IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 5 January 2024, accepted 27 January 2024, date of publication 13 February 2024, date of current version 28 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3365549

RESEARCH ARTICLE

Intelligent Personalized Recommendation Method Based on Optimized Collaborative Filtering Algorithm in Primary and Secondary Education Resource System

XIE FEIXIANG
School of Teacher Education, Taizhou University, Taizhou, Linhai, Zhejiang 317000, China
e-mail: yhxfx0718@126.com

ABSTRACT The Internet has driven the development of online education, and the vast system of educational resources has put forward higher requirements for personalized recommendation systems. In response to this issue, this study proposes a personalized recommendation system on the ground of optimized collaborative filtering algorithms. Due to the strong interaction between collaborative filtering algorithms and users, they are often used in recommendation models. However, its defects such as cold start can weaken the performance of the model. This study introduces content recommendation algorithms to address this phenomenon. A hybrid recommendation model on the ground of the two algorithms can effectively achieve personalized recommendations. Meanwhile, this study focuses on the key modules in the overall model and utilizes standardization and dimensionality reduction operations to further reduce the computational burden on the system. Finally, to verify the reliability of the model, the study compared it with other models. The experimental results showed that the accuracy of the mixed recommendation model was 2.68% higher than that of the utility recommendation model and the rule recommendation model, respectively, and 7.99%. Therefore, the personalized recommendation model on the ground of optimized collaborative filtering algorithm proposed in the study is effective.

INDEX TERMS Collaborative filtering, content recommendation, personalization, educational resources, primary and secondary schools.

I. INTRODUCTION

The innovation in the era of online information has also strengthened the widespread application of online education. Personalized recommendation systems are often used in the business field, but are relatively less commonly used in educational systems. But this model can help teachers and students find resources of interest smoothly among the vast educational resources in the system. Due to the strong support of the country for modern education, online education has also become a popular development trend in recent years. Its educational resources are gradually enriched, and education websites in various regions are also gradually developing, covering a large amount of data [1]. However, this also leads to the phenomenon of information overload and resource overflow. Users find it difficult to find resource data of interest solely on their own, which to some extent wastes users' time and a large amount of teaching resources. It can be seen that the problems of primary and secondary education resources can be divided into five categories. Firstly, users of the teaching system need to spend a lot of effort to find suitable resources. According to a large amount of data research, nearly 30% of users are unable to find data that meets their own requirements in a short period of time, thus wasting time and energy. Secondly, there are some knowledge cognitive gaps between individuals, and the development

The associate editor coordinating the review of this manuscript and approving it for publication was Gustavo Olague.

of information networks has exacerbated the accumulation of resources. However, users' cognition may not keep up with the development of the internet, and their information literacy is low, which naturally leads to a decrease in their ability to query resources. Therefore, the gap between user cognition and knowledge development will further widen, leading to the problem of survival of the fittest and weakening the fairness of education [2]. Thirdly, the self characteristics of each teaching system are not clear enough, and the user group is much higher than the individual characteristics of the users. Therefore, high-quality resources are often overlooked due to excessive data information, resulting in waste of resource information. The effective extraction of high-quality resources can help users achieve efficient teaching. The fourth is that the recommendation function lacks interactivity and real-time performance. If resources are not updated for a long time or are delayed, the recommendation system cannot accurately provide resource recommendations to users. This also means that the education system urgently needs intelligent innovation. Meanwhile, users' personal preferences will also change over time, which also requires the system to be more real-time. Finally, the maturity of recommendation models, the prevalence of human-computer interaction, and the continuous development of networks have led to the iteration of personalized recommendation technology, making potential preference recommendation a research hotspot. For teaching systems, this should also be fully utilized to help users overcome their own limitations and gain more knowledge [3]. As a mature recommendation algorithm, collaborative filtering algorithm has been widely used. Alsaadi et al. [4] proposed to use collaborative filtering algorithm to build a user-assisted model, and introduced statistical analysis of user rating data and scoring habits based on P-moments to establish a comprehensive similarity measurement method to quantify the distance between users. Using particle swarm optimization (PSO) to determine the weight of statistics and form a neighborhood set of similar users, the user-based rating prediction is realized. Experimental results show that their CFI algorithm has certain reliability. Xiong et al. proposed a variational autoencoder enhanced graph convolutional network for collaborative filtering. Compared with traditional algorithms, this optimization algorithm solves the problem of information loss and limited representation ability, overcomes the information loss problem through prior information, and uses the generalized graph Laplace convolution kernel to solve the problem of high-frequency information loss. The experimental results show that their optimized collaborative filtering algorithm achieves good performance on real data sets [5]. ZC.A et al. proposed an improved multi-objective optimization recommendation model based on matrix decomposition to solve the problem of inefficiency of traditional recommendation algorithms in complex application scenarios. The model adopts a two-layer structure, the bottom layer uses a collaborative filtering algorithm based on matrix decomposition algorithm and regularization constraint to predict the unknown item score, and the top layer uses a multi-objective evolutionary algorithm to optimize the recommendation list. The experimental results prove that their model can provide users with accurate and novel and diverse recommendation lists in a more efficient way [6]. Although the above researches have optimized the collaborative filtering algorithm to a certain extent, they have not paid much attention to the particularity of users, and have not further solved the problem of slow convergence caused by the characteristics of the algorithm itself, and the optimization is not perfect. Therefore, a CFI optimization algorithm based on content recommendation is proposed in this paper, and the user attributes are deeply mined, and the real-time request response speed of the model is improved through standardization and other operations, and the algorithm convergence is accelerated. The contribution of the research is that according to the different characteristics of secondary students and primary students, the data identification based on age can realize personalized education resource recommendation for different users, so as to improve the personalized service level of the system. At the same time, on the premise of ensuring the accuracy and stability of the recommendation system, the exploration increases the real-time performance of the system, which will bring more real-time user experience. This recommendation algorithm with the introduction of user attributes can still maintain stable recommendation effects in the context of big data, and the running time has also been greatly improved. Therefore, the research design model can provide more accurate and interesting online education resources recommendation for primary and secondary education and students. The full text is divided into four parts. The first part is the research status of personalized recommendation systems, etc. The second part introduces the optimization of CFI algorithms and designs important modules in the model. The third part is an experiment to verify the effectiveness of the model. The fourth part is a summary of the experimental results.

II. RELATED WORKS

Personalized recommendation systems can be applied in multiple fields to provide users with intelligent experiences. Y. Zhu believes that excessive online teaching resources can actually make it difficult for teachers and students to obtain interesting data information. Therefore, personalized recommendation systems are essential for new educational systems, as they can not only handle information overload but also overcome users' own limitations. In response to this issue, this study proposes a resource recommendation model on the ground of adaptive genetic algorithm, and verifies the effectiveness of the system through relevant performance tests [7]. Xu et al.'s online learning resources have improved in terms of technology, which can better help students achieve efficient learning. On the ground of this, this study proposes to apply Ideological and Political Courses (IPC) to personalized learning resource recommendation systems, ensuring the correctness of time and format, and helping students obtain preferred resources. The research results indicate that students who use the IPC system have significantly improved their grades [8]. Liang et al. believe that the traditional teaching resource recommendation model has weak equilibrium, so they propose a recommendation algorithm based on support vector machine and trust relationship, and apply it to online sports teaching. SVM algorithm was used to classify and screen all teaching resources, then Kalman filter was used to denoise, and data fusion was carried out according to similarity. In the subsequent experimental analysis, the resource balance recommended by their method reached a maximum of 96% for students [9]. Zhang believes that sharing educational resources can improve learning outcomes. This study proposes the use of optimized CFI algorithms to establish a resource management system for English teaching resource sharing platforms. Graph neural networks and recurrent neural networks are used to construct prediction models. The former can extract user hidden data to prepare for subsequent data, while the latter can extract time series information. Finally, the effectiveness of this method was verified through experiments, with an accuracy of 88% [10]. It can be seen that with the vigorous development of online education, recommendation algorithms have been widely used in online education system resource recommendation. The essence of recommendation model is to classify big data. In order to ensure the accuracy of classification, the recommendation algorithm needs to be further studied and strengthened.

Zhaoshan et al. believe that human resources are relatively diverse, but their recommendation process often leads to defects such as data loss, ultimately resulting in a shorter recommended resource length. Therefore, this study will improve the frequent itemset mining algorithm and apply it to human resource recommendation models, while introducing positive matrix decomposition to extract hidden features and calculate similarity parameters. To improve model recommendation performance and reduce data loss, this study used CFI algorithms to optimize it. The effectiveness of the recommendation model was experimentally verified, with the longest recommended resource length and the most information contained [11].

Xu et al. learned that the fog based Internet of Things can provide reliable access to virtualized resources. The fog recommendation system has the characteristics of high timeliness and accuracy, and the significant improvement of data resources also marks the importance of this system. And due to its unpredictable and highly changing work environment, there are significant challenges at the management level. This study proposes a hybrid optimization algorithm based fiber optic gyroscope IoT resource recommendation system, and selects CloudSim for environmental simulation. Finally, its effectiveness is verified through experiments. The experimental results show that compared to the artificial bee colony algorithm, its accuracy has increased by 8% [12]. Brandstetter et al.'s recommendation system is a common problem in online education, and this problem needs to be solved by combining teaching strategies and needs, rather than simply summarizing from one aspect. On the ground

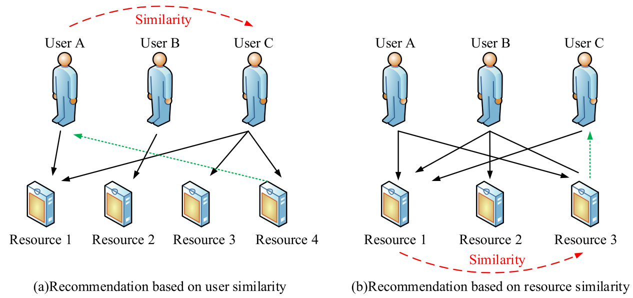


FIGURE 1. Two similarity recommendation types on the ground of CF algorithm.

of this, this study proposes to embed a recommendation model into the TEL system and design a process architecture. The effectiveness of this method has been demonstrated through experiments [13]. Wu et al. realized the impact of resource constraints on business and therefore proposed an interest behavior multiplication network. This network utilizes dynamic interaction between users and resources to predict user preferences. This study uses recurrent neural networks to extract long-term dependency features and constructs resource constrained branches to further understand the impact of project changes on user preferences. Meanwhile, this study uses mutual information to calculate similarity. Finally, the effectiveness of the model was verified through experimental analysis on the second-hand car transaction dataset and the Tmall dataset [14]. Alagarsamy et al. proposed a recommendation system on the ground of semantic fuzzy perception model, and introduced two algorithms: target keyword based data preprocessing algorithm and intelligent Anova-T residual algorithm for data preprocessing operations. The core of the system is a similarity matching algorithm on the ground of fuzzy rules, which improves the accuracy of model recommendations and verifies the effectiveness of the model through experiments [15].

A large number of studies have proven the necessity of recommendation systems. This study selects common CFI algorithms as the basis for building recommendation models and adopts content recommendation algorithms to address their shortcomings in cold start and other aspects. This can further enhance the performance of the recommendation system.

III. CONSTRUCTION OF AN INTELLIGENT PERSONALIZED RECOMMENDATION MODEL ON THE GROUND OF CFI ALGORITHM IN PRIMARY AND SECONDARY EDUCATION RESOURCE SYSTEMS

The popularization of online education has put forward higher requirements for the intelligent recommendation of educational resources, and the user objects of the system include students, teachers, and administrators. The first two mainly focus on personalized selection and recommendation of teaching resources, while the latter is the review of users and resources. However, there are still some issues with algorithm

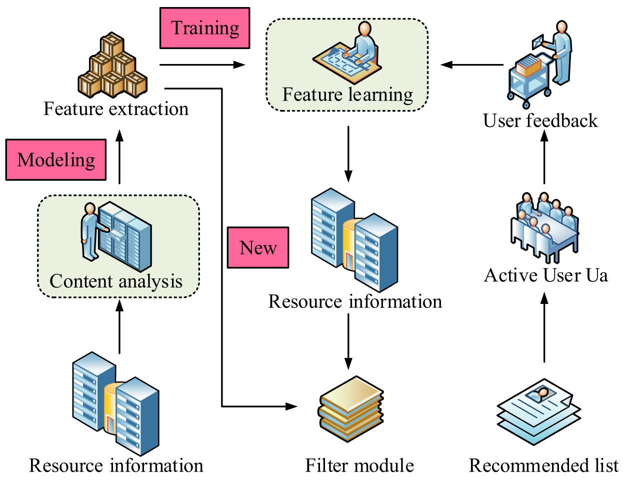


FIGURE 2. **Operation flow of content recommendation algorithm.**

accuracy and stability in the current recommendation model, which need to be further addressed and optimized.

A. OVERALL ARCHITECTURE OF PERSONALIZED RECOMMENDATION MODEL ON THE GROUND OF OPTIMIZED CFI ALGORITHM

This study selects an optimization algorithm on the ground of CFI System (CF) to achieve the intelligent recommendation function of the model. The key of this algorithm is to calculate user similarity and resource similarity, and sort and filter the data to achieve the goal of mining potential development areas for users. Therefore, the system eliminates the process of resource feature modeling and can handle unstructured resource information [16]. The CF algorithm can be divided into two types: user based and project based, as shown in Figure 1.

The algorithm in Figure 1 (a) will construct a neighborhood set with similar preferences on the ground of each user's preference for the project, and select potential interest resources to recommend to relevant users. The algorithm in Figure 1 (b) will start from the user's preference for the resource, excavate items similar to the resource, filter the items that have been followed, and finally make recommendations. The former provides users with more freshness, while the latter is relatively stable. The model should choose appropriate recommendation principle algorithms on the ground of different practical purposes. The advantages of this algorithm lie in the interaction between feedback information and users, the convenience of modeling, and the improvement of stability over time. However, this algorithm has strong interactivity with users, such as when the user's behavior data is low, the calculation accuracy of the model will plummet. Meanwhile, once a new record appears for the user, the model needs to perform repeated calculations, which increases the computational pressure of the model. When the user is a new user, it can also cause a lack of historical data for cold start phenomenon [17]. In response to the above issues, this study introduces a content recommendation algorithm, which requires the system to model the features of resources, aiming to screen and recommend on the ground of feature data. Similarity comes from the user's historical search and other records, and its basic process is shown in Figure 2.

This indicates that the system operation can be roughly divided into three major steps: first, feature extraction is performed and classified according to structured and unstructured data. For unstructured data, structural transformation is required, as shown in formula (1) [18].

$$\begin{cases} D = \{d_1, d_2, \ldots d_N\} \\ T = \{t_1, t_2, \ldots t_n\} \\ d_j = \left\{w_{1j}, w_{2j}, \ldots w_{nj}\right\} \end{cases} \tag{1}$$

Taking the selected articles as an example, in equation (1) above, D represents the set of articles. T represents the collection of articles where keywords appear. d_j represents the set of text vectors. N/n represents the number of articles and keywords, respectively. w_{nj} represents the weight of each keyword. The research selects the word frequency inverse document frequency method for weight calculation, as shown in formula (2).

$$TF - IDF(t_k, d_j) = TF(t_k, d_j) \bullet \log \frac{N}{n_k} \tag{2}$$

In equation (2) above, $TF(t_k, d_j)$ represents the frequency of the occurrence of the k-th word in article j. n_k represents the total number of articles in the set that contain k words. Therefore, the weight of this word in article j is shown in formula (3).

$$w_{k,j} = \frac{TF - IDF(t_k, d_j)}{\sqrt{\sum_{s=1}^{T} TF - IDF(t_s, d_j)^2}} \tag{3}$$

Then there is the analysis of user interests, commonly including the nearest neighbor method (KNN), decision tree algorithm (DT), linear classification algorithm (LC), and Naive Bayes algorithm (NB). The KNN algorithm lies in the comparison of attribute vectors between resources in the system. The DT algorithm is more suitable for situations where resource features are low and do not belong to unstructured attributes. The key of the LC algorithm is to find a plane in high-dimensional space that can separate two classes, and use gradient descent method to update learning parameters, as shown in formula (4) [19].

$$\overrightarrow{C_u}^{(t+1)} := \overrightarrow{C_u}^{(t)} - \eta(\overrightarrow{C_u}^{(t)} \bullet \overrightarrow{w_j} - y_{uj})\overrightarrow{w_j} \tag{4}$$

In equation (4) above, U represents the user. $\overrightarrow{w_j}$ represents the attribute vector of the item. t represents the number of iterations. y_{uj} represents the user's preference rating for resource j. η represents the learning rate, which is the size of the iteration step. $\overrightarrow{C_u}^{(t)}$ represents the user data learning parameter vector sought. $\overrightarrow{C_u}^{(t+1)}$ represents the user data learning parameter vector at the next time of update. $\overrightarrow{C_u}^{(t)} \bullet \overrightarrow{w_j}$ represents the plane composed of the item attribute vector and the user data learning parameter vector. It can be seen that

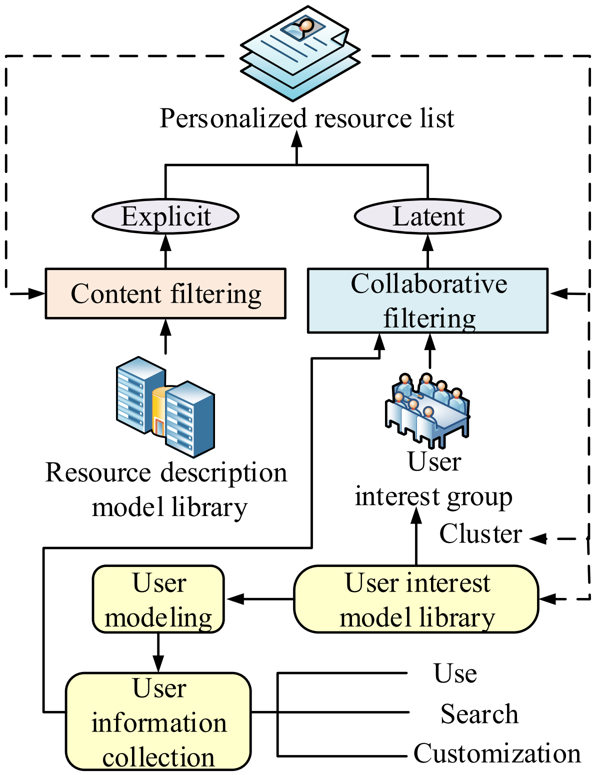


FIGURE 3. Personalized recommendation model on the ground of optimized CFI algorithm.

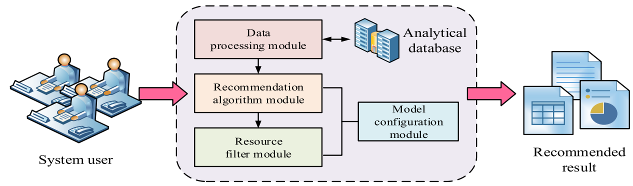


FIGURE 4. Relationship between modules in the personalized recommendation model.

LC algorithm divides data points into two different categories by constructing a linear decision boundary, and updates model parameters in continuous iteration to find the linear decision boundary that can correctly classify input data.The NB algorithm is often applied to text classification, and the attributes of each resource are independent. After completing user preference analysis, the model can recommend resources. The optimized CF algorithm can simultaneously achieve sensitivity detection of resource features and associativity of the algorithm. Common combination methods include weighted combination, feature combination, and cascading. The entire recommendation model is shown in Figure 3.

The user interest model in Figure 3 is established on the ground of user features, historical data, and natural attributes, and extracts model data through explicit or implicit information extraction. Explicit collection relies on interaction with users, who need to independently set preference options to obtain more reliable and direct data. However, this can also lead to issues with user costs, and according to convention, the number of users willing to actively provide preferences is limited, which is a breakthrough in model performance. Implicit collection does not rely on user participation, but actively extracts during system operation. Common technologies include Ajax and JS, which extract data at a low cost, but need to be improved in terms of timeliness and accuracy. Therefore, this study selected a combination of the two information extraction methods to reduce operational costs while improving data transparency and user privacy security measures. Because the implicit information extraction is based on the analysis of the anonymous behavior data of the user, it is not necessary to obtain the specific identity information or personal sensitive data of the user. Through the analysis of implicit behavior, it is possible to achieve a personalized understanding of user preferences and needs without exposing users' private information. At the same time, the user behavior data is usually aggregated and anonymized, and the statistical and desensitization of the behavioral data can reduce the sensitivity of individual users, thus reducing the privacy risk of users. The storage and processing of implicit behavioral data will also take a series of data security protection measures, such as data encryption, permission control, access audit, etc., to ensure that users' private data will not be leaked or abused. To sum up, implicit information extraction in the network model can protect users' privacy to a certain extent. This model is often represented through data structures, and research introduces Vector Space Mode (VSM) to represent text space. The core of this method is to transform the text into a vector of feature terms and weights, treating the correlation problem as a vector correlation problem. The resource description model for primary and secondary schools focuses on the consideration of subject types, age of application, and keywords. Each resource feature includes information such as name, subject, grade, and version, which needs to be integrated when making recommendations.

B. RESEARCH ON OPTIMIZATION DESIGN OF KEY MODULES IN THE OVERALL RECOMMENDATION MODEL

The personalized recommendation model roughly includes four modules, namely data processing, recommendation system, resource screening, and model conFigureuration. The interrelationships between them are shown in Figure 4.

In Figure 4, the data processing module includes data crawling and conversion, dimensionality reduction, and standardization operations. Due to the fact that the preliminary information of the model comes from third-party software such as the subject website, the model needs to further record the data in the database while crawling it to improve the convenience of subsequent queries. The data that needs to be recorded includes number, name, location, and type. When the above data is missing, regular expressions and Beautiful soap parsing should be used to extract field data from website source code [20]. The amount of user feature data is huge, so data dimensionality reduction is essential, such as directly using numerical codes to replace age or grade, which can greatly improve the operational efficiency of the model. The user profile here is developed for secondary and primary education students. The online education resource recommendation system needs to be able to make personalized recommendations according to the age and learning stage of students. For primary education students, educational resources should be more intuitive, vivid, and can arouse their interest; For students in secondary education, it needs to be more challenging and in-depth to meet their learning needs. At the same time, the content of the curriculum at the primary and secondary levels also needs to be taken into account in accordance with the syllabus and standards of each grade. For different types of learners, the online education resource recommendation system should be able to recommend suitable personalized learning paths and teaching methods according to their learning habits, interests and learning levels, so as to improve learning effects. In addition, the system should filter and screen educational resources according to the characteristics of the learning content of primary and secondary school students, according to the credibility and suitability of educational resources, and ensure the accuracy, credibility and suitability of the content. In summary, the need for online education resource recommendation systems in primary and secondary education is to provide age-appropriate and stage-appropriate course content and teaching resources, support personalized learning and teaching methods, and effectively monitor learning progress and provide feedback. Then, a large number of primary and secondary education characteristics are processed. The standardization of data, also known as binarization, converts all data into numerical values of 0 or 1. Standardization can further reduce the operational burden of the model and improve algorithm efficiency. The resource recommendation module adopts a fusion algorithm of content recommendation and CFI, and the model should use intersection for resource rating prediction when calculating similarity. The prediction process of CF algorithm and optimization algorithm is shown in formula (5).

$$\begin{cases} f_{CF} = a \cap u_1, a \cap u_2, \ldots a \cap u_n \\ f_N = a \cup u_1, a \cup u_2, \ldots a \cup u_n \end{cases} \tag{5}$$

In equation (5) above, f_{CF}/f_N represent the resource rating predictions of CF algorithm and content recommendation algorithm, respectively. a represents the user. u_n represents the set of neighboring users. This indicates that the CF algorithm first calculates the correlation between users and neighbors separately, and after obtaining the union, calculates the neighbor's rating matrix, and finally outputs an estimated rating table. The optimization algorithm used in the study is to select the intersection of users and neighbors to obtain the final output score. The latter can achieve more accurate predictions when the dataset is small, greatly improving the performance of the entire recommendation model [21]. The core of CFI is that user preferences are interrelated. Therefore, recommendations for corresponding products can be made on the ground of user similarity. The basic structure of the hybrid recommendation algorithm model is shown in Figure 5.

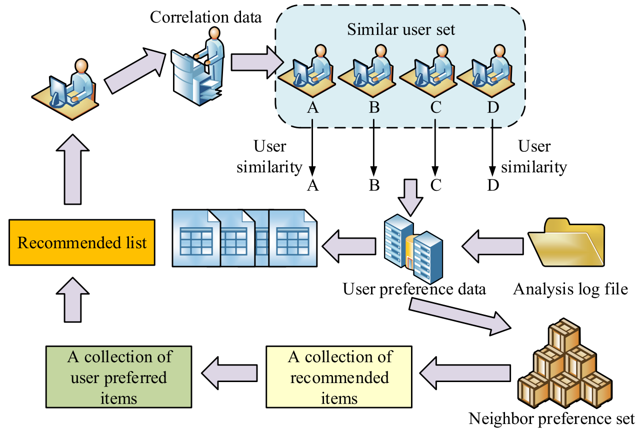


FIGURE 5. Basic structure of hybrid recommendation algorithm model.

After obtaining user browsing, retrieval, and favorite data, the model can establish a collection of neighboring users. On the ground of this data, it simulates the changing trend of user preferences influenced by neighboring users, and sets the corresponding neighbor weights according to the similarity calculation results, thereby performing the best neighbor set matching. Finally, it estimates and scores the resources. The Existing User Interest Model (EUIM) for any user, as shown in formula (6) [22].

$$EM = (w1_1, w1_2, \ldots w1_j \ldots w1_k) \tag{6}$$

In equation (6) above, $w1_j$ represents the weight value of a keyword f_j in the user interest resource data F in the model. The initial set of educational resources is represented as $D = \{d_1, d_2, \ldots d_i \ldots d_n\}$. The main feature keyword is represented as $F = \{f_1, f_2, \ldots f_i \ldots f_n\}$. The two matrix data correspond one-to-one to form a spatial vector model. When the weight w_{ij} is 0, it indicates that the keyword f_j does not exist in resource d_i. The weight matrix for the resource set is shown in formula (7).

$$DM = \begin{bmatrix} w_{11} & w_{12} & \ldots & w_{1k} \\ w_{21} & w_{22} & \ldots & w_{2k} \\ \ldots & \ldots & \ldots & \ldots \\ w_{n1} & w_{n2} & \ldots & w_{nk} \end{bmatrix} \tag{7}$$

The study selected the commonly used TF-IDF method as the weight representation method. The Potential User Interest Model (PUIM) of users is obtained by optimizing CFI algorithms, which recommend nearby user preferences with high similarity to users, as shown in formula (8).

$$PM = (w2_1, w2_2, \ldots w2_j \ldots w2_k) \tag{8}$$

In equation (8) above, $w2_j$ represents the weight of keyword f_j in the PUIM model. The difference between the PUIM model and the EUIM model is that it is independent of user initial data. In a complete education resource system, it is not enough to only recommend existing interest data to users. The huge amount of education resource data requires the model to be able to achieve potential interest recommendations. This study utilizes CFI algorithms to explore users' potential preferences. Due to the possibility of different descriptions and other information of the same resources in the system, it is unable to identify users with similar preferences. On the ground of this, this study introduces a similarity calculation method that integrates behavior and content, expressed as rating similarity $sim_{grade}(u, v)$ and content similarity, respectively. The calculation of the former is shown in formula (9).

$$sim_{grade}(u, v) = \frac{\sum i \in D_u \cap D_v \frac{1}{\log(1+|U(i)|)}}{\sqrt{|D_u|\,|D_v|}} \tag{9}$$

In equation (9) above, D_u and D_v represent the resource evaluation diversity of users u and v, respectively. $U(i)$ represents the user set that has commented on resource d_i. The content similarity $sim_{content}(u, v)$ between two users is shown in formula (10).

$$sim_{content}(u, v) = \frac{EM_u \bullet EM_v}{|EM_u| \bullet |EM_v|} \tag{10}$$

In equation (10) above, EM_u and EM_v represent the initial interest resource sets of two users, respectively. In summary, the calculation of mixed similarity is shown in formula (11).

$$sim(u, v) = \beta sim_{grade}(u, v) + (1-\beta) sim_{content}(u, v) \tag{11}$$

In equation (11) above, β represents the weighting factor, which is a similarity ratio parameter that needs to be experimentally obtained within the range of [0, 1]. A weighting factor of 0 indicates that the model only needs to consider content similarity. On the contrary, when the weighting factor is 1, it means that the model only needs to consider score similarity. The model needs to calculate the score similarity and content similarity between users separately, and then fuse the similarity values on the ground of the weighting factor values to obtain the final user mixed similarity. The user with the highest similarity to the target user will be included in the neighboring user set, and then use CFI principles to recommend potential interest resources. The weight calculation of a feature word f_i in the latent preference model is shown in formula (12) [23].

$$w2_{uj} = \sum\nolimits_{v_i \in U_u} \frac{sim(u, v_i)}{\sum_{v_i \in U_u} sim(u, v_i)} \bullet w1v_ij \tag{12}$$

In equation (12) above, $U_u = \{v_1, v_2 \ldots v_i \ldots v_k\}$ represents a set of similar users. u represents the target user. $sim(u, v_i)$ represents the similarity between the target user and any user. The Fusion User Interest Model (FUIM) is a keyword weight vector formed by the fusion of the EUIM model and the PUIM model, as shown in formula (13).

$$FM = (w3_1, w3_2, \ldots w3_j \ldots w3_k) \tag{13}$$

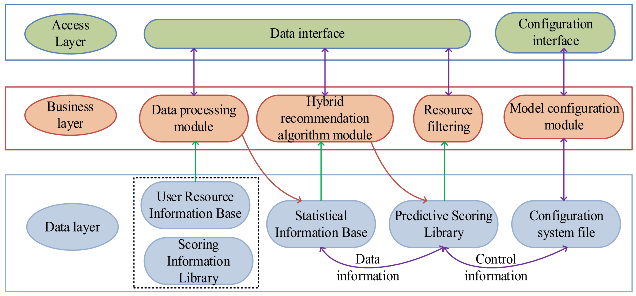


FIGURE 6. General structure of the overall personalized recommendation model.

In equation (13) above, $w3_j$ represents the weight of keyword f_j in the FUIM model. The result obtained is the weight vector of the main feature words, which can be compared with a fixed threshold to complete the output of the final recommendation list. The weight vector of the feature word f_j is shown in formula (14).

$$w3u_j = \max(w1_{u_j}, w2_{u_j}) \tag{14}$$

In equation (14) above, max represents the maximum value between the feature weights in the existing interest model and the feature weights in the potential interest model. Candidate resource $d = \{wd_1, wd_2 \ldots wd_i \ldots wd_m\}$. The general structure of the overall personalized recommendation model is shown in Figure 6.

The initial interest feature matrix is generated on the ground of the user's historical data, such as browsing, comments, likes, etc. On the ground of the existing data matrix, the current user's neighbor set can be found, and on the ground of the user behavior data in this set, the user's potential preference resource set can be predicted. According to the weight values in the initial neighbor set, the correlation between each neighbor and the target user can be obtained separately. Using the nearest neighbor set, compare its resource rating with the target user's resource rating to achieve prediction of potential preferred resources. Afterwards, it outputs a recommendation list in descending order and conducts user feedback evaluations. The study selected the Mean Absolute Error (MAE) value as the main indicator for evaluating the model, as shown in formula (15).

$$MAE = \frac{\sum_{i=1}^{n} |q_i - p_i|}{n} \tag{15}$$

In equation (15) above, p_i and q_i respectively represent the predicted user rating and the actual user rating. n represents the number of scoring resources. The magnitude of this value is inversely related to the performance of the model. F1 value, as a comprehensive evaluation index that can reflect both model accuracy and recall performance, is also often used in algorithm testing, as shown in formula (16).

$$F1 = P^*R^*2/(P+R) \tag{16}$$

In formula (16), P represents the accuracy of the algorithm, and is the ratio of the correct quantity retrieved to the relevant quantity. R represents the algorithm recall rate, which is the ratio of the correct amount of resources extracted to the sample resource review. However, the F1 value is based on the condition that the influence of the recall rate is the same. When the influence of the two cannot be treated equally, parametric measurement is required, as shown in Formula (17).

$$F = \frac{\left(1+\beta^2\right)^* P^* R}{\beta^{2*} P + R} \tag{17}$$

In formula (17), β represents the weight parameter, when the value is greater than 1, the recall rate has a greater impact; when the value is in the interval (0,1), the accuracy rate has a greater impact; when the value is equal to 1, it means that the two indicators are equally important.

IV. PERFORMANCE VERIFICATION ANALYSIS OF RECOMMENDATION MODEL ON THE GROUND OF HYBRID COLLABORATIVE FILTERING ALGORITHM

To further validate the performance of the optimized CFI recommendation model, a model simulation comparison experiment was conducted before and after optimization to verify its effectiveness. And the recommendation model was applied to the actual platform for visual analysis, clarifying the number of successful recommendations and clicks of the model within a week. Meanwhile, the optimized personalized recommendation model was compared with other recommendation models in terms of performance.

A. ANALYSIS OF MODULE PERFORMANCE VERIFICATION ON THE GROUND OF HYBRID CFI ALGORITHM RECOMMENDATION MODEL

Since different operating environments and parameters can have a significant impact on algorithm performance, the computer parameters were selected in the research, and the specific system operating environment and other parameters were selected, as shown in Table 1.

The computing resources to run the algorithm, namely CPU and GPU, will directly affect the running speed and parallelization ability of the algorithm. The size and speed of the computer's available memory will also affect the performance of the algorithm. For example, when the algorithm requires a large amount of computation and needs to store a large amount of data, but the memory of the running environment is small, the performance may be degraded or the running may fail. After the MySQL 5.5 database is created, further JDBC-ODBC connections need to be completed with Eclipse. When crawling data, it is necessary to use JAVA multithreading to improve the collection speed. The main module for optimization in this study is the recommendation algorithm part of the model. Therefore, this study takes the data processing module and resource filtering module as fixed parts, and only makes corresponding modifications to the recommendation algorithm. The structure of a recommendation system that uses CFI algorithm and content recommendation algorithm separately is shown in Figure 7.

TABLE 1. System operating environment and parameter Settings.

Environment	Argument
Operating system	Window10 (64 bit)
Visual interface environment	Eclipse integrated development
Programming environment	JAVA
Front-end interface	JavaScript+MVC+SSH
Crawling data tool	Python2.7+ Beautiful
Archive	MySQL5.5
Model deployment	Tomact7.0

In Figure 7, the boxed parts are the parts with differences between the two models. This indicates that there are differences in the data layer between the content recommendation algorithm and the CFI algorithm, both of which come from statistical information databases. However, the CFI algorithm has added a scoring information database module. On the ground of the above two model structures, the overall operation process of the hybrid recommendation algorithm can be designed. The data source of the data processing module is still the user resource information library and user rating information library. The recommendation algorithm adopts a fusion algorithm of content recommendation and CFI recommendation. The model has added a resource filtering module, whose data source is the predicted rating system, which is the estimated value of resource ratings in the recommendation algorithm. This indicates that resource rating data is a key performance indicator of the recommendation model. This study introduces Mean Absolute Error (MAE) and F-value as specific evaluation indicators. Meanwhile, selecting user data from primary and secondary education auxiliary institutions in the education recommendation system, the experiment is divided into analyzing the overall recommendation performance of the system, analyzing the recommendation performance of new users, and analyzing the stability of the model. The specific experimental parameters are shown in Table 2.

The experimental test dataset consists of 1000 rating information from 50 users on 500 resource data. The target users are 10 randomly selected users, and it is assumed that 80% of the data is known, while the remaining data represents unknown data that needs to be predicted. The performance results of the three models are shown in Figure 8.

Figure 8 (a) shows that the average absolute error of the hybrid recommendation algorithm model is significantly lower than the other two types of algorithm models, with an average MAE value of 0.895. The content recommendation

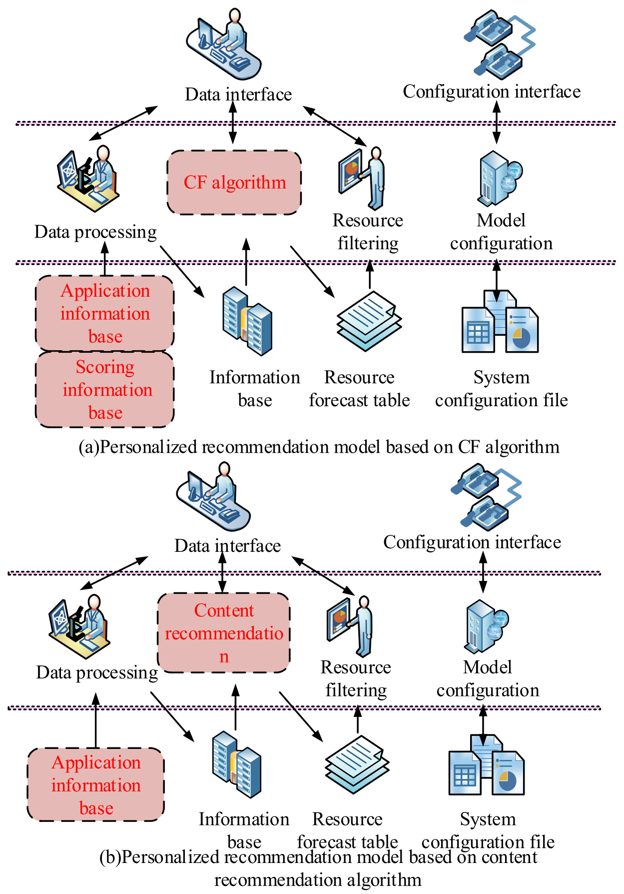


FIGURE 7. The recommended system model structure using CF algorithm and CR algorithm alone.

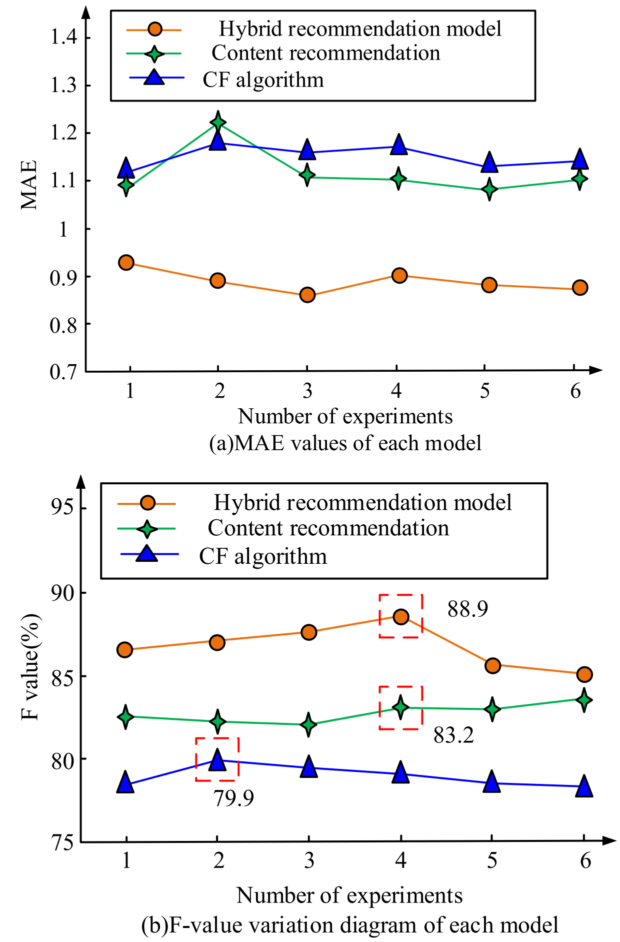


FIGURE 8. Overall recommended performance comparison diagram of each model.

TABLE 2. Specific parameters of user data.

/	Index	Argument
Initial data message	Total user base	426
	Total active users	356
	Total resources	2322
	Total score data	1980
Experimental selection data information	Number of users	50
	Quantity of educational resources	500
	Scoring information	1000
	Known information	800
	Unknown information	200

algorithm and CF algorithm achieved 1.123 and 1.141 respectively, higher than the hybrid recommendation algorithms of 0.228 and 0.246. Among them, the changes in the hybrid recommendation algorithm model and the CF algorithm model are relatively stable. The content recommendation algorithm model fluctuated significantly before and after the second experiment, but remained stable in subsequent regions and remained at a level below the error of the CF algorithm. The F-value is a comprehensive evaluation indicator used to balance accuracy and recall, which can be used to evaluate model quality. The higher the value, the better the recommendation performance of the model. Figure 8 (b) shows that the F value of the hybrid recommendation algorithm model has always been at the optimal level. In the fourth experiment, the hybrid recommendation model and content recommendation model reached the highest level, with 88.9% and 83.2%, respectively. After reaching the highest F-value in the second experiment, the CF algorithm showed a gradual downward trend, with its highest value being 9% lower than the mixed recommendation model. This study further validated and analyzed the performance of new user recommendation and the stability of the model. For the recommendation experiment for new users, the study randomly selected 60 users without rating records and 100 users with rating data as the test set and training set, respectively. For the verification of stability, the scoring data is divided into 5 parts, with a training to testing ratio of 4:1. The experimental results are shown in Figure 9.

Figure 9 (a) shows the recommendation performance of the hybrid recommendation algorithm when facing new users.

Its F-value varies within the (84,86) range and is relatively s, with a maximum and minimum difference of 1.1%. Compared with the overall recommended maximum value of F in the system, it has decreased by 3%. Overall, even for users without any rating data, the hybrid recommendation model can achieve more accurate personalized recommendations with relatively small changes. Figure 9 (b) shows the changes in MAE values of the model from day 1 to day 25. This indicates that the relative variation of the mixed recommendation model is minimal, with an average absolute error within the (0.8,1) range, and the difference between the maximum and minimum values is only 0.04. The CFI recommendation system has the highest fluctuation, showing an overall trend of first decreasing and then increasing. The MAE difference is 0.6, which is 0.56 higher than the fluctuation situation of the mixed recommendation model. The MAE value of the content recommendation model also changes over time, with a slightly lower magnitude compared to the CF algorithm, with a MAE difference of 0.2. Although the amplitude of change is only 0.16 higher than that of the hybrid algorithm, its MAE mean is still at a relatively high level. Further applying the algorithm to practical platforms can obtain visual results of the recommendation model, as shown in Figure 10.

Figure 10 (a) shows that in practical application, the number of successfully recommended users of the model is relatively stable, basically around 150000 users. However, the number of successfully recommended users fluctuates slightly, reaching a maximum of 600, which is about 190 different from the minimum value. This indicates that the stability of the model is relatively good, and there will be no sudden drop or rise phenomenon, with high robustness. Figure 10 (b) shows the relationship between the successful resource recommendation of the model and its actual click through volume, indicating that the difference between the two is relatively small. Due to the different coordinate axes, there may seem to be a significant difference in visual perception, but in reality, the difference is very small. This indicates that the recommendation system of the model is relatively excellent and can accurately grasp the potential areas of interest of users.

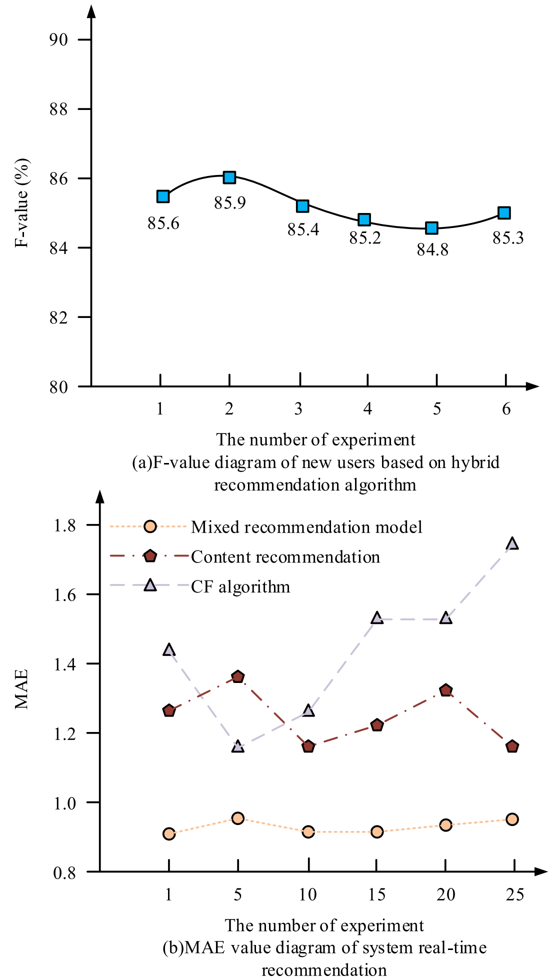


FIGURE 9. Comparison of new user recommendation and stability performance of each model.

B. PERFORMANCE VERIFICATION ANALYSIS OF HYBRID RECOMMENDATION ALGORITHM MODEL AND OTHER MODELS

This study selected the Meclass dataset as the experimental basis, with a total of 281 users and 1268 resources [24]. There were a total of 30000 rating data, with a training to test ratio of 8:2. This study introduces utility recommendation and rule recommendation models for comparison. The performance of each algorithm model is shown in Table 3. on the ground of the number of different neighbor sets K.

Table 3 shows that the performance indicators of each model increase with the increase of K value, and the hybrid recommendation algorithm used in this study has always been in the best neutral performance of the model. When the K value is 60, the accuracy of the hybrid recommendation algorithm is 2.68% and 7.99% higher than the utility recommendation model and rule recommendation model, respectively, with a recall rate of 2.75% and 8.44% higher. Under the conditions of K values of 20/40/60, the average accuracy of the hybrid recommendation algorithm model is 87.75%, and the average recall rate is 86.16%. The average accuracy of the utility recommendation model is 85.24%, and the average recall rate is 83.51%. The average accuracy of the rule recommendation model is 81.25%, and the average recall rate is 80.28%. This indicates that the performance of the rule recommendation model is the worst because the recommendation timeliness is relatively poor and there is also a problem of low personalization. The utility recommendation model is relatively less flexible due to rule recommendation, but users must input the utility function, which belongs to static recommendation. The hybrid recommendation model used in this study not only preserves the good performance of CFI algorithms in interacting with users, but also utilizes content recommendation algorithms to solve problems such as cold start. Therefore, among the three models, the performance of

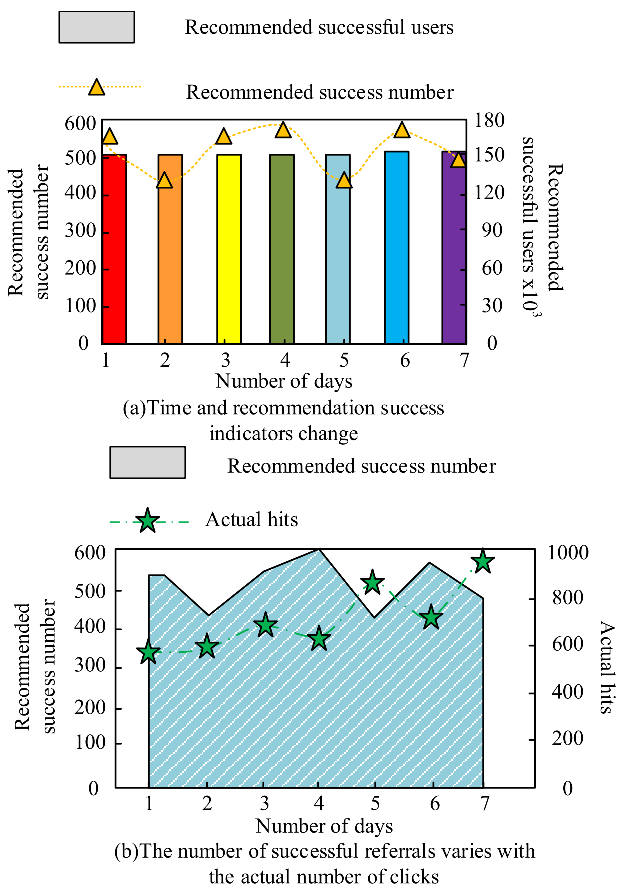


FIGURE 10. **Visual analysis of the application of the actual platform recommendation system.**

TABLE 3. **Changes of performance indexes of different models under varying K values.**

K	Model	Accuracy rate(%)	Recall rate(%)
20	Mixed recommendation	84.61	83.22
	Utility recommendation	81.47	79.65
	Rule recommendation	79.26	77.89
40	Mixed recommendation	87.49	85.24
	Utility recommendation	85.79	83.61
	Rule recommendation	82.33	81.40
60	Mixed recommendation	91.15	90.01
	Utility recommendation	88.47	87.26
	Rule recommendation	83.16	81.57

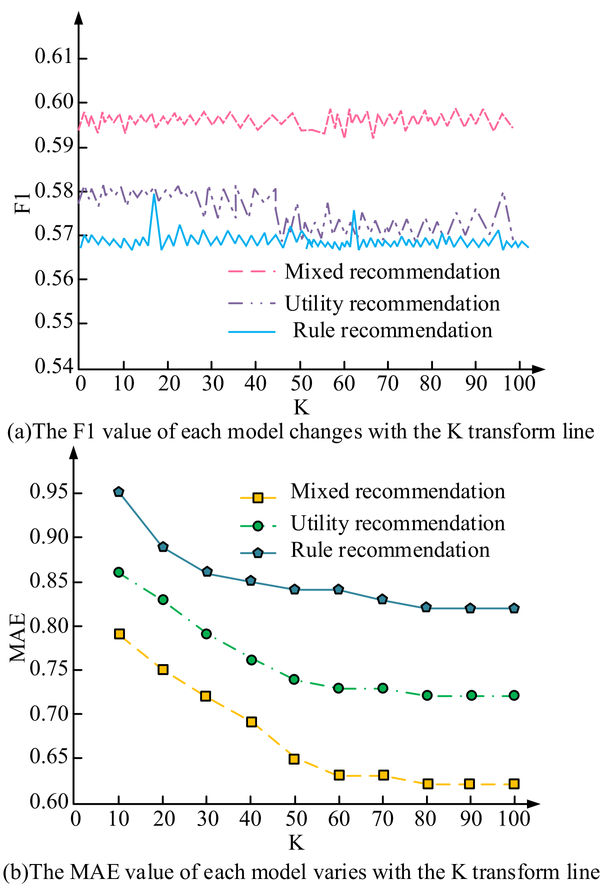
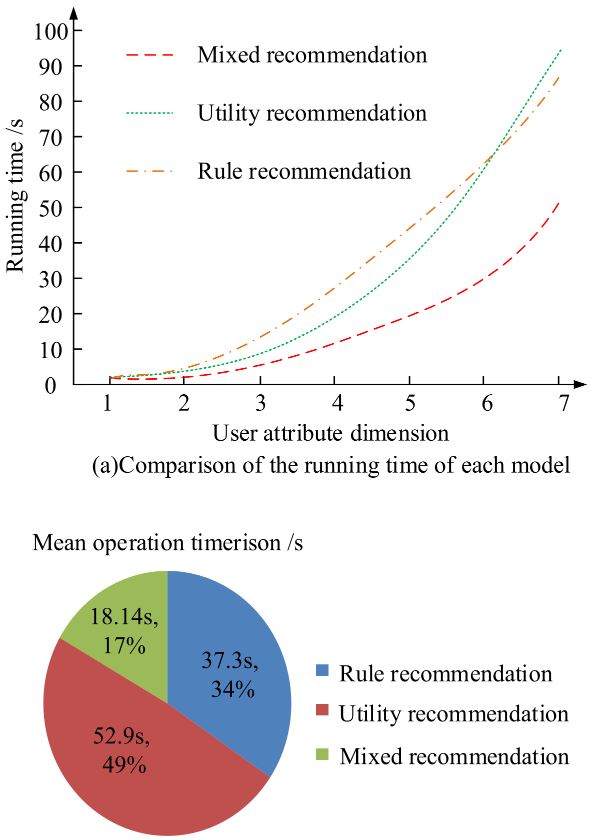


FIGURE 11. **Performance variation diagram of each model.**

this model is the best. This study further introduces the F1 parameter, and the performance changes of each model are shown in Figure 11.

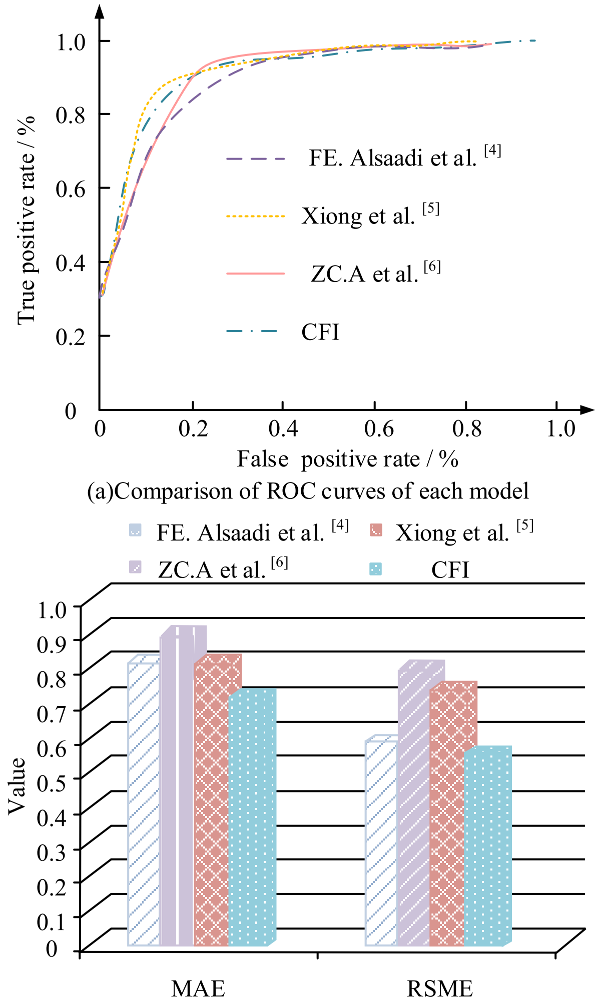
Figure 11 (a) shows that the F1 value of the mixed recommendation model is relatively stable with little fluctuation as the K value changes, and most of it is distributed in the range of 0.59 and 0.6, with an average F1 value of 0.594. The utility recommendation model and rule recommendation model both vary with the K value. Among them, the utility recommendation model shows a gradually decreasing trend, with a maximum F1 value of 0.58. Although the overall change of the rule recommendation model is not significant, there may be a sudden decrease or increase in the phenomenon, and its F1 value is always at the lowest level. Figure 11 (b) shows that the MAE values of each model decrease with the increase of K value, the variation of the mixed recommendation model is steeper, and the other two types of algorithms have relatively smoother changes. At a K value of 100, the MAE of the three models ranges from high to low, with values of 0.85/0.75 and 0.63, respectively. This indicates a significant gap among the three. The average absolute error difference between the mixed recommendation model and the rule recommendation model is 0.22. The research further compares the time complexity of each model, and the experimental results are shown in Figure 12.



(a)Comparison of the running time of each model

(b) Operation time mean comparison

FIGURE 12. **Comparison of time complexity of each algorithm model.**



(a)Comparison of ROC curves of each model

(b) Error comparison of each model

FIGURE 13. **Performance comparison of different CF optimization models.**

As can be seen from Figure 12, the collaborative filtering hybrid algorithm proposed in this study has the best time complexity performance, and its running time rises most slowly with the user attribute dimension. The average running time is 18.14s, which is 65.71% and 51.38% lower than the average running time of utility recommendation model and rule recommendation model respectively. Although the module of the hybrid model is more complex, the dimensionality of the calculated parameters is reduced due to its standardization and other processing, and the feature extraction speed of user attributes is also enhanced. The other models are relatively simple and perform well when the user attribute dimension is low. However, with the increase of the user attribute dimension, the processing of the model is difficult, which leads to the increase of the running time. Finally, this model is compared with other CF optimization models [4], [5], [6]. The experimental results are shown in Figure 13.

It can be seen from Figure 13(a) that the ROC curves of all models perform well, but the CFI model proposed in this study and the CF model proposed by Xiong et al. [5] are relatively superior to the other two CFI models and have better ability to distinguish data. However, in Figure 13(b), Xiong et al. [5] 's model has the highest mean MAE and RSME, reaching 0.87 and 0.77 respectively, while the MAE values of the other two CFI models are around 0.79. Therefore, the MAE error of the proposed CFI optimization model is 16.57% lower on average than that of the other three CFI models. The mean RSME values of CFI model of Alsaadi et a1. [4] and CFI model of ZC.A et a1. [6] are 0.53 and 0.71, respectively. Therefore, the RMSE error of the proposed CFI optimization model is 31.37% lower than that of the other three CFI models. Therefore, the CFI hybrid model proposed in this study has the best comprehensive performance. In summary, it can be seen that the hybrid algorithm on the ground of content recommendation and CFI used in this study has the best overall performance, making it the best choice for educational recommendation systems.

V. CONCLUSION

With the growth of online educational resources, personalized recommendation models have become a popular research direction. On the ground of this, this study proposes the construction of a recommendation model using CFI algorithms. Meanwhile, it introduces content recommendation algorithms to solve problems such as cold start of the model. Finally, comparative analysis experiments were conducted on the performance of various models before and after model optimization. Firstly, a comparison was made between the pre and post optimization models, and the main difference between the models was the introduction of various modules in the data layer. Subsequently, the study conducted six experimental analyses on the pre and post optimization models (content recommendation model, CFI model, and hybrid recommendation model). The experimental results show that the MAE value of the mixed recommendation model is lower than that of the content recommendation and CF models by 0.228 and 0.246, and the maximum F value of the mixed recommendation model is 9% higher than that of the CF model. In the experiments of the model's recommendation for new users and system stability, the F value of the former decreased by 3% compared to the overall recommendation, but tended to stabilize overall. The MAE difference over time is lower than the CF model by 0.56. Finally, the study introduced utility recommendation and rule recommendation models for comparative experimental analysis. The experimental results show that the accuracy of the hybrid recommendation algorithm is 2.68% and 7.99% higher than the utility recommendation model and rule recommendation model, respectively, with a recall rate of 2.75% and 8.44% higher. The mixed recommendation model has a relatively stable F1 value with an average F1 value of 0.594 as the K value changes. However, due to the small user data volume system, the model training is not perfect, so the user attribute dimension should be improved and parameters should be adjusted in the future to further make the model more mature.

REFERENCES

[1] C. Jung and L. Mertins, "Quality and accountability in the online business education environment," *Computer*, vol. 54, no. 6, pp. 49–52, Jun. 2021, doi: 10.1109/MC.2021.3069531.

[2] A. Akgül, "A resource recommendation for improving musical expression and narration in piano education: An examination of Loeschhorn Op. 65 etudes," *Educ. Res. Rev.*, vol. 16, no. 5, pp. 189–201, May 2021.

[3] Y. Dai and J. Xu, "Study of online learning resource recommendation based on improved BP neural network," *Int. J. Embedded Syst.*, vol. 14, no. 2, p. 101, Mar. 2021, doi: 10.1504/ijes.2021.113834.

[4] F. E. Alsaadi, Z. Wang, N. S. Alharbi, Y. Liu, and N. D. Alotaibi, "A new framework for collaborative filtering with P-moment-based similarity measure: Algorithm, optimization and application," *Knowl.-Based Syst.*, vol. 248, Jul. 2022, Art. no. 108874, doi: 10.1016/j.knosys.2022.108874.

[5] X. Xiong, X. Li, Y. Hu, Y. Wu, and J. Yin, "Handling information loss of graph convolutional networks in collaborative filtering," *Inf. Syst.*, vol. 109, Nov. 2022, Art. no. 102051, doi: 10.1016/j.is.2022.102051.

[6] Z. Cui, P. Zhao, Z. Hu, X. Cai, W. Zhang, and J. Chen, "An improved matrix factorization based model for many-objective optimization recommendation," *Inf. Sci.*, vol. 579, pp. 1–14, Nov. 2021, doi: 10.1016/j.ins.2021.07.077.

[7] Y. Zhu, "Personalized recommendation of educational resource information based on adaptive genetic algorithm," *Int. J. Rel., Quality Saf. Eng.*, vol. 30, no. 2, Apr. 2023, Art. no. 2250014, doi: 10.1142/S0218539322500140.

[8] Y. Xu and T. Chen, "The design of personalized learning resource recommendation system for ideological and political courses," *Int. J. Rel., Quality Saf. Eng.*, vol. 30, no. 1, pp. 2250020–2250030, Feb. 2023, doi: 10.1142/S0218539322500206.

[9] X. Liang and J. Yin, "Recommendation algorithm for equilibrium of teaching resources in physical education network based on trust relationship," *J. Internet Technol.*, vol. 23, no. 1, pp. 133–141, Jan. 2022, doi: 10.53106/160792642022012301014.

[10] J. Zhang, "An English teaching resource recommendation system based on network behavior analysis," *Sci. Program.*, vol. 2021, pp. 1–9, Nov. 2021, doi: 10.1155/2021/6191543.

[11] Z. Liu, Y. Ma, H. Zheng, D. Liu, and J. Liu, "Human resource recommendation algorithm based on improved frequent itemset mining," *Future Gener. Comput. Syst.*, vol. 126, pp. 284–288, Jan. 2022, doi: 10.1016/j.future.2021.08.017.

[12] Z. Xu, H. Qin, S. Yang, and S. M. Arefzadeh, "A new approach for resource recommendation in the fog-based IoT using a hybrid algorithm," *Comput. J.*, vol. 66, no. 3, pp. 692–710, Oct. 2021, doi: 10.1093/comjnl/bxab189.

[13] C. Brandstetter, F. Suda, L. Papariello, F. Dopler, and B. Göschlberger, "Exploring learning resource recommendation approaches for secondary education," in *Proc. 23rd Int. Conf. Inf. Integr. Web Intell.*, Dec. 2021, pp. 28–33, doi: 10.1145/3487664.3487774.

[14] Q. Wu, T. Zhang, Z. Cui, and J. Yang, "Interest-behaviour multiplicative network for resource-limited recommendation," 2020, *arXiv:2009.13249*.

[15] R. Alagarsamy, R. Arunpraksh, S. Ganapathy, A. Rajagopal, and R. J. Kavitha, "A fuzzy content recommendation system using similarity analysis, content ranking and clustering," *J. Intell. Fuzzy Syst.*, vol. 41, no. 6, pp. 6429–6441, Dec. 2021, doi: 10.3233/jifs-210246.

[16] B. Alhijawi, G. Al-Naymat, N. Obeid, and A. Awajan, "Novel predictive model to improve the accuracy of collaborative filtering recommender systems," *Inf. Syst.*, vol. 96, Feb. 2021, Art. no. 101670, doi: 10.1016/j.is.2020.101670.

[17] S. Samantaray and A. Sahoo, "Prediction of runoff using BPNN, FFBPNN, CFBPNN algorithm in arid watershed: A case study," *Int. J. Knowl.-Based Intell. Eng. Syst.*, vol. 24, no. 3, pp. 243–251, Sep. 2020, doi: 10.3233/kes-200046.

[18] W. Zhu, J. Zhang, and J. Romagnoli, "General feature extraction for process monitoring using transfer learning approaches," *Ind. Eng. Chem. Res.*, vol. 61, no. 15, pp. 5202–5214, Apr. 2022, doi: 10.1021/acs.iecr.1c04565.

[19] P. Trajdos and R. Burduk, "Linear classifier combination via multiple potential functions," *Pattern Recognit.*, vol. 111, Mar. 2021, Art. no. 107681, doi: 10.1016/j.patcog.2020.107681.

[20] W. Qi and D. Li, "A user experience study on short video social apps based on content recommendation algorithm of artificial intelligence," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 35, no. 2, Feb. 2021, Art. no. 2159008, doi: 10.1142/s0218001421590084.

[21] J. Wang, A. K. Sangaiah, and W. Liu, "A hybrid collaborative filtering recommendation algorithm: Integrating content information and matrix factorisation," *Int. J. Grid Utility Comput.*, vol. 11, no. 3, p. 367, Jan. 2020, doi: 10.1504/ijguc.2020.107616.

[22] N. Van Dat, P. Van Toan, and T. M. Thanh, "Solving distribution problems in content-based recommendation system with Gaussian mixture model," *Appl. Intell., Int. J. Artif. Intell., Neural Netw. Complex Problem-Solving Technol.*, vol. 52, no. 2, pp. 1602–1614, May 2022.

[23] C. Ma, "Content feature extraction-based hybrid recommendation for mobile application services," *Comput., Mater. Continua*, vol. 71, no. 3, pp. 6201–6217, 2022, doi: 10.32604/cmc.2022.022717.

[24] C. Wu, S. Liu, Z. Zeng, M. Chen, A. Alhudhaif, X. Tang, F. Alenezi, N. Alnaim, and X. Peng, "Knowledge graph-based multi-context-aware recommendation algorithm," *Inf. Sci.*, vol. 595, pp. 179–194, May 2022, doi: 10.1016/j.ins.2022.02.054.

XIE FEIXIANG was born in Yuhuan, Zhejiang, in 1973. He graduated from Zhejiang Normal University. He is with the School of Teacher Education, Taizhou University. His main research interests include primary education and teacher education.