z. Han, P. Liu, S. Teng and M. Song, "SIIR: Symmetrical Information Interaction Modeling for News Recommendation," in IEEE Transactions on Neural Networks and Learning Systems, doi: 10.1109/TNNLS.2023.3299790.
35. K. Abbas et al., "Co-Evolving Popularity Prediction in Temporal Bipartite Networks: A Heuristics Based Model," in IEEE Access, vol. 11, pp. 37546-37559, 2023, doi: 10.1109/ACCESS.2023.3262587.
36. Hui B, Zhang L, Zhou X, et al. Personalized recommendation system based on knowledge embedding and historical behavior[J]. Applied Intelligence, 2022: 1-13.
37. Wu C, Wu F, Qi T, et al. Feedrec: News feed recommendation with various user feedbacks[C]//Proceedings of the ACM Web Conference 2022. 2022: 2088-2097.
38. Wei T, Feng F, Chen J, et al. Model-agnostic counterfactual reasoning for eliminating popularity bias in recommender system[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021: 1791-1800.
39. Long L, Yin Y, Huang F. Graph-aware collaborative filtering for top-N recommendation[C]//2021 International Joint Conference on Neural Networks (IJCNN). IEEE, 2021: 1-8.

Jing Yu is with the Lu Xun Academy of Fine Arts, Shenyang 110004, China. She is an associate professor, has published many journal and conference papers related to her research. Research direction: Art Design, Multimedia art, interest recommendation, human-computer interaction.

Lu Zhao is with the Lu Xun Academy of Fine Arts, Shenyang 110004, China. She is an associate professor, has published many journal and conference papers related to her research. Research direction: Art Design, Multimedia art, interest recommendation, human-computer interaction.

Shoulin Yin received the B.S. and M.S. degrees in image processing from Software College, Shenyang Normal University, Shenyang, China, in 2013 and 2015, respectively. He is currently a PHD with School of Information and Communication Engineering, Harbin Engineering University. He has published more than 20 technical articles in scientific journals and conference proceedings. His research interests include image fusion, object detection, and image recognition.

Mirjana Ivanovic (Member, IEEE) has been a Full Professor with the Faculty of Sciences, University of Novi Sad, Serbia, since 2002. She has also been a member of the University Council for informatics for more than 10 years. She has authored or coauthored 13 textbooks, 13 edited proceedings, 3 monographs, and of more than 440 research articles on multi-agent systems, e-learning and web-based learning, applications of intelligent techniques (CBR, data and web mining), software engineering education, and most of which are published in international journals and proceedings of high-quality international conferences. She is/was a member of program committees of more than 200 international conferences and general chair and program committee chair of numerous international conferences. Also, she has been an invited speaker at several international conferences and a visiting lecturer in Australia, Thailand, and China. As a leader and researcher, she has participated in numerous international projects. She is currently an Editor-in-Chief of Computer Science and Information Systems Journal.

Received: January 15, 2024; Accepted: April 10, 2024.