

Optimization research on collaborative filtering-based personalized educational resource recommendation algorithm

Wang, Hai ✉

Nanning University, China (whai2079@gmail.com)



ISSN: 2243-7770
Online ISSN: 2243-7789

Received: 13 March 2025

Revised: 20 April 2025

Accepted: 23 April 2025

OPEN ACCESS

Available Online: 25 April 2025

DOI: 10.5861/ijrsm.2025.25044

Abstract

With the rapid development of educational informatization, various online education platforms have accumulated massive educational resources. However, this has also led to the problem of educational resource overload, making it difficult for learners to quickly find content that meets their needs and interests. To address this issue, this study focuses on the application and optimization of collaborative filtering algorithms in personalized educational resource recommendation. Aiming at the challenges faced by collaborative filtering algorithms such as data sparsity, cold start problems, and insufficient interpretability of recommendation results, this study proposes an optimized model that integrates multi-source data with graph neural networks. By introducing multi-source data such as student interest tags, course metadata, and learning behavior records, and combining a time decay factor to dynamically adjust user interest weights, the model effectively alleviates the problem of data sparsity and improves the accuracy of recommendations. Meanwhile, the use of course description text to generate embedding vectors solves the cold start problem for new users or new resources. Additionally, by constructing a course knowledge graph, the interpretability of recommendation results is significantly enhanced. Experimental results show that the optimized algorithm achieves significant improvements in evaluation metrics such as accuracy, recall, coverage, and NDCG, providing learners with more accurate and interpretable personalized educational resource recommendation services.

Keywords: collaborative filtering algorithm, personalized educational resource recommendation, data sparsity, multi-source data fusion, knowledge graph

Optimization research on collaborative filtering-based personalized educational resource recommendation algorithm

1. Introduction

1.1 Research Background and Significance

With the rapid development of educational informatization, various online education platforms have emerged, providing learners with a rich variety of educational resources. However, the massive growth of such resources has also brought new problems—educational resource overload. Faced with a dazzling array of courses and learning materials, learners often find it difficult to quickly find content that meets their learning needs and interests, which not only reduces learning efficiency but may also trigger learning anxiety. Therefore, how to effectively screen out suitable content for learners from massive educational resources has become an urgent problem to be solved in the field of educational informatization.

Personalized recommendation technology provides an effective way to solve this problem. By analyzing learners' historical learning behaviors, interest preferences, and learning needs, personalized recommendation algorithms can accurately push educational resources that meet their individualized needs to learners, thereby improving learning efficiency and learning experience. Among numerous recommendation algorithms, collaborative filtering algorithms have become the core algorithm in personalized recommendation of educational resources due to their good recommendation effects and wide application scenarios.

Collaborative filtering algorithms mainly recommend based on the similarity of user behavior or item content, which has the advantages of simple implementation and stable recommendation effects. However, in practical applications, collaborative filtering algorithms also face some challenges, such as data sparsity, cold start problems, and the interpretability of recommendation results. These problems are particularly prominent in educational resource recommendation scenarios because learners' learning behaviors are relatively sparse, and the addition of new users or new resources often leads to unstable recommendation effects. Therefore, optimizing collaborative filtering algorithms to improve the accuracy and interpretability of educational resource recommendations has important research significance and application value.

1.2 Research Objectives

This study aims to optimize collaborative filtering algorithms to solve the problems of data sparsity, cold start, and improve the interpretability of recommendation results in personalized educational resource recommendation. The specific research objectives include: (1) By integrating multi-source data (such as student interest tags, course metadata, learning behavior records, etc.), enriching user-item interaction information, alleviating data sparsity problems, and improving the accuracy of recommendations. (2) Utilizing pre-trained models to generate embedding vectors from course description texts and other sources to solve the cold start problem for new users or new resources, improving the coverage and practicality of the recommendation system. (3) Enhancing the interpretability of recommendation results by constructing knowledge graphs and other strategies, enabling learners to understand recommendation reasons more intuitively and improving the learning experience. In summary, this study aims to provide learners with more accurate and interpretable personalized educational resource recommendation services by optimizing collaborative filtering algorithms, promoting the development of educational informatization and improving learning efficiency.

2. Review of Related Work

2.1 Principles of Collaborative Filtering Algorithms

As a classic algorithm in the field of personalized recommendation, collaborative filtering algorithms mainly include user-based collaborative filtering (User-based CF) and item-based collaborative filtering (Item-based CF). User-based CF recommends resources liked by similar users to the target user by calculating the similarity between users. Item-based CF recommends resources similar to those previously liked by the target user by analyzing the similarity between items. In addition, with the development of technology, matrix factorization and embedding representation methods are also widely used in collaborative filtering algorithms, further improving the accuracy of recommendations by utilizing latent vectors to model user-item interactions.

2.2 Optimization Directions of Algorithms

To address the problems faced by collaborative filtering algorithms in practical applications, such as data sparsity, cold start problems, and the interpretability of recommendation results, researchers have proposed multiple optimization directions and methods.

(1) Data Sparsity: Due to the sparsity of user-item interaction data, it is often difficult for recommendation algorithms to accurately capture user interests and item features. To solve this problem, researchers have introduced methods such as social networks, item metadata, and matrix completion techniques to enrich the user-item interaction matrix by integrating multi-source information and improve the accuracy of recommendations.

(2) Cold Start Problem: For new users or new resources, due to the lack of sufficient historical interaction data, traditional collaborative filtering algorithms are often difficult to provide effective recommendations. To solve this problem, researchers have proposed methods such as hybrid content recommendation, pre-trained models, and domain knowledge to generate initial recommendation results by utilizing descriptive information or domain knowledge of new users or new resources, improving the coverage and practicality of the recommendation system.

(3) Interpretability: The recommendation results generated by collaborative filtering algorithms often lack clear interpretability, making it difficult for users to understand the recommendation reasons, which affects the recommendation effect and user satisfaction. To solve this problem, researchers have proposed methods such as feature learning, metadata description, and dynamic interpretation to provide clear explanations and bases for recommendation results by constructing knowledge graphs or extracting item features, enhancing users' trust and satisfaction.

(4) Graph Neural Network Optimization: In recent years, with the development of graph neural network technology, researchers have begun to apply it to collaborative filtering algorithms to further improve the accuracy and interpretability of recommendations by mining high-order relationships in user-item interaction relationships. For example, graph models such as NGCF and LightGCN have achieved remarkable results in recommendation systems.

In summary, collaborative filtering algorithms have broad application prospects and research value in the field of personalized recommendation. However, in practical applications, it is still necessary to optimize and improve them for specific problems to enhance the accuracy, interpretability, and practicality of recommendations.

3. Algorithm Optimization Methods and Model Design

This study proposes the following optimization methods and model designs for the challenges faced by collaborative filtering algorithms in personalized educational resource recommendation.

3.1 Hybrid Recommendation Model Framework

To alleviate the problem of data sparsity, this study designs a hybrid recommendation model framework that integrates multi-source data to enrich user-item interaction information. Specifically, we combine multi-dimensional data such as student interest tags, course metadata, and learning behavior records, which can more comprehensively reflect users' learning needs and interest preferences. Through a comprehensive analysis of these data, we can more accurately capture user interests and item features, thereby improving the accuracy of recommendations.

In terms of similarity calculation, traditional collaborative filtering algorithms often only consider the direct similarity between users or items. However, in practical applications, user interests may change over time. Therefore, this study introduces a time decay factor to dynamically adjust user interest weights. In this way, we can more accurately reflect users' current interest preferences and further improve the accuracy and timeliness of recommendations.

To address the cold start problem for new users or new resources, this study proposes a method of generating embedding vectors using course description texts. Specifically, we utilize pre-trained models to encode course description texts and generate course embedding vectors. Then, by calculating the similarity between new users or new resources and existing resources, we recommend similar resources to them. This method can effectively utilize descriptive information to solve the cold start problem for new users or new resources and improve the coverage and practicality of the recommendation system.

3.2 Interpretability Enhancement Strategies

To improve the interpretability of recommendation results, this study constructs a knowledge graph to associate recommendation reasons. Specifically, we analyze the relationships between courses and construct a course knowledge graph. Then, based on users' historical learning behaviors and interest preferences, we recommend relevant courses to them in the knowledge graph. At the same time, the system displays recommendation reasons, such as "Recommended 'Machine Learning' because you have studied 'Python Programming'", enabling users to understand the recommendation basis more intuitively and enhancing their trust and satisfaction.

3.3 Experimental Dataset and Evaluation Metrics

To verify the effectiveness of the optimization methods and model designs proposed in this study, we selected user behavior logs from an online education platform of a university as the experimental dataset. The dataset includes multi-dimensional information such as users' course ratings, learning duration, and collection records, which can comprehensively reflect users' learning behaviors and interest preferences.

In terms of evaluation metrics, we selected commonly used metrics such as accuracy (Precision), recall (Recall), coverage (Coverage), and NDCG to evaluate the performance of the recommendation system. These metrics can reflect the accuracy, coverage, and practicality of the recommendation system from different angles and provide strong support for us to optimize the recommendation algorithm.

4. Experiments and Data Analysis

To verify the effectiveness of the optimization methods and model designs proposed in this study, we conducted a series of experiments and conducted an in-depth analysis of the experimental results. The following are the details of the experimental setup, experimental results, and interpretability analysis.

4.1 Experimental Setup

To comprehensively evaluate the performance of the optimized algorithm, we designed three groups of experiments, with the traditional User-based CF algorithm as the control group and two optimized algorithms as

the optimized groups. The experimental groups and algorithm descriptions are shown in the table below:

Table 1

Experimental Groups and Algorithms

Experimental Group	Algorithm Description
Control Group	Traditional User-based CF algorithm
Optimized Group 1	Hybrid model integrating course metadata
Optimized Group 2	LightGCN model with time decay

The experimental dataset adopts user behavior logs from an online education platform of a university, which includes multi-dimensional information such as users' course ratings, learning duration, and collection records, and can comprehensively reflect users' learning behaviors and interest preferences.

4.2 Experimental Results

In terms of accuracy, we compared the performance of the control group and the optimized groups in sparse data scenarios. The experimental results are shown in the table below:

Table 2

Experimental Results

Experimental Group	Accuracy (Precision)
Control Group	Baseline value
Optimized Group 1	Increased by 12.3%
Optimized Group 2	Increased by 15.7%

As can be seen from the table above, both Optimized Group 1 and Optimized Group 2 achieved significant improvements in accuracy, with Optimized Group 2 performing the best. In terms of cold start effects, we compared the changes in new user recommendation coverage. The experimental results showed that the new user recommendation coverage increased from 38% in the control group to 67% in the optimized group, indicating the effectiveness of the optimized algorithm in solving the new user cold start problem.

4.3 Interpretability Analysis

To verify the improvement of the optimized algorithm in interpretability, we conducted a case demonstration. A user was recommended the "Data Structures" course, and the system displayed the recommendation reason as "Because you frequently accessed algorithm-related resources and completed the C language course". By constructing a knowledge graph to associate recommendation reasons, users can understand the recommendation basis more intuitively, enhancing their trust and satisfaction.

In summary, the experimental results show that the collaborative filtering optimization model proposed in this study, which integrates multi-source data with graph neural networks, has significant advantages in improving the effectiveness of educational resource recommendations and has also made progress in interpretability.

5. Discussion

5.1 Adaptability to Educational Scenarios

The optimized model of collaborative filtering-based personalized educational resource recommendation algorithm proposed in this study demonstrates good adaptability in educational scenarios. Educational resource recommendation systems not only need to accurately meet learners' individualized needs but also need to balance the contradiction between "exploratory learning" and "accurate recommendation". Over-reliance on accurate recommendations may lead learners to fall into information silos, limiting the expansion of their knowledge horizons. Therefore, this study takes this into full consideration when optimizing the algorithm, ensuring the accuracy of recommendations while encouraging learners' exploratory behavior to a certain extent by integrating multi-source data and introducing time decay factors.

In addition, the cold start problem of educational resources is particularly prominent in educational scenarios. When new users join the platform, due to the lack of sufficient historical interaction data, traditional recommendation algorithms are often difficult to provide effective recommendations. This study provides initial recommendations for new users by utilizing embedding vectors generated from course description texts, effectively solving the cold start problem and improving the coverage and practicality of the recommendation system.

5.2 Limitation Analysis

Despite the achievements made in the optimization of collaborative filtering algorithms, this study still has some limitations. First, this study relies on high-quality course metadata annotations. However, in practical applications, the metadata of some courses may be incomplete or inaccurate, which may affect the performance of the recommendation system. In the future, natural language processing technology can be combined to automatically extract course features to reduce dependence on high-quality metadata annotations.

Second, the experimental design of this study is mainly based on user behavior log data from an online education platform of a university. Although this dataset is somewhat representative, it may not fully reflect the characteristics of all educational scenarios. Therefore, in the future, experimental verification needs to be conducted on a broader dataset to further improve the universality of the algorithm.

Furthermore, when constructing the knowledge graph, this study mainly focuses on the relationships between courses. However, in practical applications, social relationships between learners may also have an impact on recommendation results. In the future, exploring the integration of social network information into recommendation algorithms can further improve the accuracy and personalization of recommendations.

In summary, this study has achieved certain results in the optimization of collaborative filtering algorithms and provided new ideas and methods for personalized educational resource recommendations. However, there are still some limitations that need to be further studied and improved in the future.

6. Conclusion and Future Directions

6.1 Research Results

This study proposes an optimized model that integrates multi-source data with graph neural networks for the challenges faced by collaborative filtering algorithms in personalized educational resource recommendations and verifies its effectiveness through experiments. Specifically, this study mainly achieved the following results:

(1) Optimized algorithm design: Designed a hybrid recommendation model framework that integrates multi-source data such as student interest tags, course metadata, and learning behavior records, effectively alleviating the problem of data sparsity. At the same time, a time decay factor is introduced to dynamically adjust user interest weights, improving the accuracy and timeliness of recommendations. In addition, utilizing course description texts to generate embedding vectors solves the cold start problem for new users or new resources.

(2) Enhanced interpretability: Constructed a course knowledge graph to significantly enhance the interpretability of recommendation results by associating recommendation reasons, enabling users to understand the recommendation basis more intuitively and improving the learning experience and user satisfaction.

(3) Experimental verification: Experiments were conducted on the user behavior log dataset of an online education platform of a university, and the results showed that the optimized algorithm achieved significant improvements in evaluation metrics such as accuracy, recall, coverage, and NDCG, especially in sparse data scenarios and new user cold start problems.

6.2 Future Directions

Despite the achievements made in the optimization of collaborative filtering algorithms, this study still has some limitations, which also provide directions for future research:

(1) Combining natural language processing technology: Current research relies on high-quality course metadata annotations. In the future, natural language processing technology can be combined to automatically extract course features, reducing dependence on manual annotations and improving the universality and automation of the algorithm.

(2) Multi-source data fusion: In addition to the integrated student interest tags, course metadata, and learning behavior records, in the future, exploring the integration of social network information, learning progress, academic performance, and other multi-source data into recommendation algorithms can further improve the accuracy and personalization of recommendations.

(3) Cross-platform recommendation: With the increasing number of online education platforms, in the future, cross-platform recommendations under the federal learning framework can be explored to realize data sharing and collaborative model training between different platforms, improving the coverage and practicality of the recommendation system.

(4) Privacy protection: In the process of cross-platform recommendation and multi-source data fusion, privacy protection is an important issue. In the future, it is necessary to study how to strengthen the privacy protection of user data while ensuring the recommendation effect and enhance users' trust in the recommendation system.

In summary, this study provides new ideas and methods for personalized educational resource recommendations, but there are still many areas worthy of in-depth exploration and research. In the future, we will continue to focus on the optimization and improvement of collaborative filtering algorithms to provide learners with more accurate, interpretable, and personalized educational resource recommendation services.

Funding Project: 2022 China Ministry of Education Industry-University-Research Cooperation Project "Research on the Construction of Computer Education Resource Platform Based on Personalized Recommendation Algorithms" (Project Number: 220600136140317)

7. References

- Chen, J., & Zhang, H. (2022). A review of solutions to the cold start problem in recommendation systems. *Journal of Software*, 33(4), 1015-1033.
- Li, N., & Chen, W. (2022). Recent advances of graph neural networks in recommendation systems. *Artificial Intelligence and Pattern Recognition*, 10(4), 67-76.
- Liu, X., & Li, X. (2023). A review of interpretable recommendation systems. *Application Research of Computers*, 40(1), 1-7.
- Wang, X., & Li, M. (2022). A review of personalized recommendation systems based on deep learning. *Computer Science and Applications*, 12(3), 123-134.
- Yang, F., & Zhang, X. (2022). Application of multi-source data fusion in personalized recommendation. *Big Data and Artificial Intelligence*, 5(2), 34-41.
- Zhang, L., & Liu, T. (2023). Application and optimization of collaborative filtering algorithms in educational resource recommendation. *Educational Informatization Forum*, 7(1), 45-52.
- Zhao, L., & Wang, X. (2022). Improved collaborative filtering algorithm under data sparsity conditions. *Computer Engineering and Applications*, 58(15), 110-116.

