Construction of Educational Resource System Based on Mahout Collaborative Filtering Algorithm

Guiyong Zhu
Xingzhi College, Zhejiang Normal University, China
zgyznu@163.com

Abstract: *The Internet has provided great convenience for the development of the education industry, but it has also brought about problems such as difficulties in selecting educational resources, difficulty in searching information, etc. To address these issues, the research constructs an educational resource Recommendation System (RS) grounded on Mahout Collaborative Filtering (CF) hybrid algorithm to provide efficient resource recommendations for users. During the construction of the system, the research also combines Multi-Dimensional Feature Fusion (MDFF) and deep learning personalized course recommendation methods to optimize courses and enhance the system's ability to integrate multiple data. The experiment outcomes indicate that the hybrid algorithm has higher recommendation accuracy compared to CF algorithm, Fuzzy C-Means (FCM) algorithm, and the combination of knowledge graph completion and RS algorithm. The average recommendation accuracy of the four algorithms is 88.88%, 79.11%, 71.11%, and 65.53%, respectively. In addition, empirical analysis of the constructed educational resource RS reveals that the proposed hybrid algorithm has a lower Receiver Operating Characteristic (ROC) curve area value of 0.8984 and an F1-value of 0.8298, indicating good recommendation performance and superior performance compared to other comparative educational resource RSs. The above information indicates that the educational resource RS grounded on Mahout CF hybrid algorithm has certain stability and can offer customized suggestions for resources tailored to individual users in a timely manner. This research provides a practical method for online education on the Internet, which will help further improve online education in the education platform in the future*.

Keywords: *Multi-dimensional feature fusion, collaborative filtering algorithm, personalized recommendation, educational resource system, deep learning*.

Received April 3, 2025; accepted September 15, 2025
https://doi.org/10.34028/iajit/23/2/1

1. Introduction

Taking advantage of the Internet, intelligent education has also entered the forefront of the advancement of the era [6]. With the emergence of Online Education (OE) platforms, it has also brought huge challenges to the construction of educational resources, such as how to achieve accurate educational information recommendation and how to solve problems such as educational information overload, which urgently require systematic research [2]. At present, school education is not limited to online and offline classes, but more uses the "internet plus education" model to output knowledge to students. Relying on various OE platforms, people can learn in real-time anytime and anywhere through their mobile phones and computers. The currently popular OE platforms include Massive Open Online Course (MOOC) university, bilibili, tencent classroom, etc., all of which provide rich learning resources [19]. Learning resources not only include university courses in different subjects, but also practical skills teaching in society, which have effectively promoted OE. With the continuous advancement of OE, an increasing quantity of users prefer online course learning to meet their different course learning needs [13]. With the increase of OE resources and users, education platforms have encountered problems such as information overload and inaccurate personalized recommendations [8]. In response to these issues, scholars have also attempted to improve the Recommendation System (RS) by using knowledge graph-based learning resource recommendation methods and priority sampling techniques for path selection. However, there are still problems such as unclear semantic analysis and untimely search engines. Therefore, in response to the current shortcomings of educational resource RSs, research aims to effectively integrate Multi-Dimensional Feature Fusion (MDFF) and deep learning to furnish tailored and precise recommendations for users. Grounded on this, a hybrid educational resource RS is constructed by combining Mahout Collaborative Filtering (CF) algorithm. It is expected the system to improve the quality of resource recommendations and adjust user personalized needs at any time. The innovation of the research mainly lies in the following three aspects. Firstly, the deep combination of MDFF and CF enhances the accuracy and personalization of recommendations by extracting rich features of course text, attributes, and user preferences. The second is to introduce a deep reinforcement learning, Deep Q-Network (DQN) module to capture changes in user interests in real time and dynamically adjust recommendation strategies, improving the real-time and adaptability of recommendations. The third is to adopt an end-to-end training approach to optimize the collaborative work of MDFF, DQN, and Mahout CF modules, achieving overall performance improvement.

2. Related Works

Against the backdrop of rapid digitalization and informatization, educational resources are undergoing unprecedented changes [5]. OE platforms can offer a diverse range of resources, but they cannot handle the large amount of traffic and customized personalized learning plans. Therefore, it is needful to enhance the education resource system. Sanabria *et al*. [16] conducted research on the types of open licenses for 5R activities in order to promote Open Educational Resource (OER) platforms that meet students' needs. The research results indicated that the adaptation and reuse distribution of OERs could promote the cultivation of complex thinking skills. Chen *et al*. [3] used Machine Learning (ML) algorithms to optimize the personalized recommendation ability of educational RSs. The research results indicated that ML algorithms could provide timely feedback to adjust learning plans and were applicable to various types of education. Raj and Renumol [15] proposed a learning path recommendation method that focused on knowledge construction and learning performance analysis in order to help students find learning paths in a timely manner. Research showed that this method had good performance based on an expected learning duration of 27.8%. Mikroyannidis and Papastilianou [11] conducted a study on leveraging OERs to broaden and enhance curriculum and resource sharing within Greek public management, aiming to tackle challenges like the insufficient openness in the public sector. The results of the case study indicated that civil servants strengthened existing OERs through active participation. Oladokun *et al*. [12] investigated the transformative effects of emerging technologies like the Metaverse and Meta AI on libraries within higher education institutions, with the objective of enhancing their service delivery. Their findings suggested that these libraries must adapt to change and prioritize user-centric design approaches.

With the swift advancement of the education industry, users have put forward higher requirements for educational resource RSs. They not only require recommended courses to meet user preferences, but also to provide timely feedback on relevant resources during search. Therefore, optimization is needed for recommendation algorithms related to the education system. Among them, Xiang *et al*. [20] used artificial intelligence to align user interests with video metadata to tackle the problem of poor user experience in video RSs, thereby promoting similar video recognition. The experiment outcomes showed that this method improved the recommendation accuracy. Li *et al*. [9] introduced an RS approach that utilizes a quantum nearest neighbor CF algorithm to address challenges like the cold start problem and data sparsity. Their research revealed that this method attained the fastest recommendation speed across 8 public datasets. To raise the personalized ability of RSs, Shang *et al*. [17] combined semantic understanding with user preferences using large-scale language models. This system utilized advanced natural language processing capabilities of Large Language Models (LLMs) to solve traditional recommendation methods. The experiment outcomes showed that compared to the best performing baseline, the LLM driven system achieved 8.6% Normalized Discounted Cumulative Gain (NDCG) growth. Hasan and Ferdous [7] adopted text to number conversion and cosine similarity methods to enhance the advantages of hybrid movie RSs. The experiment outcomes showed that the accuracy of this method was improved by 97%, providing promising results for personalized recommendation for users. Qin *et al*. [14] adopted a joint recommendation model learning framework to address the risk of data leakage in multimedia course recommendation. When protecting privacy, a federated learning system aggregation model was used to coordinate round partitioning. The experiment outcomes showed that this method could effectively operate in privacy protection mode.

In conclusion, both domestic and international researchers have carried out thorough investigations into education system recommendations, and have made achievements in personalized recommendations. However, currently there is not much research on educational RSs for improved intelligent recommendation algorithms in terms of audiovisual and book resources. Therefore, a new educational resource RS grounded on Mahout CF hybrid algorithm is studied to optimize the system by utilizing the advantages of personalized and timely recommendation of Mahout CF hybrid algorithm, so as to allocate resources reasonably and further improve user learning efficiency. Compared with existing deep or hybrid RSs, the research method demonstrates significant innovation in multiple aspects. Firstly, the MDFF module extracts rich semantic features from multiple dimensions such as course text description, course attributes, and user preferences through deep learning techniques. These features not only enhance the similarity calculation between users and projects, but also improve the accuracy and personalization of recommendations. Secondly, the introduction of DQN module enables the RS to capture changes in user interests in real time and dynamically adjust recommendation strategies based on users' immediate feedback. This dynamic adjustment mechanism significantly improves the real-time and adaptability of recommendations, enabling the RS to better adapt to changes in user interests and provide more personalized recommendation services. Finally, the system ensures the collaborative work between MDFF, DQN, and Mahout CF modules through end-to-end training and optimization. During the training process, the study jointly optimized feature extraction, dynamic adjustment, and similarity calculation, thereby achieving performance improvement of the entire RS. This end-to-end training approach not only improves the overall performance of the RS, but also ensures collaborative optimization between various modules.

3. Combining Multiple Methods to Optimize Mahout CF Algorithm for Educational Resource System

3.1. Personalized Course Recommendation Model Grounded on Multi-Dimensional Feature Fusion (MDFF) and Deep Learning

With the swift advancement of "internet-plus education," the personalized recommendation function of extensive open online learning programs has actively promoted the scientization and intellectualization of education in China. It has also brought many problems to OE, such as massive user information overload, inaccurate new user recommendation information, etc., [1]. In response to the above issues, research is being conducted on the use of MDFF methods for extracting and processing course feature preferences from different dimensions. MDFF has multiple advantages such as data integration, image recognition, and sentiment analysis. The multi-dimensional feature extraction diagram is shown in Figure 1.

In Figure 1, multi-dimensional feature extraction can capture the positive and negative information of the text during the preprocessing stage and encode it to obtain preliminary feature vectors when using bidirectional Long Short-Term Memory (LSTM) networks. During Deep Pyramid Convolutional Neural Network (DPCNN) extraction, local semantic features of the text can be gradually extracted and perceived to expand. Finally, multi-head attention mechanism is utilized for rich semantic capture. Multi-dimensional feature extraction is used for course text data processing, as shown in Equation (1).

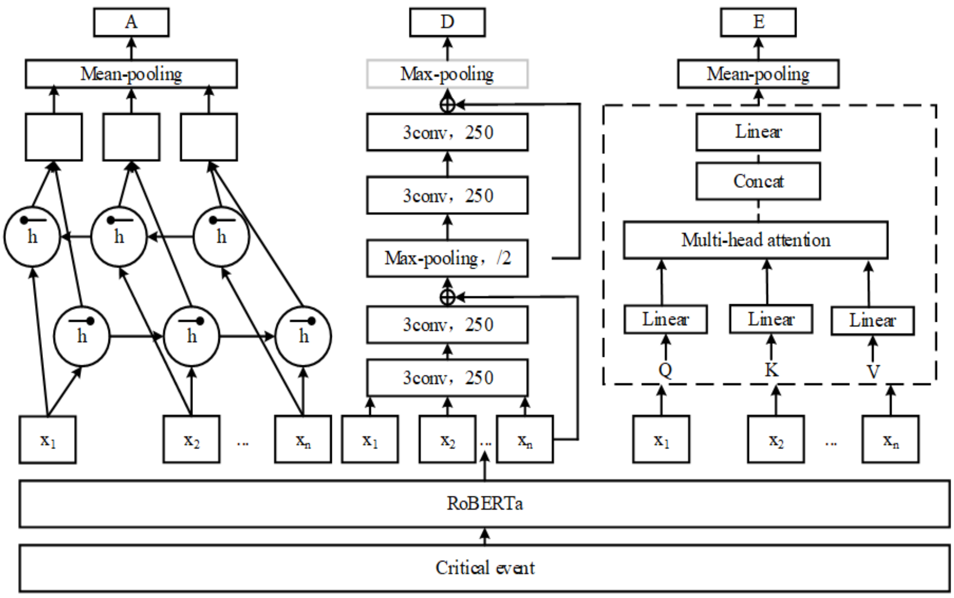


Figure 1. Multi-dimensional feature extraction diagram.

$$V = \frac{1}{2}\left(\frac{1}{m}\sum_{1}^{m} V_i^k + \frac{1}{n}\sum_{1}^{n} V_j^k\right) \tag{1}$$

In Equation (1), m and n mean the total number of words contained in the course introduction and course name, and V_j^k and V_i^k respectively represent the word vectors of the kth word obtained by segmenting the course name and course introduction. Multi-dimensional features can also be used for course attribute data processing, as shown in Equation (2).

$$v_j = T_j * W \tag{2}$$

In Equation (2), the vector after embedding the attribute data of the j-th course is represented by v_j, and the j-th row in the matrix is represented by T_j. In order to fuse the extracted features and achieve information supplementation between features, an encoder is used for calculation, as shown in Equation (3).

$$y^{(i)} = \sigma\left(W_i y^{(i-1)} + b_i\right) \tag{3}$$

In Equation (3), $y^{(i)}$ represents the output vector of the encoder after passing through the i-th layer, $y^{(i-1)}$ represents the hidden layer mapping at the input, and W_i and b_i are the weight matrix and bias vector, respectively. Therefore, MDFF can not only effectively alleviate information overload on OE platforms, but also improve students' learning efficiency. To further improve the validity of MDFF in course recommendation, deep learning technology is used to personalize the course RS. The DQN algorithm, which combines deep learning and reinforcement learning, can capture user interest shifts in a timely manner and adjust personalized recommendation strategies in real-time. Its framework is shown in Figure 2.

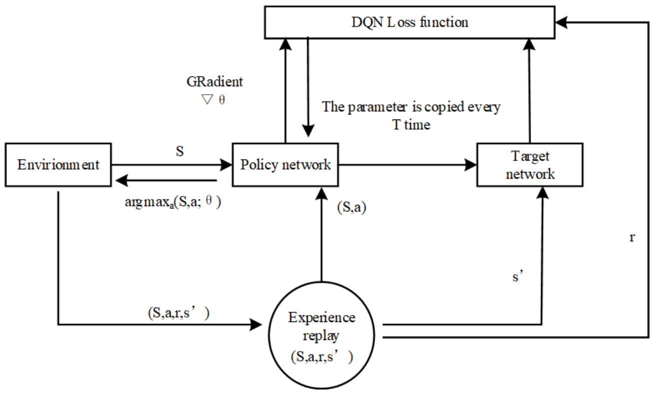


Figure 2. DQN frame diagram.

From Figure 2, the DQN algorithm adopts a multi-layer perceptron as the decision network structure, and the DQN loss function is performed through the policy network. The parameters are copied through the target network every T time, and then continuously replayed through experience. This framework is used to optimize the training cost. The square difference loss function of the multi-layer perceptron is represented in Equation (4).

$$L(\theta) = \frac{1}{2n}\sum_{i=1}^{n}|\hat{y}_i - y_i|^2 \quad (4)$$

In Equation (4), θ is the set parameter, the number of samples is represented by n, $\hat{y}_i$ means the predicted value of the samples, and i is the sample. Equation (4) can be used for parameter gradient optimization and improve data prediction and extraction capabilities. When capturing course interests, research also utilizes attention mechanisms in deep learning techniques to effectively process sequence data to simulate human attention focus. The expression for its attention distribution is shown in Equation (5).

$$\delta_i = \frac{exp(S_i)}{\sum_{j=1}^{n} exp(S_i)} \quad (5)$$

In Equation (5), the attention distribution is represented by δ_i and normalized to obtain. S_i is the similarity score. The weighted average can be calculated from Equation (5), as shown in Equation (6).

$$y = \sum_{i}^{n} S_i x_i \quad (6)$$

In Equation (6), y represents the weighted average of the input vector. In summary, attention mechanisms can be applied to complex models and capture global sequence information. At present, personalized courses for educational resources cannot be accurately recommended, and users are complex and difficult to handle. Therefore, research combines the advantages of MDFF and key semantic feature extraction, as well as deep learning technology to timely capture the complex correlation between users and courses, to construct a recommendation model for personalized courses for educational resources. The research proposes a personalized course recommendation model grounded on MDFF and deep learning, as shown in Figure 3.

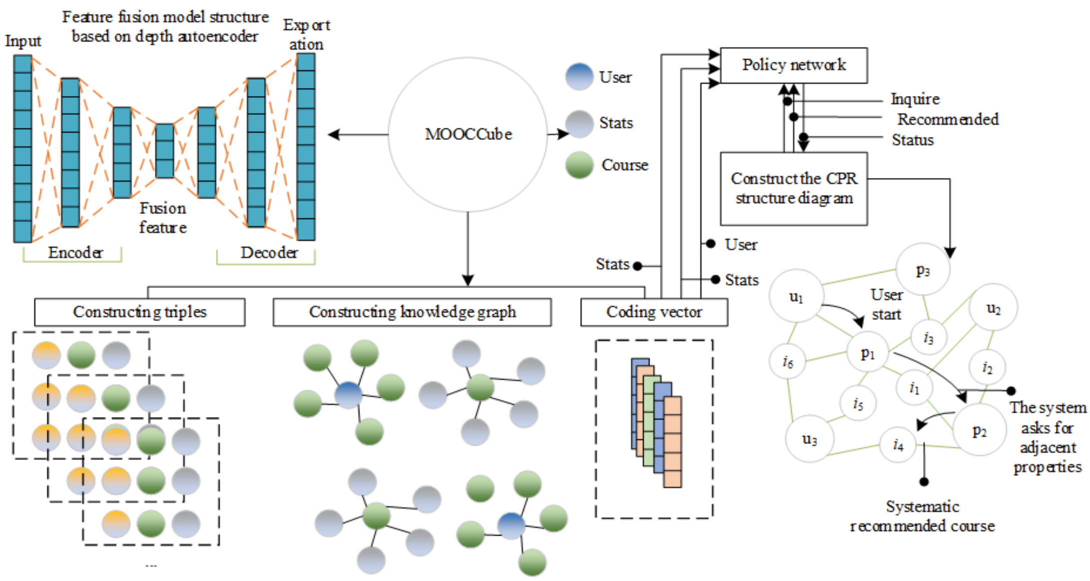


Figure 3. Personalized course recommendation model.

As shown in Figure 3, the course text data is extracted for semantic features through bidirectional LSTM, while the course attribute data is transformed into vector representations through embedding layers. Both are dynamically weighted through gate units during the encoder stage. The multi head attention mechanism not only analyzes the complex correlations within the course, but also receives real-time learning status feedback from the DQN strategy network, adjusting the importance distribution of different feature dimensions. The logical relationship between courses in knowledge graph construction, such as the introduction of prerequisite course dependencies as hard constraints into the action space of DQN, ensures that the generated recommendation sequence conforms to teaching rules and forms a closed-loop optimization from feature extraction to recommendation strategy. This design enables the system to capture users' fine-grained preferences while ensuring the teaching rationality of recommended courses.

3.2. Construction of an Educational Resource RS Based on Mahout CF Hybrid Algorithm

The educational resource system includes various resources, such as English audio-visual resources, language book resources, and computer information resources. Therefore, educational resources are redundant, and the educational resource system is prone to problems such as inaccurate user preference recommendations and insufficient correlation between users and course resources. A personalized course recommendation model grounded on MDFF and deep learning is suggested to address the above issues. However, this model still has certain shortcomings in recommending professional courses for educational platforms. Therefore, based on MDFF and personalized course recommendation through deep learning, a Mahout CF hybrid algorithm-based educational resource RS is constructed. As an open-source project, Mahout's taste component can store user preferences, as shown in Figure 4.

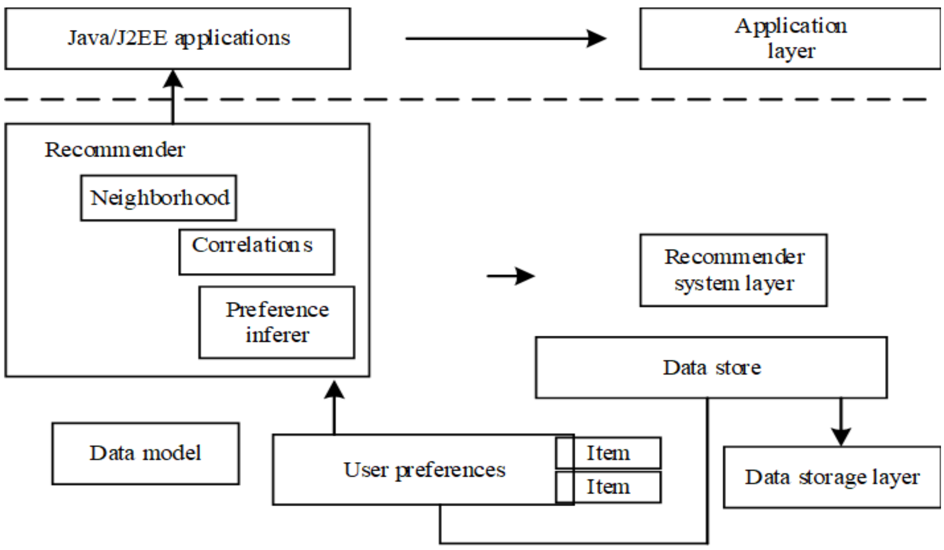


Figure 4. Taste component structure diagram.

As shown in Figure 4, the Taste component mainly consists of the data access model interface, calculating item similarity and user similarity, determining strategies for neighboring users, and using Recommender as the recommendation engine interface. The data access model interface can read and access data, and the recommender, as the most important content in the Taste component, can achieve CF of items and users for recommendations. Due to the abundance of educational learning resources and the tendency for information overload on educational platforms, research is being conducted on using the CF algorithm in Mahout to improve personalized recommendations for educational resources. CF algorithm, as a recommendation algorithm, can handle large-scale data and provide personalized services [18, 21]. The CF algorithm uses similar user data for similarity calculation, as represented in Equation (7).

$$sim(u,v) = \frac{|N(u) \cap N(v)|}{|N(u) \cap N(v)|} \tag{7}$$

In Equation (7), $sim(u, v)$ represents similarity, and $N(u)$ and $N(v)$ mean the collection of items that users u and v like. The Jaccard formula is used. $N(u) \cap N(v)$ means the total number of items that the user likes. The similarity is calculated using Equation (7), and then the rating of the item is predicted, as shown in Equation (8).

$$P_{u,i} = \bar{r}_u + \frac{\sum_{u \in N} sim(u,v) \times (r_{u,i} - \bar{r}_u)}{\sum_{u \in N} |sim(u,v)|} \tag{8}$$

In Equation (8), $\bar{r}_u$ is the average evaluation score of the user, and N and v mean the nearest neighbor set and the user set, respectively. The CF algorithm can also utilize the similarity between items for recommendation, and its users' predicted ratings for items are shown in Equation (9).

$$P_{u,j} = \sum_{i \in N(u) \cap S(j,K)} sim(j,i) \cdot r_{u,i} \tag{9}$$

In Equation (9), the user is represented by u. The item is represented by i and j. The degree of behavior of u towards j and i is $P_{(u,j)}$ and $r_{(u,i)}$, respectively. $S(j, K)$ is the closest set of Kth items. The User-based CF (UCF) and Item-based CF (ICF) are shown in Figure 5.

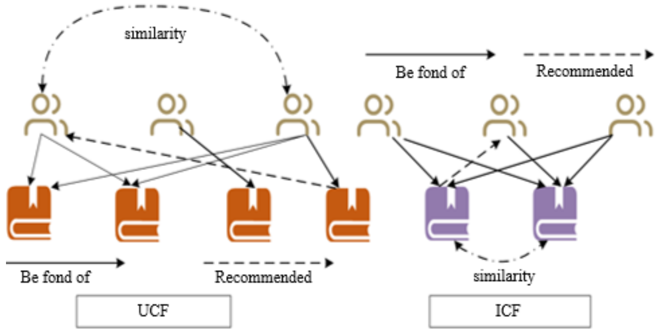


Figure 5. UCF and ICF.

From Figure 5, UCF relies on user historical behavior. If user similarity is calculated using rating data, the higher the similarity, the higher the likelihood of liking similar items. At the same time, the more users there are, the more diverse the results of recommending users will be, and they will be applicable to multiple fields. The biggest difference between ICF and UCF lies in distinguishing the similarity between items and recommending them to users [4, 10]. In order to better solve the problem of inaccurate learning goal recommendations for educational audiovisual resources, personalized feature matching is also performed on educational audiovisual resource users based on the Mahout CF algorithm. The distance formula between the determined samples is shown in Equation (10).

$$d(x, L) = \sqrt{\sum_{j=1}^{m} (x_j - L_j)^2} \quad (10)$$

In Equation (10), x_j is the sample object, and the j-th cluster center is represented by L_j. The sample dimension is shown by m. The probability model learning method can be used to mine personalized user preference features of educational audiovisual resources, as shown in Equation (11).

$$Q_j = \eta_1 \cdot \frac{Dy_j}{w(a_d)} + \eta_2 \frac{A_k p_j}{p_{int}} \quad (11)$$

In Equation (11), η_1 and η_2 represent the observation window width values for different features, respectively. The frequent feature term is represented by Dy_j. p_{int} is a personalized set of untrusted attributes. In order to gain user trust and provide personalized recommendations for educational resources, immune evolution is used to update the weights and provide optimal recommendations for audiovisual resources, as shown in Equation (12).

$$\begin{cases} V(R_j) = wR_j + C(pBest - X) \\ E = \sigma_k + Nx_{id} \end{cases} \quad (12)$$

In Equation (12), $V(R_j)$ represents weight update, and E represents the optimal recommendation result obtained. C and X represent the particle update speed and the limit speed range, respectively. The group size is represented by x_{id}. In summary, Mahout CF algorithm has strong interpretability and can perform semi-structured and unstructured processing on complex data. Therefore, in response to the problems of insufficient educational resource search engines and inaccurate personalized recommendations, based on MDFF and deep learning personalized course recommendations, the study constructs an educational resource RS grounded on Mahout CF hybrid algorithm to provide users with personalized intelligent services. The general structure of the educational resource RS based on Mahout CF hybrid algorithm is shown in Figure 6.

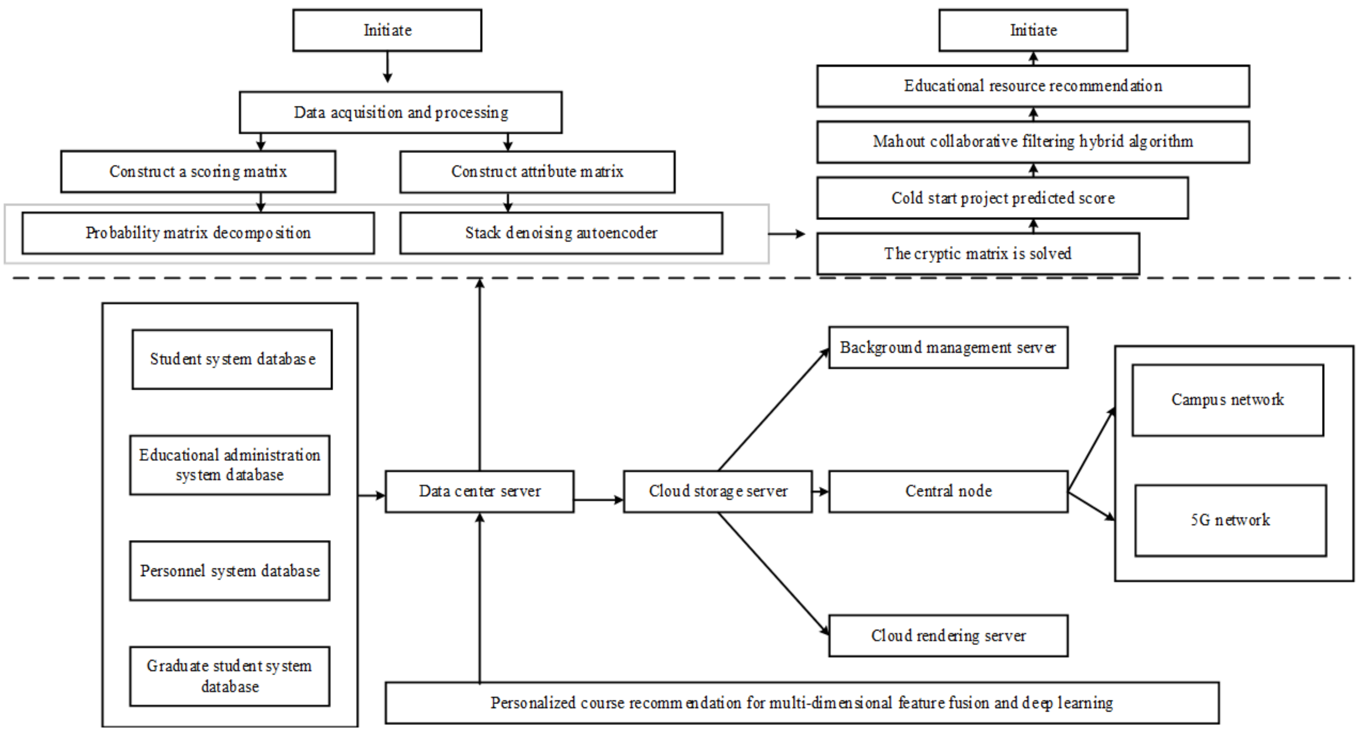


Figure 6. Overall structure of educational resource RS.

As shown in Figure 6, its innovative integration is mainly reflected in the deep coupling of data, algorithms, and services. At the data level, the raw data from various systems on campus is standardized and used for both Mahout's traditional similarity calculation and providing rich semantic context for DQN. At the algorithmic level, a bidirectional interaction mechanism has been constructed: DQN dynamically adjusts the similarity calculation threshold of Mahout by analyzing the real-time learning status of users, while the set of neighboring users generated by Mahout provides reliable exploration space constraints for DQN's policy network. The service layer intelligently sorts based on contextual factors such as terminal devices and network environment in the final recommendation, achieving differentiated display strategies between mobile and Personal Computer (PC) devices. The system has specially designed the curriculum prior relationship verification module and the learning behavior buried point collection channel, so that the entire recommendation process not only maintains the progressiveness technology, but also conforms to the basic laws of education and teaching.

4. Performance Verification of Educational Resource System Based on Mahout CF Hybrid Algorithm

4.1. Performance Comparison and Analysis of Mahout-Based CF Hybrid Recommendation Algorithm

To highlight the superior performance of Mahout CF hybrid recommendation algorithm, this study compared it with CF algorithm, Fuzzy C-Means (FCM) algorithm, and Knowledge graph completion is combined with RS (RippleNet) algorithm. This experiment used MovieLens-100k and Chinese Educational Corpus (CEC) Chinese experimental set as test datasets, with Windows 10 as the operating system, Python 3.7 as the experimental platform, and Intel (R) Xeon (R) Platinum 8255C CPU@2.50GHz+24G memory as the hardware configuration. The MovieLens-100K dataset is used as the benchmark dataset to verify the generalization of the hybrid algorithm, which includes 943 users' ratings of 100000 for 1682 movies, with a sparsity of 93.7%. The CEC is used for testing in practical scenarios, containing 50000 students and 3500 courses with a sparsity of 88.2%. Its features include course ratings, clickstream data, course selection records, and metadata such as subjects and difficulty levels. In addition, systematic optimization experiments were conducted on DQN parameter settings, and the key hyperparameter configurations were ultimately determined as follows. The learning rate adopted a segmented decay strategy, with an initial value of 0.001 and a decay rate of 0.8 times the original every 5000 steps; The discount factor was determined to be 0.95 through grid search, achieving a balance between long-term returns and immediate rewards; The experience replay buffer size is set to 20000 records to ensure training stability while avoiding interference from outdated data; In the ϵ-growth strategy, the initial value is 0.9 and linearly decays to 0.1. The trade-off curve between exploration and utilization is validated through A/B testing. For the MDFF feature extraction process, the Chinese course description text is preprocessed using Jieba word segmentation tool. After removing stop words, a pre trained BERT wwm model is used to obtain a 384 dimensional dynamic word vector; numerical attribute data is standardized by MinMax and input into a three-layer fully connected network (256-128-64 nodes), with each hidden layer followed by a LayerNorm layer and a 0.3 ratio dropout; user preference features are processed through a bidirectional Gated Recurrent Unit (GRU) network (hidden layer dimension 128) to process behavior sequences, and finally interact with course features in a 768 dimensional joint embedding space for attention. The entire feature engineering pipeline is implemented using PyTorch, with a specially designed heterogeneous data adaptation layer. Text, numerical, and sequential data are processed separately through independent preprocessing branches and then gate feature selection is performed in the fusion layer. In the specific implementation, two layers of cross attention mechanisms are set up to capture cross modal correlations. The supplementation of these technical details makes the method part fully reproducible, while highlighting the unique processing strategies for educational data. The study compared the loss values of four recommendation algorithms, Mahout CF hybrid, CF, FCM, and RippleNet, for different recommendation K-values. The outcomes are in Figure 7.

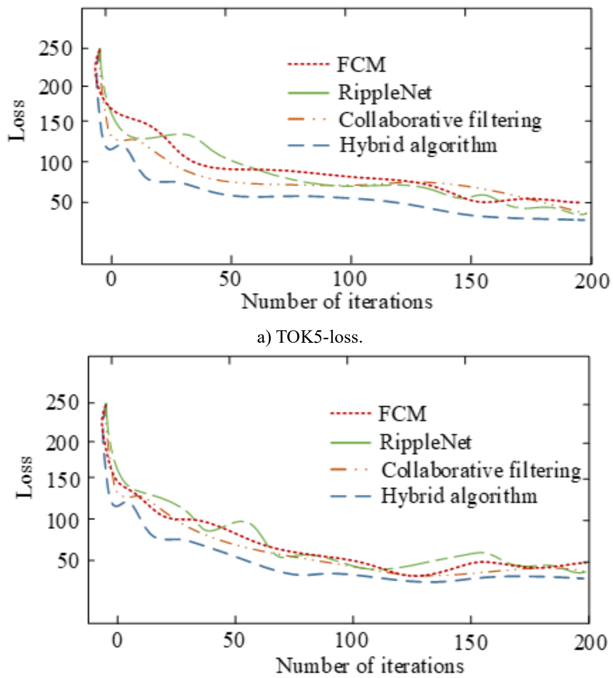
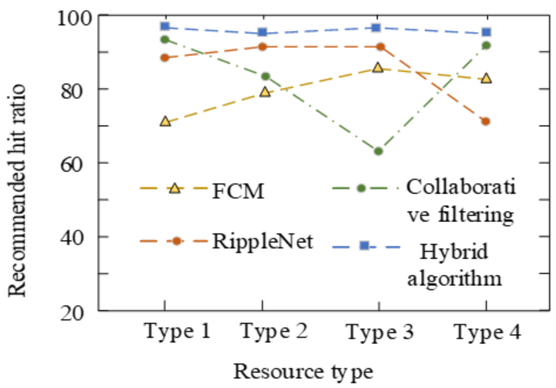
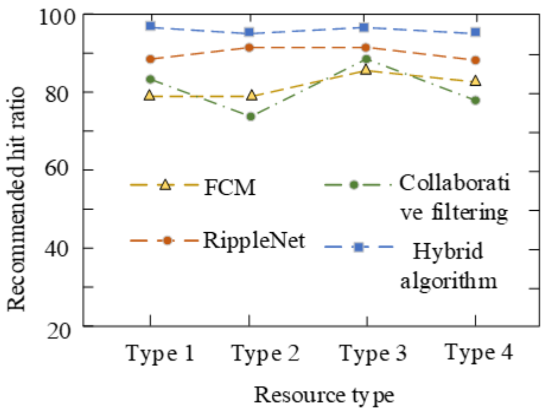


Figure 7. Loss test for different recommended K-values.

From Figure 7-a), in the TOK5 Loss comparison test, the Mahout CF hybrid recommendation algorithm had a loss value below 50 after 150 iterations, and the loss value tended to balance after 200 iterations without significant changes. From Figure 7-b), the overall iterative training loss value of Mahout CF hybrid recommendation algorithm was lower than that of CF, FCM, and RippleNet recommendation algorithms. In summary, the hybrid recommendation algorithm based on Mahout CF performed well and could determine the educational resources that users are truly interested in when evaluating search engines. In order to better demonstrate the accuracy of the Mahout CF hybrid recommendation algorithm in resource recommendation, recommendation hit tests were also conducted on four recommendation algorithms: Mahout CF hybrid, CF, FCM, and RippleNet. The test results are shown in Figure 8.



a) MovieLens-100k.

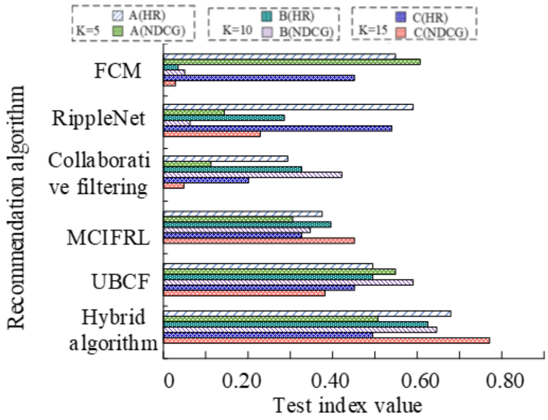


b) CEC Chinese experimental set.

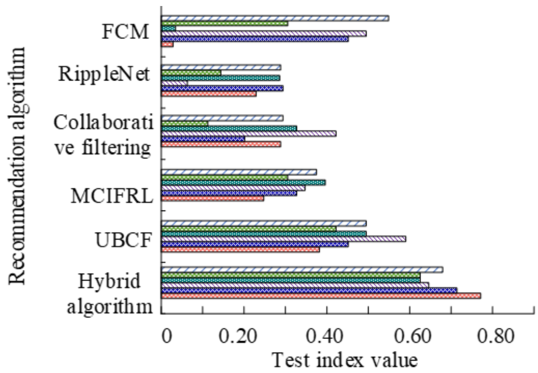
Figure 8. Recommended hit test.

According to Figure 8-a), in the MovieLens-100k test set, the Mahout CF hybrid recommendation algorithm had a hit accuracy of 98.23%. According to Figure 8-b), the average hit accuracy of the RippleNet recommendation algorithm was 91.56%, second only to the Mahout CF hybrid recommendation algorithm. In summary, the hybrid recommendation algorithm based on Mahout CF performed the best in personalized hit testing, with an overall hit rate of 99.48%, improving the reliability of resource personalized recommendations. The study not only tested the recommendation accuracy of the four algorithms mentioned above, but also compared HR and NDCG indicators. In order to make the test results more diverse, Multi-Criteria Item Filtering Recommendation Learning (MCIFRL) and User-Based Collaborative Filtering (UBCF) recommendation algorithms were added on the basis of the four algorithms, and the test results are shown in Figure 9.

According to Figure 9-a), when the recommendation list length k was 5, the HR value of Mahout CF hybrid recommendation algorithm was 0.68. As shown in Figure 9-b), the overall average NDCG index of Mahout CF hybrid recommendation algorithm was 0.75, which improved the overall user experience. In summary, the hybrid recommendation algorithm based on Mahout CF had certain stability and to some extent solved the problem of massive data.



a) MovieLens-100k.



b) CEC Chinese experimental set.

Figure 9. HR and NDCG indicator testing.

4.2. Performance Analysis of Educational Resource RS Based on Mahout CF Hybrid Algorithm

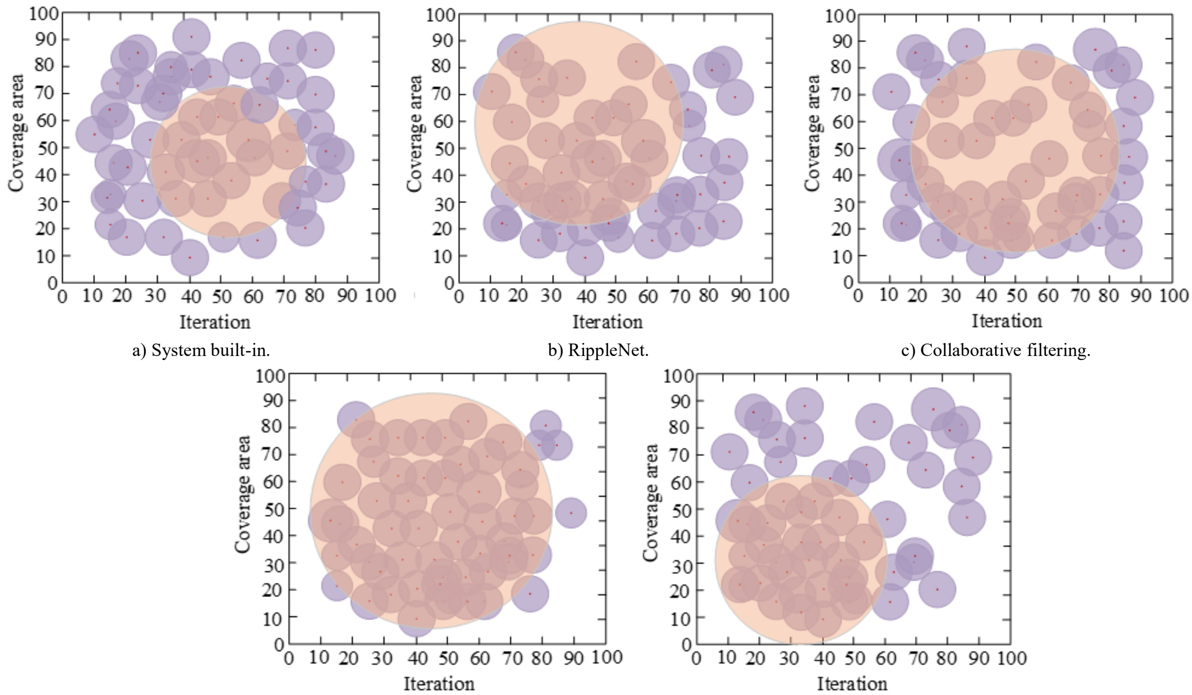
To further confirm the capability of the educational resource RS constructed based on the Mahout CF hybrid algorithm, the educational resource RSs constructed using CF, FCM, and RippleNet recommendation algorithms were also compared and tested for corresponding indicators on a simulation platform. Due to the large amount of educational resource data, Python was used to clean and statistically analyze the data before conducting experiments. In this experiment, 8 different users and 8 different courses were selected.

The similarity matrix was used as the target to find suitable courses and make recommendations. Finally, a recommendation list was generated, and the outcomes are in Table 1.

Table 1. Recommendation list result.

Subscriber number	Course number	Recommended values of different RSs			
		RippleNet	FCM	Hybrid algorithm	CF
42336	11	0.658	0.772	0.812	0.846
42246	12	0.665	0.713	0.915	0.786
41156	23	0.632	0.746	0.923	0.816
41856	48	0.612	0.741	0.947	0.810
23378	21	0.625	0.746	0.846	0.813
25569	33	0.654	0.645	0.856	0.798
27454	89	0.756	0.642	0.965	0.741
29786	45	0.641	0.690	0.847	0.723

According to Table 1, the FCM education resource RS had an overall average course recommendation value of 0.711 for 8 different users, the Mahout CF hybrid education resource RS had an overall average recommendation value of 0.888, and the CF education resource RS had an overall average recommendation value of 0.791. In summary, the educational resource RS constructed based on Mahout CF hybrid algorithm had high recommendation values and was in line with user preferences. In order to observe the coverage effect of Mahout CF hybrid educational resource RS on recommended items more intuitively, four educational resource RSs including Mahout CF hybrid, CF, FCM, and RippleNet were tested for item recommendation coverage. The test results are shown in Figure 10.



a) System built-in. b) RippleNet. c) Collaborative filtering. d) Hybrid algorithm. e) FCM.

Figure 10. Item recommendation coverage test.

According to Figure 10-a), after 100 iterations, the educational resource RS's built-in item recommendations had a smaller coverage area and were concentrated in the central region. According to Figure 10-b), the RippleNet educational resource RS covered a large number of items in the recommendation list, with a coverage rate of 45.23% compared to the total number of items. As shown in Figure 10-c), the overall coverage of the CF education resource RS was concentrated in the central area, but the recommended list items within the coverage area were relatively scattered. According to Figure 10-d), the Mahout CF hybrid educational resource RS had a concentrated coverage area and a recommended list item coverage ratio of 95.48%. As shown in Figure 10-e), the FCM education resource recommendation coverage rate was 44.15%, and most of the items were scattered in various areas. In summary, the educational resource RS based on Mahout CF hybrid algorithm had a high coverage rate, and the recommended results covered a wide range of items, achieving good personalized recommendations for users. In order to further demonstrate the performance of the educational resource RS based on Mahout CF hybrid algorithm in search, click through rate prediction experiments were conducted on four educational resource RSs: Mahout CF hybrid, CF, FCM, and RippleNet. On this basis, MCIFRL and UBCF educational RSs were also added for Area Under the Curve (AUC) and F1-value comparison. The experiment outcomes are in Table 2.

Table 2. Click-through rate experiment.

Educational resource RS	AUC	F1	p(Compared to hybrid algorithm)
RippleNet	0.7004	0.7465	<0.001
FCM	0.7246	0.6849	<0.001
Hybrid algorithm	0.8984	0.8298	/
CF	0.7846	0.7146	0.003
MCIFRL	0.6768	0.7148	<0.001
UBCF	0.7746	0.6823	0.005

According to Table 2, the Mahout CF hybrid educational resource RS achieved significant improvements in AUC and F1-values, with indices of 0.8984 and 0.8298, respectively. The AUC index of MCIFRL educational resource RS was 0.6768, which was the lowest in the entire AUC test. The F1-value of FCM educational resource RS was 0.6849, which was the lowest in the entire F1 test. In addition, the study used a two-sample t-test to compare the differences between various algorithms and hybrid algorithm. The results showed that the p-values of all compared algorithms were less than 0.05, indicating that the advantages of the hybrid algorithm were not caused by random errors, further supporting the reliability of the hybrid algorithm in educational resource recommendation scenarios. In summary, the Mahout CF hybrid educational resource RS could effectively measure the ranking ability of educational resources, without being affected by other categories, and had a good click through rate effect.

To explore the contribution of each module to the research methodology, ablation experiments were conducted. Among them, the hyperparameters of the baseline method are set as follows: the number of neighbors is 50 and the similarity threshold is 0.65. Maintain consistency between other hyperparameter settings and performance testing settings. In order to ensure the fairness and consistency of the experimental results, all methods used the same data partitioning method and evaluation indicators. Specifically, the study used a five-fold cross validation method, with the training set, validation set, and test set being 70%, 15%, and 15%, respectively. In the evaluation process, the study uses accuracy NDCG@10, the HR@5, the performance of the model is comprehensively evaluated based on the coherence of the learning path, and the same evaluation process is strictly followed in each experiment. The results are shown in Table 3.

Table 3. Results of ablation experiment.

Model	HR@5	NDCG@10	Average accuracy/%	Learning path coherence
Complete hybrid system	0.78	0.82	88.9	0.71
Without DQN	0.7	0.74	76.6	0.58
Without MDFF	0.69	0.73	80.2	0.49
Without Mahout CF	0.65	0.68	73.7	0.63
Static weight fusion	0.72	0.75	83.4	0.65
Mahout CF only (baseline)	0.68	0.71	79.1	0.32
DQN only (baseline)	0.63	0.69	71.1	0.45

According to Table 3, removing the DQN module resulted in a sudden drop of 12.3% in cold start course coverage and a decrease of 0.13% in learning path coherence, highlighting its dynamic optimization ability based on a three-dimensional reward function. By analyzing knowledge coherence, difficulty matching, and learning pace, it significantly improved the adaptability of educational recommendations. The lack of MDFF feature fusion module reduces the hit rate of text based course recommendations by 8.7%, verifying its key role in cross modal semantic understanding and effectively solving the problem of heterogeneous data fusion. The baseline system relying solely on traditional Mahout CF maintains basic recommendation capabilities, but its 23.5% cold start coverage and 0.32 path coherence score fully expose the limitations of static CF in educational settings. It is particularly noteworthy that the introduction of dynamic weighting mechanism brings about a 7% improvement in NDCG, and its designed formula for balancing course popularity and user activity has become the core hub of the coordination algorithm module. The innovation of this hierarchical fusion architecture is reflected in the fact that the candidate set generated by Mahout CF is verified for teaching logic through DQN real-time policy network, while the deep features provided by MDFF support fine-grained content matching, ultimately forming a hybrid system that combines CF interpretability, reinforcement learning dynamism, and deep learning semantic understanding ability. The complete system has improved recommendation accuracy by 15.7% compared to the optimal baseline, due to the significant impact generated by the collaborative efforts of various modules.

Based on the above content, it can be concluded that although the Mahout CF hybrid algorithm itself has good scalability, actual education platforms need to meet the real-time recommendation needs of millions of users and course resources. The research method can optimize the computational complexity to near linear complexity through Mahout’s distributed computing framework; The DQN module requires iterative training, which increases the time required for a single recommendation by about 15%, but can be alleviated through offline training and online inference separation. The user course interaction matrix adopts sparse storage format, reducing memory usage by 60%. After dimensionality reduction through Principal Component Analysis (PCA), MDFF reduces feature storage to 30% of the original data. The data preprocessing and similarity calculation stage relies on Hadoop MapReduce to achieve horizontal scaling, and it has been tested that processing 1TB of data on a 20 node cluster takes less than 2 hours.

5. Conclusions

Addressing the issue of inaccurate personalized recommendation of educational resources in the education system, this study constructed an educational resource RS based on Mahout CF hybrid algorithm, which can better recommend educational resource information. The system also adopted MDFF and deep learning for personalized course recommendation during construction, which addresses issues such as poor user preferences and enhances intelligent services. In order to demonstrate the superior performance of the educational resource RS constructed by Mahout CF hybrid algorithm, it was compared with three recommendation algorithms: CF, FCM, and RippleNet. The experimental results demonstrated that the Mahout CF hybrid recommendation algorithm tended to a stable state after 200 iterations, with a loss value below 50, while the CF, FCM, and RippleNet recommendation algorithms had lower loss values than the hybrid recommendation algorithm. In addition, empirical analysis of the constructed educational resource RS found that the average recommendation value of the educational resource RS based on Mahout CF hybrid algorithm was 0.888. The average recommendation values of the educational resource recommendation models constructed using CF, FCM, and RippleNet recommendation algorithms were 0.791, 0.711, and 0.655, respectively, which were lower than those of the educational resource RS proposed by the research. Overall, the educational resource RS based on Mahout CF hybrid algorithm improved the personalized recommendation ability of educational audio-visual resources and provided greater allocation value to educational resources. However, there are still certain limitations to the research, as it has not yet conducted specialized tests on cold start issues such as new course launches, new user registrations, and extreme data sparsity scenarios. Therefore, in future research, a test set for simulating cold start can be added, such as blocking historical data of some users/courses and quantifying the performance degradation of algorithms under sparse conditions.

References

[1] Bhaskaran S. and Marappan R., "Enhanced Personalized Recommendation System for Machine Learning Public Datasets: Generalized Modeling, Simulation, Significant Results and Analysis," *International Journal of Information Technology*, vol. 15, no. 3, pp. 1583-1595, 2023. https://doi.org/10.1007/s41870-023-01165-2

[2] Carpenter J. and Shelton C., "Educators' Perspectives on and Motivations for Using TeachersPayTeachers.com," *Journal of Research on Technology in Education*, vol. 56, no. 2, pp. 218-232, 2024. https://doi.org/10.1080/15391523.2022.2119452

[3] Chen W., Shen Z., Pan Y., Tan K., and Wang C., "Applying Machine Learning Algorithm to Optimize Personalized Education Recommendation System," *Journal of Theory and Practice of Engineering Science*, vol. 4, no. 1, pp. 101-108, 2024. https://doi.org/10.53469/jtpes.2024.04(01).14

[4] Chen Y., "Improving Student Engagement in College Education with Fuzzy Control Algorithm," *The International Arab Journal of Information Technology*, vol. 22, no. 4, pp. 651-664, 2025. https://doi.org/10.34028/iajit/22/4/2

[5] Cotton D., Cotton P., and Shipway J., "Chatting and Cheating: Ensuring Academic Integrity in The Era of ChatGPT," *Innovations in Education and Teaching International*, vol. 61, no. 2, pp. 228-239, 2024. https://doi.org/10.1080/14703297.2023.2190148

[6] Familoni B. and Onyebuchi N., "Augmented and Virtual Reality in US Education: A Review: Analyzing the Impact, Effectiveness, and Future Prospects of AR/VR Tools in Enhancing Learning Experiences," *International Journal of Applied Research in Social Sciences*, vol. 6, no. 4, pp. 642-663, 2024. https://doi.org/10.51594/ijarss.v6i4.1043

[7] Hasan M. and Ferdous J., "Dominance of AI and Machine Learning Techniques in Hybrid Movie Recommendation System Applying Text-to-Number Conversion and Cosine Similarity Approaches," *Journal of Computer Science and Technology Studies*, vol. 6, no. 1, pp. 94-102, 2024. https://doi.org/10.32996/jcsts.2024.6.1.10

[8] Jacob O., Evans O., and Ayeni E., "Vice Chancellors and University Education Sustainability in Nigeria: Challenges and Way Forward," *Journal of Higher Education and Academic Advancement*, vol. 1, no. 7, pp. 12-22, 2024. https://doi.org/10.61796/ejheaa.v1i1.284

[9] Li J., Shi J., Zhang J., Lu Y., and et al., "Quantum Nearest Neighbor Collaborative Filtering Algorithm for Recommendation System," *ACM Transactions on Knowledge Discovery from Data*, vol. 18, no. 8, pp. 1-28, 2024. https://doi.org/10.1145/3674982

[10] Li Y., Chen H., Xu S., Ge Y., and et al., "Fairness in Recommendation: Foundations, Methods, and Applications," *ACM Transactions on Intelligent Systems and Technology*, vol. 14, no. 5, pp. 1-48, 2023. https://doi.org/10.1145/3610302

[11] Mikroyannidis A. and Papastilianou A., "Open Educational Resources in Public Administration: A Case Study in Greece," *Open Learning: The Journal of Open, Distance and e-Learning*, vol. 39, no. 3, pp. 226-240, 2024. https://doi.org/10.1080/02680513.2021.1950526

[12] Oladokun B., Ajani Y., George N., and Okeke O., "From Metaverse to Meta AI: A Dynamic Disruption in Libraries in Higher Education Institutions," *Library Hi Tech News*, vol. 41, no. 9, pp. 12-14, 2024. https://doi.org/10.1108/LHTN-04-2024-0059

[13] Phulpoto S., Oad L., and Imran M., "Enhancing Teacher Performance in E-Learning: Addressing Barriers and Promoting Sustainable Education in Public Universities of Pakistan," *Pakistan Languages and Humanities Review*, vol. 8, no. 1,

pp. 418-429, 2024. https://doi.org/10.47205/plhr.2024(8-I)38

[14] Qin Y., Li M., and Zhu J., "Privacy-Preserving Federated Learning Framework in Multimedia Courses Recommendation," *Wireless Networks*, vol. 29, no. 4, pp. 1535-1544, 2023. https://doi.org/10.1007/s11276-021-02854-1

[15] Raj N. and Renumol V., "An Improved Adaptive Learning Path Recommendation Model Driven by Real-Time Learning Analytics," *Journal of Computers in Education*, vol. 11, no. 1, pp. 121-148, 2024. https://doi.org/10.1007/s40692-022-00250-y

[16] Sanabria J., Ponce B., Perez L., and Montoya M., "Reusable Educational Resources for Developing Complex Thinking on Open Platforms," *Education and Information Technologies*, vol. 29, no. 1, pp. 1173-1199, 2024. https://doi.org/10.1007/s10639-023-12316-0

[17] Shang F., Zhao F., Zhang M., Sun J., and Shi J., "Personalized Recommendation Systems Powered by Large Language Models: Integrating Semantic Understanding and User Preferences," *International Journal of Innovative Research in Engineering and Management*, vol. 11, no. 4, pp. 39-49, 2024. https://doi.org/10.55524/ijirem.2024.11.4.6

[18] Stray J., Halevy A., Assar P., Ashar A., and et al., "Building Human Values into Recommender Systems: An Interdisciplinary Synthesis," *ACM Transactions on Recommender Systems*, vol. 2, no. 3, pp. 1-57, 2024. https://doi.org/10.1145/3632297

[19] Ubogu R., "Supervision of Instruction: A Strategy for Strengthening Teacher Quality in Secondary School Education," *International Journal of Leadership in Education*, vol. 27, no. 1, pp. 99-116, 2024. https://doi.org/10.1080/13603124.2020.1829711

[20] Xiang Y., Huo S., Wu Y., and Zhu M., "Integrating AI for Enhanced Exploration of Video Recommendation Algorithm Via Improved Collaborative Filtering," *Journal of Theory and Practice of Engineering Science*, vol. 4, no. 2, pp. 83-90, 2024. https://doi.org/10.53469/jtpes.2024.04(02).12

[21] Zhang Y., Zhang Y., Yan D., Deng S., and Yang Y., "Revisiting Graph-Based Recommender Systems from the Perspective of Variational Auto-Encoder," *ACM Transactions on Information Systems*, vol. 41, no. 3, pp. 1-28, 2023. https://doi.org/10.1145/3573385

Guiyong Zhu obtained a Bachelor of Science degree from the Department of Mathematics, Nanjing University in 2004, and a Master of Engineering degree from the School of Computer Science, National University of Defense Technology in 2006. From 2007 to 2017, he worked as System Architect of ArcherMind Technology Co., Ltd., being responsible for system architecture design, development and project management of various information system projects. Since 2018, he has been working as a lecturer at Xingzhi College, Zhejiang Normal University. His research interests include the Design and Development of Various Intelligent Information Systems, Natural Language Processing, and the Application of Large Language Models. Mr. Zhu is a professional member of the China Computer Federation (CCF).