Personalized Recommendation for Online News Based on UBCF and IBCF Algorithms

Wei Shi[1]*, Yitian Zhang[2]
School of Humanities and Law, Nanchang HangKong University, Nanchang 330063, China[1]
The Centre for Translation and Intercultural Studies, The University of Manchester, Manchester, M139PL, UK[2]

***Abstract*—With the popularization of the Internet and the widespread use of mobile devices, online news has become one of the main ways for people to obtain information and understand the world. However, the increasing number and variety of news often cause users to feel troubled when searching for content of interest. To solve this problem, the first step is to design a personalized recommendation model for online news. Based on this model, a new personalized recommendation model is designed by combining the item-based collaborative filtering (IBCF) and the user-based collaborative filtering (UBCF). The experimental results showed that the average scores of the volunteers for the performance indicators, coverage indicators, and satisfaction indicators of the model were 85 and 93, 86, respectively. This system has high accuracy, low resource consumption, and higher user satisfaction, providing a new algorithmic approach for the field of recommendation models. The contribution of research is not only improving the accuracy of recommendations, but also increasing the diversity of recommendations, effectively solving the problem of data sparsity and real-time news. By introducing a tag propagation network for clustering analysis of users and projects, the recommendation results are further optimized and user satisfaction is improved. In addition, the research also realizes efficient data processing and storage through real-time user data collection and distributed data processing technology, which significantly improves the performance and response speed of the system.**

Keywords—IBCF algorithm; UBCF; collaborative filtering; news recommendations; label promotion network

I. Introduction

With the rapid development of the Internet, online news has become an important way for people to obtain information. However, faced with massive news information, how to effectively filter out content that users are interested in and improve their reading experience has become an urgent problem to be solved. Personalized recommendation systems have emerged, which analyze the interests, preferences, and behavioral habits of users to recommend relevant news content, thereby improving user satisfaction and loyalty [1-2]. Collaborative filtering algorithm is an important method in personalized recommendation systems, mainly divided into user-based collaborative filtering (UBCF) and items-based collaborative filtering (IBCF). However, single UBCF or IBCF algorithms have certain problems in the recommendation process [3]. For example, the UBCF algorithm, in news platforms with a large user base and diverse preferences, may be less accurate due to the addition of new users, due to the lack of understanding of the interests of new users. On the other hand, IBCF may not be able to effectively recommend diverse content when users have a low interest in specific news topics, because it mainly relies on the similarity between news projects and ignores the personalized needs of users. In view of the above issues, the study first designs a network news recommendation index model based on the characteristics of network news recommendation. In response to the inherent shortcomings of UBCF and IBCF algorithms, the two are innovatively combined. The Label Promotion network is used to perform clustering analysis on users and projects. By utilizing known user behavior information and news attribute information, the data sparsity is solved. The research question focuses on how to combine item-based collaborative filtering (IBCF) and user-based collaborative filtering (UBCF) algorithms to improve the accuracy and diversity of recommendations. The purpose of the research is to integrate IBCF and UBCF algorithms and introduce the label promotion network for cluster analysis, optimize the recommendation results, improve user satisfaction, and provide an efficient algorithm method for the personalized recommendation field of online news. The paper mainly has four parts. The first part is the research status of recommendation models. The second part combines the IBCF algorithm and UBCF algorithm to design a new IBCF-UBCF network news personalized recommendation model. The third part conducts comparative experiments on the algorithm performance. The fourth part summarizes the research content.

The innovation of this method is mainly reflected in the following aspects: Firstly, by integrating effectiveness, coverage, and user satisfaction, a comprehensive and scientific online news personalized recommendation index model has been constructed, providing an effective evaluation system for the continuous improvement of recommendation systems. Secondly, this study adopts streaming distributed data collection technology for real-time user data collection, and combines the Apache Spark framework with the ZooKeeper cluster to achieve efficient data processing and storage, effectively solving the bottleneck problem of data processing and storage. In addition, this method innovatively combines the UBCF algorithm based on user social relationship strength and interest similarity with the IBCF algorithm based on attribute similarity, improving the accuracy and diversity of recommendations. Finally, by categorizing data streams and establishing indexes, the pertinence of data processing and retrieval efficiency have been further improved. These innovative points collectively enhance the interactivity and personalization of news recommendations, providing users with higher quality services.

The main contribution of this study is to propose a novel online news personalized recommendation model that combines item-based collaborative filtering (IBCF) and user-based collaborative filtering (UBCF). By integrating these two collaborative filtering algorithms, this study not only improves the accuracy of recommendations, but also increases the diversity of recommendations, effectively solving the problems of data sparsity and real-time news. In addition, by introducing label promotion networks for clustering analysis of users and items, the recommendation results were further optimized and user satisfaction was improved. This study not only provides new algorithmic methods for personalized recommendation systems, but also provides strong technical support for the development of the online news industry.

II. Related Works

With the continuous development of the economy and society, more researchers are paying attention to injecting more intelligent elements into personalized services. Mizgajski et al. proposed a emotional perceived recommendation system method to address the effectiveness of emotional factors in recommendation systems. The emotional response of user self-evaluation was used to recommend news. The results showed that incorporating pleasant emotions into collaborative filtering recommendations had the best performance. It had further research value in selecting emotional response recommendations [4]. Goyani et al. combined two methods to improve recommendation performance to address the limitations of collaborative filtering and content filtering in movie recommendation systems. To solve the user similarity calculation, experimental results showed that this method could improve the accuracy. In addition, this paper also reviewed the different technological applications of recommendation systems to promote research progress in this field [5]. To explore the user preferences for specific news, Symeonidis et al. combined the intra session and inter session item transfer probabilities of users, revealing both short-term and long-term intentions. Experimental evaluation showed that this method could better capture the similarity between items jointly selected by users within and between consecutive sessions. It was superior to state-of-the-art algorithms [6]. Tewari et al. proposed a recommendation method based on a semi-automatic encoder, which integrated user ratings and other additional information to address the information overload and user interest matching in recommendation systems. This method had improved accuracy, recall, and F-value evaluation indicators compared to other popular methods. The top 10 recommendation results were more accurate [7]. Wang et al. aimed to address the large data volume, cold start, and data sparsity in modern commercial website recommendation systems by transforming large data volume into a large user group. The k-means clustering was applied to partition user groups. Then it was combined with collaborative filtering and content-based recommendation algorithms. When the accuracy and recall were about 0.4 and 0.8, the F value was the highest [8].

Gao, H et al. investigated the implicit knowledge in Industrial Internet of Things (IIoT) using collaborative learning techniques to address the difficulty in selecting suitable APIs. A recommendation method for enhancing matrix factorization models was proposed. This method was effective and superior in both real datasets and industrial system scenarios [9]. Zhu et al. proposed a news recommendation method based on deep attention neural network (DAN) to address the existing recommendation methods being unable to handle the dynamic diversity of news and user interests. This method used parallel convolutional neural networks (CNN) and recurrent neural networks (RNN) to aggregate user interest features and capture hidden order features of user clicks. The results showed that this method had superiority and effectiveness, with an accuracy rate of 95.5% in comparative experiments [10]. Huang et al. proposed an spatiotemporal long short-term memory network based on attention mechanism (ATST-LSTM) method to address the lack of spatiotemporal contextual information in Point of Interest (POI) recommendation. The results showed that it outperformed other recommendation methods [11]. Li and others proposed a new method for jointly processing new users and long tail recommendations in recommendation systems. By learning auxiliary information such as user attributes and social relationships, the cold start for new users is solved. Experimental results showed that this method outperformed existing best methods in social recommendation on real datasets such as images, blogs, videos, and music [12]. Zhao et al. proposed a NeuNext framework to address the complex sequence patterns and rich context in sparse user check-in data. Joint learning utilized POI context prediction to assist in the next one. Experimental results showed that this method outperformed other recommendation methods [13], as shown in Table I.

Although the above methods have made significant progress in personalized recommendation systems, there are still some shortcomings. In the previous study, most personalized recommendation systems face the problems of sparse data, cold start and insufficient diversity of recommendations. For example, the emotive perception recommendation system proposed by Mizgajski et al. fails to fully consider the dynamic changes of users' interests, resulting in the inability to match the recommended content with users' real-time needs. Goyani et al.'s work, while combining collaborative filtering and content filtering, still faces the challenge of reduced accuracy in sparse data. Symeonidis et al.'s research focused on capturing short - and long-term user intentions, but failed to effectively address the long tail recommendation problem. This paper proposes a method to combine UBCF and IBCF to better capture the interest of new users and solve the cold start problem by introducing social relationships between users, while IBCF can improve the diversity of recommended content and avoid the uniformity of recommendation results by focusing on project attributes. At the same time, the research also adopts the flow distributed data collection technology to ensure the acquisition of real-time user data, and further enhance the response ability of the system in the dynamic environment. Through the above advantages, the research can effectively overcome the shortcomings in the previous work and provide an efficient personalized recommendation solution for online news. This method not only better meets the growing individual needs of users, but also provides new ideas for research and application in related fields.

TABLE I A REVIEW OF RELATED STUDIES

Researchers	Proposed Method	Experimental Results
Mizgajski et al. [4]	Emotion-aware recommendation system method	Best performance by integrating pleasant emotions into collaborative filtering recommendations
Goyani et al. [5]	Combining two methods to improve recommendation performance	Improved accuracy of recommendations
Symeonidis et al. [6]	Combining user's intra-session and inter-session item transition probabilities	Better capture of item similarities commonly chosen by users within and across consecutive sessions
Tewari et al. [7]	Recommendation method based on semi-automatic encoders	improved accuracy, recall, and F-measure evaluation metrics
Wang et al. [8]	Converting big data into a large user group, combining collaborative filtering and content-based recommendation algorithms	Highest F-measure when precision and recall are approximately 0.4 and 0.8 respectively
Gao et al. [9]	Enhancing matrix factorization models using collaborative learning techniques	Effective and superior performance in real datasets and industrial system scenarios
Zhu et al. [10]	A news recommendation method based on deep attention neural networks	Superior and effective achieving 95.5%accuracy in comparative experiments
Huang et al. [11]	An attention-based spatio-temporal long short-term memory network approach	Superior to other recommendation methods
Li et al. [12]	A new method for jointly handling new users and long-tail recommendations	Superior performance compared to the best available methods in social recommendations on real datasets (eg, images, blogs videos, and music]
Zhao et al. [13]	NeuNext framework	Superior to other recommendation methods

III. A PERSONALIZED RECOMMENDATION MODEL FOR ONLINE NEWS BASED ON COLLABORATIVE FILTERING

This chapter is mainly divided into two sections. The first section first designs a personalized recommendation index system for online news. Based on this, a network news recommendation model based on the IBCF is designed. The second section mainly combines the IBCF with the UBCF. A series of improvements are made to the two algorithms, and an IBCF-UBCF network news recommendation model is designed.

A. *Personalized Recommendation Index System and Algorithm Design for Online News*

The personalized recommendation index model for online news is an important tool for measuring the performance of recommendation systems, mainly composed of three aspects. Firstly, there are performance indicators, including accuracy, recall, F1 value, and AUC value. These indicators can effectively measure the accuracy and reliability of recommendation algorithms, ensuring that users receive high-quality news recommendations [14-15]. Secondly, there are coverage indicators that cover popular content, long tail content, diversity, and personalization. These indicators aim to evaluate the coverage ability of recommendation systems for various types of news, to meet the diverse interests and needs of users [16-17]. Finally, there are user satisfaction indicators, including timeliness, user retention rate, user rating, and user feedback. These indicators can reflect the satisfaction and acceptance of users towards recommendation results. It is an important basis for optimizing recommendation systems. In summary, the personalized recommendation index model for online news integrates three aspects: effectiveness, coverage, and user satisfaction, providing a comprehensive and scientific evaluation system for the continuous improvement of recommendation systems [18-19]. The personalized recommendation index model for online news is shown in Fig. 1.

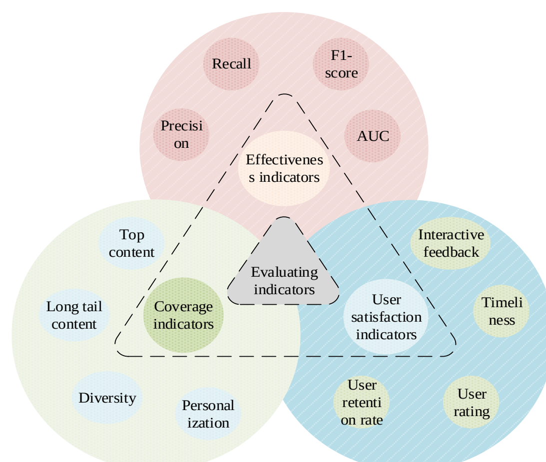

Fig. 1. A personalized recommendation index model for online news.

Based on the personalized recommendation index model for online news, the study first uses the IBCF algorithm to calculate the similarity between news, which is the core of the algorithm. The general IBCF algorithm uses chord similarity, as shown in Eq. (1).

$$sim(i,j) = \frac{i * j}{\| i \|_2 * \| j \|_2} \tag{1}$$

In Eq. (1), i represents a vector in the user space. j represents another vector in user space. However, cosine similarity assumes that the relationship between features is linear. For news data with non-linear relationships, it may not be the best choice. The adjusted cosine similarity takes this into account. Therefore, the mathematical expression for the modified cosine similarity is shown in Eq. (2).

$$sim(i,j)=\frac{\sum_{u\in U_{ij}}(r_{u,i}-\bar{r}_1)*(r_{u,j}-\bar{r}_j)}{\sqrt{\sum_{u_i}(r_{u,i}-\bar{r}_i)^2}*\sqrt{\sum_{u_j}(r_{u,j}-\bar{r}_j)^2}} \tag{2}$$

In Eq. (2), $r_{u,j}$ stands for the rating of u user on i. $\bar{r}_i$ represents the mean of the project i rating vector. $\bar{r}_j$ represents the mean of the project j rating vector. Fig. 2 displays the IBCF.

In Fig. 2, it is assumed that similar users A, B, and C exist, and items a, b, c, and d represent different news types. Each user's interest in a particular project forms an interaction matrix. User A's preference for item a, user B's preference for item b and item c, and user C's preference for item a form the core basis of the recommendation system. The reason why the study divided users A, B, and C into three categories lies in their similarities in interests and behaviors. User similarity is calculated by assessing their ratings or preferences for shared items. Specifically, user A likes item a, which belongs to a particular interest category, possibly news about a particular topic. User B is interested in both project b and Project c, reflecting A preference for similar topics, but may be somewhat different from user A's interest. User C: Liking item a may indicate significant differences with user B's interests and similarities with user A's interests. Through the above analysis, users' behavior in social networks or news apps can characterize their interests. In news recommendation, data sparsity is mainly reflected in the interaction matrix between users and news. This matrix is usually very large because it contains all user ratings or preference information for all news. However, in practical situations, users often only browse or evaluate a small portion of news, which results in most elements in the matrix being zero, forming a high-dimensional and sparse matrix. To address this issue, the study uses Singular Value Decomposition (SVD) for prediction filling, transforming the original high-dimensional sparse matrix into a low dimensional dense matrix. This can to some extent reduce the data sparsity. The algorithm flow is shown in Fig. 3.

Data preprocessing collects user-news interaction data, forms a rating matrix, and fills missing values (unobserved user news-interactions) with 0 or other appropriate default values. SVD is calculated. The SVD function in the linear algebra library is used to decompose the scoring matrix R, resulting in three matrices: U, $\sum$, and $V \wedge T$. SVD is truncated. According to the requirements, the first k largest singular values are selected to be retained, while the remaining smaller singular values are ignored. The U, $\sum$, and $V \wedge T$ matrices are updated to include the first singular value. The user hidden vector is calculated. For each user, its corresponding left singular vector is divided by the square root of the corresponding singular value to obtain the user hidden vector. The news hidden vector is calculated. For each news, its corresponding right singular vector is divided by the square root of the corresponding singular value to obtain the news hidden vector. User-news rating is predicted. For a given user and news, the predicted rating between them is calculated. Missing values are filled in. The missing values (i.e. elements with 0) in the original scoring matrix are replaced with predicted scores. For evaluation and optimization, the dataset is divided into training and testing sets. The parameter k and possible regularization parameters are adjusted to optimize the performance of the model. Assuming there is a user news interaction matrix R for $m \times n$ (m represents the number of users and n represents the number of news), SVD can decompose R into three matrices multiplied by each other, as shown in Eq. (3).

$$R=U\sum V\wedge T \tag{3}$$

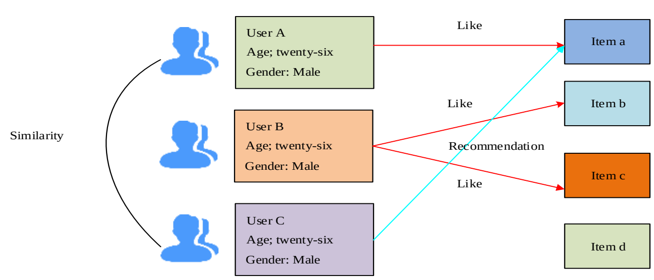

Fig. 2. Schematic diagram of IBCF algorithm.

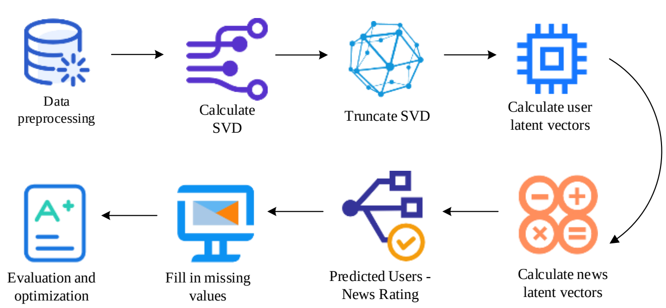

Fig. 3. The basic process of data filling based on clustering filling method.

In Eq. (3), U represents a unitary matrix of $m \times n$. Its column vector is the left singular vector of R. $\sum$ is a diagonal matrix of $k \times k$, with its non-zero elements (i.e. singular values) arranged in decreasing order. $V \wedge T$ is a unitary matrix of $k \times n$, and its column vector is the right singular vector of R. To reduce computational complexity and solve the data sparsity, the first k largest singular values are retained and the remaining smaller singular values are ignored. The purpose of SVD is to reduce computational complexity and solve data sparsity problems by preserving the first k largest singular values. Keeping these largest singular values ensures that we still capture most of the important information, while reducing unnecessary calculations by ignoring smaller singular values. Thus, while the dimensions of U and $V \wedge T$ are related to the number of users and news, respectively, the dimension of $\sum$ is determined by the number of retained singular values k. This method effectively realizes the reduction and approximation and solves the problem of data sparsity. The approximate expression of the original R matrix is displayed in Eq. (4).

$$R \approx U\_k \sum\_kV \wedge T\_k \tag{4}$$

In Eq. (4), U_k is the matrix composed of the first k columns of U. $\sum\_k$ is a matrix composed of the first k columns and first k rows of $\sum$. $V \wedge T\_k$ is a matrix composed of the first k columns of $V \wedge T$. This process is called Truncated SVD. Then, U_k and $V \wedge T\_k$ can be used for prediction filling. Unobserved user news interaction values can be estimated based on existing user behavior data and news content information, as shown in Eq. (5).

$$r\_\{ui\} \approx p\_u \wedge Tq\_i \tag{5}$$

In Eq. (5), p_u is the hidden vector of user u. q_i is the hidden vector of news i. The p_u is shown in Eq. (6).

$$p\_u = \frac{U\_k[:,u]}{sqrt(\sigma^2)} \tag{6}$$

In Eq. (6), σ is the k-th singular value. The q_i is shown in Eq. (7).

$$q\_i = \frac{V\_k[:,i]}{sqrt(\sigma^2)} \tag{7}$$

B. Design and Improvement of UBCF-IBCF

The IBCF algorithm has certain advantages in news recommendation, such as capturing the similarity between news and alleviating the cold start problem for new users and news. However, it also has some shortcomings. Although research has solved the data sparsity through SVD, news usually has timeliness and real-time, with a large number of news updates every day [20-21]. Therefore, it is necessary to frequently calculate the similarity matrix between news, which will increase the computational burden of the system. For uncommon news, due to the lack of sufficient interaction data, they may be ignored, resulting in recommendation results being too focused on popular news. Therefore, the study combines the UBCF algorithm to improve it. The overall structure diagram of the scheme is displayed in Fig. 4.

The model is mainly divided into two main branches, namely the UBCF branch and the IBCF branch. These two branches are independent of each other but work together, aiming to provide more comprehensive and accurate recommendations from both user and project perspectives. Firstly, there is the UBCF branch, whose main task is to utilize the relationship data between users to improve the accuracy. Specifically, the inputs of the UBCF branch include user-item rating data and user relationship data. After processing these data, the system starts calculating the strength of relationships between users. Then the system further calculates the fusion similarity of users. This step obtains a more comprehensive user similarity indicator by comprehensively considering the similarity of user ratings and the strength of social relationships. Based on this indicator, the system constructs a user similarity network that describes the similarity between users in terms of interests and social relationships. To further optimize the recommendation results, the system uses the Label Propagation network to perform clustering analysis on users. Through this method, the system can gather users with similar interests and social relationships together to form user clusters, as shown in Fig. 5.

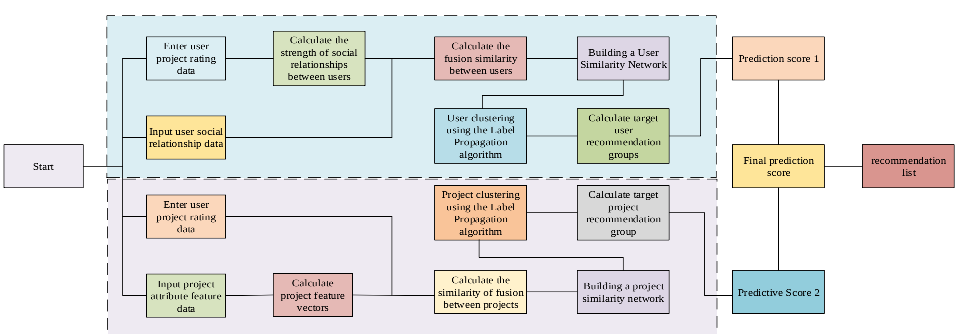

Fig. 4. Schematic diagram of news recommendation model structure combining IBCF algorithm and UBCF.

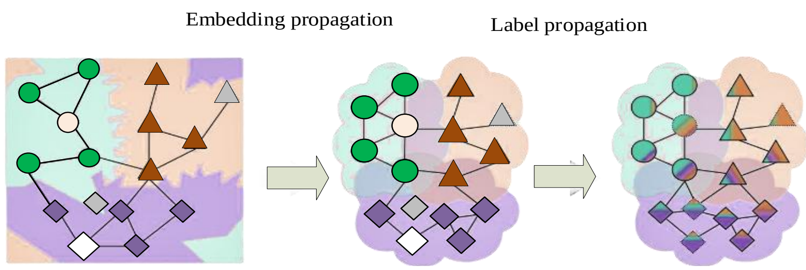

Fig. 5. Cluster analysis of users using the Label Propagation network.

The system identifies the user group that is most similar in interests and social relationships to the target user, namely the recommendation group. The UBCF branch calculates predicted scores based on this and rating data, reflecting the potential interest of the target user in the project. The IBCF branch utilizes project attribute feature data to improve recommendation accuracy. After processing user-project ratings and project attribute feature data, the system calculates project feature vectors and fuses similarity to obtain more comprehensive project similarity indicators. Due to the strong timeliness of news, traditional IBCF is unable to complete cold start recommendations when new projects appear. Therefore, the study adopts a new calculation method based on attribute similarity to solve this problem, as shown in Eq. (8).

$$Sim(I_1, I_2) = \beta Sim\_V(I_1, I_2) + (1-\beta) Sim\_P(I_1, I_2) \tag{8}$$

In Eq. (8), β represents the tuning parameter. Sim_P represents the similarity of items considering user rating preferences. Sim_V represents the similarity of attributes between two items, as shown in Eq. (9).

$$Sim\_V(I_1, I_2) = \frac{\sum_{t=1}^{n} Sim(p_t, q_t)}{n} \tag{9}$$

In Eq. (9), p_t stands for the t-th feature vector in project I_1. q_t stands for the t-th feature vector in project I_2. $Sim(p_t, q_t)$ represents the similarity of the t-th feature vector between project I_1 and project I_2. To simplify the calculation, it can be expressed as Eq. (10).

$$Sim(p_t, q_t) = \frac{(\max_t - \min_t) - |p_t - q_t|}{\max_t - \min_t} \tag{10}$$

In Eq. (10), $\max_t$ stands for the maximum value of the t-th eigenvector. $\min_t$ represents the minimum value of the t-th feature vector. In news recommendation models, this component may be simple text data. Therefore, the similarity of $Sim(p_t, q_t)$ needs to be determined based on different types, as shown in Eq. (11).

$$Sim(p_t, q_t) = \begin{cases} 1/T & p_t \neq q_t \\ 1 & p_t = q_t \end{cases} \tag{11}$$

In Eq. (11), T stands for the type of value under the t-th feature vector. Based on this indicator, a project similarity network is constructed, which describes the similarity between projects in terms of attributes and user evaluations. Similar to the UBCF branch, to optimize recommendation results, the IBCF branch uses a Label Propagation network to cluster items, forming a recommendation group. The predicted score is derived from the recommendation group and rating data, reflecting the user's interest in the project. UBCF and IBCF are fused to calculate the final prediction score, ensuring recommendation accuracy and diversity. The user recommendation groups is divided into three steps. The system generates a news recommendation list based on predicted scores, presenting the most relevant and interesting items to users, as shown in Fig. 6.

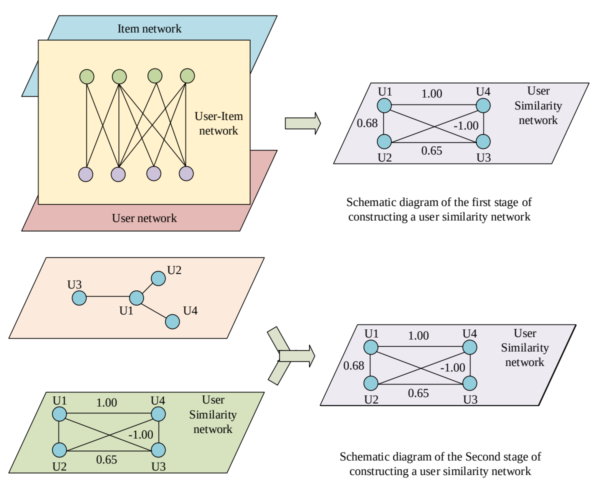

Fig. 6. Schematic diagram of the calculation process for user recommendation groups.

Firstly, based on the similarity of user rating preferences, the user-item bipartite graph (used to represent the preference relationship between users and items) is mapped to the original user similarity network. In the user-item bipartite graph, nodes are divided into two groups, one representing users and the other representing items. The second step is to use the Label Promotion network to aggregate user groups. Finally, similar users are grouped together to discover their group behavior patterns. The K-means is applied to calculate the user's belonging degree to the community, which requires calculating the category center and the center point of each cluster. Assuming c_j represents the center of cluster c_j, the cluster center is shown in Eq. (12).

$$\min j \in (1,2,...,k) \| x_i - c_j \| 2 \tag{12}$$

In Eq. (12), x_i represents sample i. The criterion function is shown in Eq. (13).

$$J = \sum i = 1n \min j \in (1,2,...,k) \| x_i - c_j \| 2 \tag{13}$$

In Eq. (13), n represents the number of samples. k represents clustering. The key to this method is to continuously adjust and optimize the center position of each category to ensure that they best represent the sample points in that category, thus forming clusters. As the number of users and news increases, data processing and storage can become a bottleneck. Given this, distributed database and cloud computing technology are used to disperse data to multiple nodes for storage and processing to improve data throughput and processing speed. When user and news numbers proliferate, similarity calculation and recommendation algorithms can become very time-consuming. Using parallel computing techniques, such as using the Apache Spark processing framework, to assign computing tasks to multiple processing units, thus accelerating algorithm execution and improving the scalability of the system. In order to improve the interactivity of personalized recommendations for online news, real-time user data collection is required. Research has adopted streaming distributed data collection technology for real-time user data collection. Based on the Apache Spark framework, configure hardware resources, install operating systems, and necessary software dependencies. The main reason for studying based on Apache Spark is its powerful distributed processing capabilities. Spark supports memory computing, reduces disk I/O operations, improves data processing speed, and is very suitable for real-time user data collection and analysis. Install and configure the ZooKeeper cluster for coordinating Spark jobs to ensure the normal operation of the server. Then assign sub server roles, divide the server into different roles based on system requirements and load balancing strategies, such as data extraction nodes, data processing nodes, etc., and study the use of HAProxy strategy. Create a Kinesis data collection queue again and establish a dedicated data collection queue to receive raw data obtained from the data source, ensuring that the data enters the system in an orderly manner. The next step is to classify the data flow, categorize the original data, and label the incoming data by user ID, news ID, or other attributes based on data type and source, in order to process specific data flows more efficiently in subsequent operations. Finally, establish an index for the stored data to improve the efficiency of subsequent queries and retrieval. Through this method, real-time user data collection and efficient processing can be achieved. By categorizing data streams and establishing indexes, the pertinence and retrieval efficiency of data processing can be improved, thereby enhancing the interactivity and personalization of news recommendations.

IV. Performance Testing and Applicability Analysis of IBCF-UBCF Algorithm

This chapter is mainly divided into two sections. The first section mainly conducts a series of algorithm comparison analysis and performance testing on the proposed IBCF-UBCF. The second section mainly focuses on the application of the IBCF-UBCF model in practical news recommendation experiments.

A. *Performance Testing of Personalized Recommendation Models for Online News*

To improve the quality of personalized recommendation content for online news, a new personalized recommendation algorithm for online news is designed by combining the IBCF algorithm and UBCF algorithm. To fully verify the excellent performance of the algorithm, the Collaborative Filtering (CF), Wisdom of the Crowd-based Recommendation (WCR), and Self-Attention Recommendation (SAR) are introduced to compare with the IBCF-UBCF algorithm, The above four algorithms are trained using the Microsoft News Dataset (MIND) and MSNBC.com dataset, respectively. The study took into account several characteristics of the sample, including historical behavioral data of users, the subject and type of news content, and users' social connections. Specifically, users' historical behavior describes their past clicks and comments on various types of news, reflecting their personal interests and preferences. The characteristics of news content, including title, keywords, publication time, etc., can help the algorithm understand the similarity between different news. In addition, the user's social relationship features are used to capture the interaction between users, and further enhance the personalization and accuracy of recommendations by analyzing the interaction between users in the social network. Fig. 7 the experimental results.

Fig. 7 shows the accuracy curves during the training process. From Fig. 7(a), in the MIND dataset, the IBCF-UBCF had the fastest convergence speed. After 140 iterations, it converged to 96.3%. The accuracy of WCR, SAR, and CF algorithms was slightly lower, at 90.3%, 85.4%, and 77.8%. From Fig. 7(b), the accuracy performance of the IBCF-UBCF didn’t changed much in the MSNBC.com dataset. The accuracy of the other three algorithms varied significantly. The improved UBCF algorithm proposed in the study had strong generalization ability and high accuracy. To further verify the superiority of the proposed network news personalized book intelligent recommendation model based on the IBCF-UBCF, the root mean square error (RMSE) and training time of the four algorithms are shown in Fig. 8.

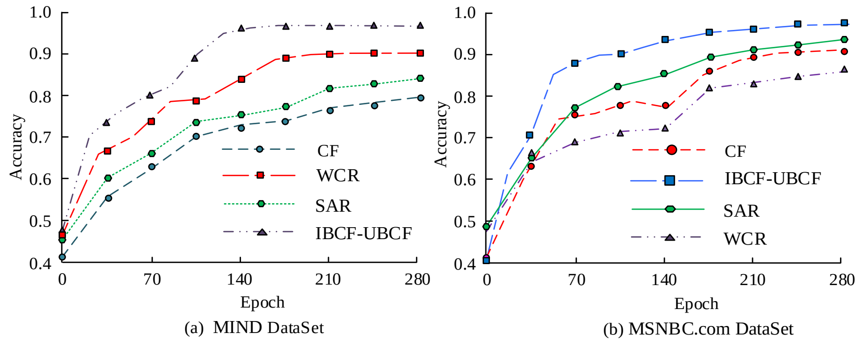

Fig. 7. Accuracy curves during the training process.

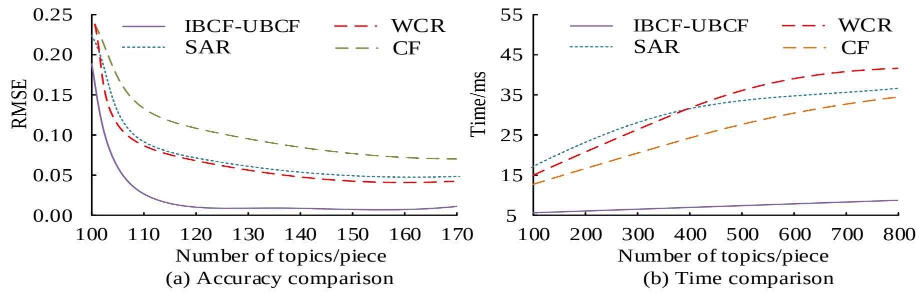

Fig. 8. Comparison results of root mean square error and running time.

Fig. 8(a) shows the RMSE comparison results of the four algorithms. The RMSE of the four algorithms decreased continuously with the increase of sample size. The minimum RMSE of the IBCF-UBCF was 0.010. The lowest RMSEs of WCR, SAR, and CF algorithms were 0.051, 0.053, and 0.068, respectively. Fig. 8(b) shows the running time. When the sample size was 800, the running time of the four algorithms was 6.3ms, 30.5ms, 35.4ms, and 38.9ms, respectively. In the above experimental results, the comparison results of the proposed algorithm on different data sets are different, which is due to the differences in the characteristics and structure of the data sets, such as the user behavior pattern, the diversity of news content and the degree of data sparsity. By combining UBCF and IBCF, the proposed algorithm can perform well in scenarios with strong user social influence, so it is more suitable for data sets with rich social network information. In addition, the performance of the algorithm can be fully utilized when the news content attributes in the data set are more diverse and the user's historical behavior records are sufficient. Conversely, in scenarios where data is sparse or social interaction is lacking, the effectiveness of the algorithm may be limited. To verify the effectiveness of each component in the IBCF-UBCF, ablation experiments are designed. Table II displays the experimental results.

TABLE II IBCF-UBCF ALGORITHM ABLATION EXPERIMENT

Model	**MSE**	**RMSE**	**MAE**
IBCF-UBCF	0.000032	0.0057	0.0022
WCR	0.000145	0.0131	0.0069
SAR	0.000078	0.0091	0.0044
CF	0.000052	0.0056	0.0010
Missing IBCF module	0.000168	0.0173	0.0061
Missing UBCF module	0.000173	0.0121	0.0056

From the results presented in Table II, among numerous network news recommendation algorithms, the IBCF-UBCF had excellent performance. The values of mean square error (MSE), RMSE, and mean absolute error (MAE) were 0.000032, 0.0057, and 0.0022. These data are significantly superior to other algorithms, fully demonstrating the superiority of the IBCF-UBCF in network news recommendation tasks. Furthermore, to verify the effectiveness of each module in the IBCF-UBCF, ablation experiments are conducted. When the IBCF module was missing, the MSE, RMSE, and MAE were 0.000168, 0.0173, and 0.0061, indicating a decrease in recommendation performance. This indicates that the IBCF module plays a crucial role in the algorithm. Similarly, when the UBCF module was missing, the MSE, RMSE, and MAE values of the algorithm were 0.000173, 0.0121, and 0.0056, respectively, showing a certain performance decline. These two experiments further confirm the importance of the IBCF and UBCF modules in the IBCF-UBCF, both of which are indispensable and together ensure the high performance of the algorithm.

B. Application Analysis of IBCF-UBCF in Network News Recommendation Model

The experiment repeatedly verifies the superiority of the IBCF-UBCF in the field of online news recommendation. To further prove its equally good performance in practical applications, it is applied in actual news recommendation software. The research method is compared with the methods proposed in references [22] and [23] to explore the resource consumption of each algorithm during operation. The CPU used in the experiment is Intel i5-2500K, and the experimental platform is Windows 10. The dataset used is the news recommendation data obtained from the ByteDance's official website. These data have been carefully cleaned, deweighted, marked and annotated to ensure the quality and accuracy of the data. In the process of processing, we also paid special attention to the correlation of user behavior and news content, thus constructing a highly personalized news recommendation data set, namely, Personalized News Recommendation Datasets (PNRD). The experimental results are shown in Fig. 9.

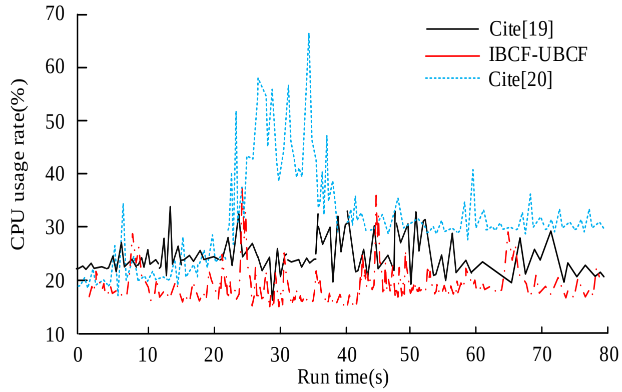

Fig. 9. The resource consumption of three algorithms in actual news recommendations.

From the graph, the proposed IBCF-UBCF had a relatively low CPU usage during system operation, with little fluctuation, maintaining a CPU usage rate of around 20. The performance of the algorithm proposed in research [23] was poor. Its CPU usage fluctuated greatly during the 10-30s period, and the average CPU usage during operation reached 30%. The performance of the method proposed in study [22] fell between the two, with an average CPU usage rate of 25% during runtime. In addition, the study also records the ROC curves of three models. Fig. 10 displays the results.

From the graph, the proposed IBCF-UBCF had the highest ROC curve, which fully demonstrated its outstanding performance among all the algorithms involved in the comparison. In addition, the area under the ROC curve of each algorithm, i.e. the AUC value, is experimentally calculated. The AUC value of the IBCF-UBCF was as high as 0.936, highlighting its superior news recommendation performance. In contrast, the AUC value of the reference [22] was 0.901, while the AUC of the reference [23] was 0.878, which once again proved the excellent performance of the IBCF-UBCF in news recommendation tasks. In addition, the experiment randomly interviews 1000 users in the age group (children, youth, middle-aged, quinquagenarian, old age), who evaluates the recommended content of the IBCF-UBCF algorithm using a percentage system, and compared with the proposed method in literature [23], the statistical results of the evaluation are shown in Fig. 11.

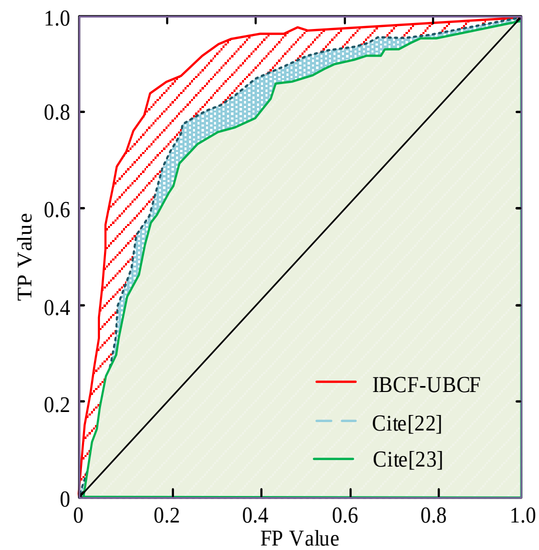

Fig. 10. ROC curves during the use of three algorithms.

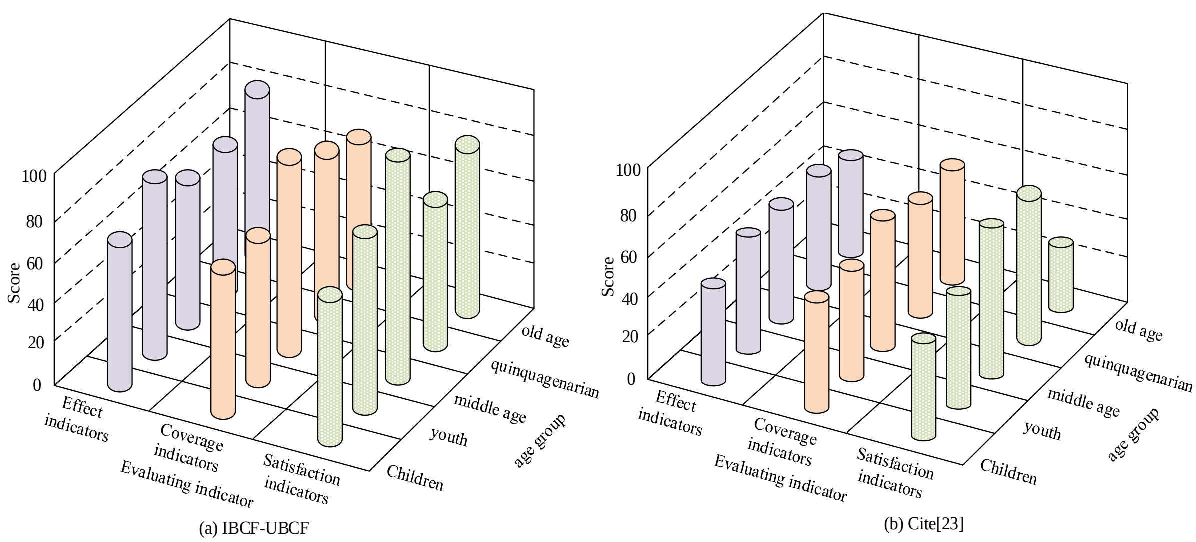

Fig. 11. The score of individual indicators in the IBCF-UBCF.

From the provided chart data, it is evident that volunteers from different age groups have relatively high evaluations of the three key indicators of the news recommendation system - performance, coverage, and satisfaction. Especially in terms of performance indicators, the news recommendation system achieved a high score of 95.3, demonstrating its outstanding performance in news push. Not only do young users give it a high rating, but other age groups also give it a rating of 85 or above, indicating that the system can meet the needs of users of all age groups. In terms of coverage, the middle-aged population gave the highest rating, with an average score of 94.3 points. Volunteers from other age groups also received positive reviews, with an average score of over 93, indicating that the news recommendation system has done a fairly comprehensive job in content coverage. As for satisfaction, even for the lowest rated 50-year-old population, their score is above 80 points, indicating that users are generally satisfied with the system. However, the performance of the methods proposed in reference [23] is relatively low. In summary, the news recommendation system has received high recognition among users of different age groups. In addition, the study tested the performance of the system in low, medium and high load conditions, recording response time, throughput and resource utilization, School of Humanities and Law, Nanchang HangKong University, Nanchang 330063, China. Subsequently, the three load conditions were tested again by increasing the system resources to observe the impact of the increased resources on the system performance. The experimental results are shown in Table III.

TABLE III COMPREHENSIVE PERFORMANCE ANALYSIS OF ONLINE NEWS PERSONALIZED RECOMMENDATION SYSTEM

Experimental group	**Response time (ms)**	**Throughput (requests/second)**	**Resource utilization rate (%)**
Low load	50	1000	20
Medium load	80	800	30
High load	120	600	40
Increase resources - low load	40	1200	25
Increase Resources - Medium Load	65	1000	35
Increase resources - high load	90	800	45

Under low, medium, and high load conditions in the experimental group, as the load increases, the response time gradually increases while the throughput gradually decreases. This indicates that under high loads, the system's processing speed of requests slows down, and the number of requests it can process also decreases. The resource utilization rate increases with the increase of load, indicating that the system fully utilizes resources under high loads. When resources are increased, response time decreases, throughput improves, and resource utilization improves under low, medium, and high load conditions. This indicates that increasing resources can effectively improve system performance, enabling the system to respond to requests faster, process more requests, and utilize resources more efficiently. In summary, the table data shows the impact of load and resources on system performance, as well as the performance improvement that increasing resources can bring. This provides valuable reference information for system optimization and resource allocation. Validation measures were included in the study, and the results were compared with previous studies, as shown in Table IV.

TABLE IV RESULTS OF ALGORITHM PERFORMANCE VERIFICATION

Reference	Method	Accuracy rate (%)	Recall rate (%)	RMSE
Mizgajski et al. [4]	Emotion perception recommendation system	82.5	78.0	0.052
Goyani et al. [5]	Combine collaborative filtering with content filtering	80.3	75.5	0.063
Symeonidis et al. [6]	User intent analysis	83.1	80.2	0.049
Research method	IBCF-UBCF algorithm	96.3	91.5	0.010

Table IV shows the performance verification results of the proposed IBCF-UBCF algorithm compared with previous studies. As can be seen from the table, IBCF-UBCF algorithm is significantly superior to other methods in the accuracy rate (96.3%) and recall rate (91.5%), indicating that it achieves higher user satisfaction and relevant content recommendation ability in the recommendation system. At the same time, the root-mean-square error (RMSE) of the algorithm is only 0.010, indicating that its error in predicting user preferences is very small, which further proves its excellent performance. In contrast, the accuracy and recall rates of previous research methods were both below 85%, and RMSE values were generally higher than 0.05, reflecting their instability and potential limitations in handling recommendation tasks. These results fully demonstrate the effectiveness and superiority of the IBCF-UBCF algorithm in the field of personalized recommendation, and verify its potential in improving user experience and recommendation system performance.

V. DISCUSSION

In order to evaluate the novelty of the proposed algorithm, the proposed IBCF-UBCF model was compared with other recommended algorithms (such as CF, WCR and SAR). By comparing the convergence rate, accuracy and RMSE, the performance of the proposed model was evaluated. The results show that the algorithm improved the accuracy and diversity of the recommendation. By evaluating the ability of the recommendation system to cover popular content, long tail content, diversity and personalization, we can determine whether the system can meet the interests and needs of different users. By considering user satisfaction indicators such as timeliness, user retention, user ratings, and user feedback, researchers are able to measure user satisfaction and acceptance of recommended results. Retest the three load conditions by increasing the system resources, and observe the effect of the increased resources on the system performance. The results show that the system realizes efficient data processing and storage, and effectively solves the bottleneck problem of data processing and storage. A comprehensive evaluation index system realizes the efficient data processing and storage, and improves the interactivity and personalization degree of the recommendation system.

The advantages of the study are that it effectively solve the data sparsity problem and improve the personalization of recommendations. In addition, the method realizes efficient data processing and storage through the flow of distributed data acquisition technology and cloud computing framework, which significantly improves the performance and response speed of the system. The disadvantage of research is that because the system mainly relies on regular calculations and updates, it may not reflect the latest interests of users in real time and the changes of news. Moreover, the average satisfaction of the model among middle-aged and elderly groups is relatively low, indicating that the system may need to be further optimized to better meet the needs of different age groups.

To overcome these defects, real-time data flow processing technology can be introduced in the future to realize just-in-time analysis of user behavior data, which can improve the real-time performance of the recommendation system and better capture the latest interests of users, such as the Apache Kafka technology adopted in literature [24]. Develop multimodal learning algorithms, combining user historical behavior and immediate feedback, and multi-dimensional features of news content to enhance the adaptability of the model to the needs of users of different ages. The user feedback mechanism is designed to allow users to evaluate the recommendation results, and then these feedback data is used to further train and optimize the algorithm to improve the relevance of recommendations and users' satisfaction, such as the scheme based on social relations and behavioral characteristics adopted in literature [25].

VI. Conclusion

To provide readers with more accurate network news recommendation results, a new network news recommendation algorithm is designed by combining the IBCF and UBCF. In algorithm performance testing, the IBCF-UBCF had the fastest convergence speed, reaching convergence after 140 iterations, and finally converging to 96.3%. The accuracy of WCR, SAR, and CF algorithms was slightly lower, at 90.3%, 85.4%, and 77.8%, respectively. The improved UBCF proposed in the study had strong generalization ability and high accuracy. In addition, the RMSE of the four algorithms decreased continuously with the increase of sample size. The minimum RMSE of the IBCF-UBCF was 0.010. The lowest RMSE of WCR, SAR, and CF were 0.051, 0.053, and 0.068, respectively. The IBCF-UBCF proposed in the study had a relatively low CPU usage during system operation, with little fluctuation. It generally maintained a CPU usage rate of around 20. The performance of the algorithm proposed by the comparative study is poor. In the 10-30 second period, its CPU usage fluctuates greatly, and the average CPU usage reaches 30% during operation. The results of the satisfaction assessment of the volunteers showed that they scored an average of 85, 93 and 86 points on the model's performance indicators, coverage indicators and satisfaction indicators, respectively. These high scores indicate the effectiveness of the model in meeting the personalized needs of users and enhancing the experience of the recommendation system. The experimental results show that the proposed method is significantly superior to the existing recommendation algorithms in many performance indexes, demonstrating excellent performance and low resource consumption. In addition, through the introduction of real-time user data collection and distributed processing technology, the response speed and scalability of the recommendation system are improved, and the effectiveness of the research method in practical applications is verified. The proposed algorithm effectively overcomes the common problems of sparse data, cold start and insufficient diversity of recommendation in traditional personalized recommendation methods, and significantly improves the accuracy and user satisfaction of recommendation by comprehensively considering users' historical behaviors, social relations and news content characteristics.

References

[1] Mokayed, H., Quan, T. Z., Alkhaled, L., & Sivakumar, V. Real-time human detection and counting system using deep learning computer vision techniques.//Artificial Intelligence and Applications. 2023, 1(4): 221-229.

[2] Pal, S., Roy, A., Shivakumara, P., & Pal, U. Adapting a Swin Transformer for License Plate Number and Text Detection in Drone Images.//Artificial Intelligence and Applications. 2023, 1(3): 145-154.

[3] Groumpos P P. A Critical Historic Overview of Artificial Intelligence: Issues, Challenges, Opportunities, and Threats.//Artificial Intelligence and Applications. 2023, 1(4): 197-213.

[4] Mizgajski J, Morzy M. Affective recommender systems in online news industry: how emotions influence reading choices. User Modeling and User-Adapted Interaction, 2019, 29(2): 345-379.

[5] Goyani M, Chaurasiya N. A review of movie recommendation system: Limitations, Survey and Challenges. ELCVIA: electronic letters on computer vision and image analysis, 2020, 19(3): 0018-37.

[6] Symeonidis, P., Chaltsev, D., Berbague, C., & Zanker, M. Sequence-aware news recommendations by combining intra-with inter-session user information. Information Retrieval Journal, 2022, 25(4): 461-480.

[7] Tewari, A. S., Parhi, I., Al-Turjman, F., Abhishek, K., Ghalib, M. R., & Shankar, A. User-centric hybrid semi-autoencoder recommendation system. Multimedia Tools and Applications, 2022, 81(16): 23091-23104.

[8] Wang W. Application of E-Commerce Recommendation Algorithm in Consumer Preference Prediction. Journal of Cases on Information Technology (JCIT), 2022, 24(5): 1-28.

[9] Gao, H., Qin, X., Barroso, R. J. D., Hussain, W., Xu, Y., & Yin, Y. Collaborative learning-based industrial IoT API recommendation for software-defined devices: the implicit knowledge discovery perspective. IEEE Transactions on Emerging Topics in Computational Intelligence, 2020, 6(1): 66-76.

[10] Zhu, Q., Zhou, X., Song, Z., Tan, J., & Guo, L. Dan: Deep attention neural network for news recommendation.//Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33(01): 5973-5980.

[11] Huang, L., Ma, Y., Wang, S., & Liu, Y. An attention-based spatiotemporal lstm network for next poi recommendation. IEEE Transactions on Services Computing, 2019, 14(6): 1585-1597.

[12] Li, J., Lu, K., Huang, Z., & Shen, H. T. On both cold-start and long-tail recommendation with social data. IEEE Transactions on Knowledge and Data Engineering, 2019, 33(1): 194-208.

[13] Zhao, P., Luo, A., Liu, Y., Xu, J., Li, Z., Zhuang, F., ... & Zhou, X. Where to go next: A spatio-temporal gated network for next poi recommendation. IEEE Transactions on Knowledge and Data Engineering, 2020, 34(5): 2512-2524.

[14] Ammen E W, Al-Salihi S, Al-Salhi R. Gas Chromatography–Mass Spectrometry Combined with Successive Dilution for the Determination of Preservatives in Pharmaceuticals. Journal of Analytical Chemistry, 2021, 76(5): 621-629.

[15] Zhang, Z., Dong, M., Ota, K., Zhang, Y., & Ren, Y. LBCF: A link-based collaborative filtering for overfitting problem in recommender system. IEEE Transactions on Computational Social Systems, 2021, 8(6): 1450-1464.

[16] Brentano F E, Montaudié H, Marqueste G C, et al. Algorithmes de prise en charge thérapeutique du mélanome du stade I au stade IV. Recommandations de prise en charge du Groupe de cancérologie cutanée de la Société française de dermatologie. Annales de Dermatologie et de Vénéréologie - FMC, 2024, 4 (4): 281-288.

[17] Vijayakumar, R. Asokan, Design of Extended Hamming Code Technique Encryption for Audio Signals by Double Code Error Prediction, Journal of Information Technology and Digital World, 2021, 3(3), 179-192

[18] S.R Mugunthan, T Vijayakumar, Design of improved version of sigmoidal function with biases for classification task in ELM domain, Journal of Soft Computing Paradigm (JSCP), 2021, 3 (02), 70-82,

[19] Huimin L, Yongyi C, Hongjie N, et al. Dual-path recommendation algorithm based on CNN and attention-enhanced LSTM. Cyber-Physical Systems, 2024, 10 (3): 247-262.

[20] Kataria S, Batra U. Co-clustering neighborhood—based collaborative filtering framework using formal concept analysis. International Journal of Information Technology, 2022, 14(4): 1725-1731.

[21] Wang, F., Zhu, H., Srivastava, G., Li, S., Khosravi, M. R., & Qi, L. Robust collaborative filtering recommendation with user-item-trust records. IEEE Transactions on Computational Social Systems, 2021, 9(4): 986-996.

[22] El Fazziki, A., El Alami, Y. E. M., Elhassouni, J., El Aissaoui, O., & Benbrahim, M. Employing opposite ratings users in a new approach to collaborative filtering. Indonesian Journal of Electrical Engineering and Computer Science (IJEECS), 2022, 25(1): 450-459.

[23] Phua E J, Batcha N K. Comparative analysis of ensemble algorithms' prediction accuracies in education data mining. Journal of Critical Reviews, 2020, 7(3): 37-40.

[24] Lee K, Zeitlin A R ,Svrakic M . Making Recommendations for an Evaluation and Treatment Algorithm for Patients with Ear Fullness and No Objective Abnormalities. Otology & neurotology: official publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology, 2024, 45 (4): 447-453.

[25] Chunjing Y. Application of Recommendation Algorithms Based on Social Relationships and Behavioral Characteristics in Music Online Teaching. International Journal of Web-Based Learning and Teaching Technologies (IJWLTT), 2024, 19 (1): 1-18.