Learning recommendation based on hybrid collaborative filtering algorithm

Xuejian Huang¹, Gensheng Wang¹,²

¹Computer Practice Teaching Center, Jiangxi University of Finance and Economics, Nanchang 330013, China
²School of Humanities, Jiangxi University of Finance and Economics, Nanchang 330013, China

Corresponding author and e-mail: Xuejian Huang, 761780598@qq.com

Abstract. To solve the sparse matrix problem of learning resource recommendation algorithm based on collaborative filtering, a hybrid recommendation algorithm combining user learning behaviour and rating is proposed. First of all, learning behaviour characteristics are defined and different characteristics are given weight values; then, according to the user's learning behaviour log, the method of weight accumulation to calculate the user's learning behaviour value of resources is used; finally, the learning behaviour matrix and rating matrix is used to build a hybrid collaborative filtering recommendation algorithm. Experimental results show that the hybrid recommendation algorithm has lower root mean square error than the traditional collaborative filtering algorithm.

1. Introduction
Recommendation technology is an important method to solve the problem of network learning resource overload [1]. At present, network learning resource recommendation algorithms mainly include content-based filtering, association rules and collaborative filtering [2]. Collaborative filtering recommendation algorithm does not need to analyze the characteristics of resources and users [3], which is the most popular algorithm used in personalized recommendation of network learning resources. However, the traditional collaborative filtering algorithm is faced with the problem of matrix sparsity [4]. Many scholars have also proposed relevant improvement methods to solve this problem. For example, reference [5] adds other implicit information (such as browsing, purchase, click history, etc.) to fill in the User-Item rating matrix; reference [6] proposes an improved collaborative filtering recommendation algorithm based on implicit scoring and similarity transfer; reference [7] proposes an improved collaborative filtering recommendation algorithm which integrates the trust between learners in social networks. It is found that mining other potential features of users and adding implicit information between users and item are the main methods to solve the problem of matrix sparsity in collaborative filtering algorithm. This paper proposes an improved collaborative filtering recommendation algorithm based on user's learning behavior and rating, which can alleviate the problem of sparse rating matrix and realize more accurate personalized online learning resource recommendation.

2. Hybrid collaborative filtering recommendation algorithm
To solve the problem of sparse matrix in collaborative filtering algorithm, a hybrid collaborative
filtering recommendation algorithm is proposed, which combines rating data and learning behavior data. The specific algorithm flow is shown in Figure 1.

![Figure 1. Algorithm flow.](image)

2.1. Collaborative filtering based on rating matrix

2.1.1. User-Resource rating matrix. Convert the user's historical rating of resources into a rating matrix of \( M \times N \), as shown in Figure 2.

![Figure 2. User-Resource rating matrix.](image)

\( r_{mn} \) is the rating of user \( u_m \) on resource \( i_n \), and the value range is 0-10. When the number of resources is large, users may only rating a small part of the resources, resulting in the whole matrix sparse.

2.1.2. User similarity based on rating. The collaborative filtering recommendation algorithm based on rating establishes user interest similarity model according to user's rating data of resources, and recommends resources through users with similar interests [8]. Similarity calculation is the core part of recommendation algorithm, mainly including cosine similarity, Pearson similarity and Jaccard...
Cosine similarity is the most commonly used method, and its calculation is shown in Eq. (1):

\[ sim\_rating(u, v) = \cos(\overline{s_u}, \overline{s_v}) = \frac{\sum_{i \in I} (r_u - \overline{r_i}) \cdot (r_v - \overline{r_i})}{\sqrt{\sum_{i \in I} (r_u - \overline{r_i})^2} \cdot \sqrt{\sum_{i \in I} (r_v - \overline{r_i})^2}} \]  

\[ \overline{s_u} \] and \[ \overline{s_v} \] represent the rating vectors of users \( u \) and \( v \) for the same resource respectively. \( I \) is the resource collection. \( r_u \) and \( r_v \) represent user \( u \) and \( v \) ratings for resource \( i \), respectively. \( \overline{r_i} \) represents the average rating of user \( u \) and \( v \) for all rating resources, respectively.

### 2.2. Collaborative filtering based on learning behavior matrix

#### 2.2.1. Behavior characteristics and weight calculation

In addition to the explicit rating, users will also generate a lot of implicit data such as browsing, collecting, downloading, commenting, sharing, etc. These implicit data also reflect the user's interest in learning resources. The related learning behaviors of users to resources are transformed into corresponding interest weights. The weight value of interest corresponding to each behavior is determined by expert judgment. The final result is (browse=1, collection=2, download=3, comment=3, praise=4, share=5).

#### 2.2.2. User-Resource behavior matrix

According to the user learning behavior log, behavior characteristics and weight, the User-Resource behavior matrix is constructed, as shown in Figure 3.

![Figure 3. User-Resource behavior matrix.](image)

\( s_{mn} \) is the cumulative weight of user \( m \) learning behavior to resource \( n \).

#### 2.2.3. User similarity based on behavior

In order to keep consistent with user similarity based on rating matrix, cosine similarity is still used to calculate user behavior similarity, as shown in Eq. (2):

\[ sim\_behavior(u, v) = \cos(\overline{b_u}, \overline{b_v}) = \frac{\sum_{i \in I} s_{ui} \cdot s_{vi}}{\sqrt{\sum_{i \in I} s_{ui}^2} \cdot \sqrt{\sum_{i \in I} s_{vi}^2}} \]  

\( \overline{b_u} \) and \( \overline{b_v} \) represent the cumulative weight of user \( u \) and \( v \) to the learning behavior of the same resource, respectively.

### 2.3. Fusion user similarity
User similarity consists of two parts: user rating similarity and user learning behavior similarity. The fusion user similarity is calculated by linear weighting. The calculation is shown in Eq. (3):

\[
\text{sim}(u,v) = \lambda \text{sim}_{\text{rating}}(u,v) + (1 - \lambda) \text{sim}_{\text{behavior}}(u,v)
\]

(3)

Where \( \lambda \) is the fusion weight factor and the value range is \([0,1]\). \( \text{sim}_{\text{rating}}(u,v) \) is the rating similarity of user resources. See Eq.(1) for calculation. \( \text{sim}_{\text{behavior}}(u,v) \) is the similarity of user learning behavior. See Eq.(2) for calculation.

2.4. Rating forecast and recommended results

According to the similarity degree of users, the most similar \( k \) users are selected as the neighboring user set \( S_u = \{s_{u1}, s_{u2}, s_{u3}, \ldots, s_{uk}\} \) of the target user \( u \). Based on the neighboring user set \( S_u \), the target user \( u \) rating of resources is predicted. The specific calculation is shown in Eq. (4):

\[
\hat{r}_u = \bar{r}_v + \frac{\sum_{v \in S_u \cap v \neq u} \text{sim}(u,v) \times (r_v - \bar{r}_v)}{\sum_{v \in S_u \cap v \neq u} \text{sim}(u,v)}
\]

(4)

Where \( \hat{r}_u \) represents the prediction rating of target user \( u \) to resource \( i \). \( \bar{r}_v \) is the average value of the historical resource rating of the target user \( u \). \( \text{sim}(u,v) \) is the similarity between user \( u \) and user \( v \). \( r_v \) is the rating of similar user \( v \) on resource \( i \). \( \bar{r}_v \) is the mean value of similar users' \( v \) history rating. Finally, the top-N resources with the highest prediction rating are recommended to the target user \( u \).

3. Experimental verification and result analysis

3.1. Experimental data

The experimental data comes from the background management data of JXUFE University's online teaching platform, which includes 1215 courses and 42452 student users, from which 3000 users and 500 courses are selected as the experimental data.

3.2. Evaluating indicator

In the experiment, RMSE (Root Mean Square Error) is used to evaluate the performance of the algorithm. The smaller the RMSE value is, the better the performance of the algorithm is. RMSE calculation is shown in Eq. (5):

\[
\text{RMSE} = \sqrt{\frac{\sum_{(u,i) \in T} (r_{ui} - \hat{r}_{ui})^2}{|T|}}
\]

(5)

Where \( T \) is the test data. \(|T|\) is the size of the test data set. \( r_{ui} \) is the actual rating of user \( u \) for \( i \). \( \hat{r}_{ui} \) is the predicted rating of user \( u \) for \( i \).

3.3. Experimental result analysis

3.3.1. Experimental comparison of the number \( k \) of different neighboring users. During the experiment, the fusion weight factor \( \lambda \) in Eq. (3) is set as 0.7. The number of neighboring users \( k \) is increased from 10 to 50, each step is 5. There are 9 groups of experiments, and RMSE values of each group of experiments are calculated. The experimental results are shown in Figure 4.
According to the experimental results, when the number of neighboring users $k$ is 30, the RMSE value is the minimum.

3.3.2. Experimental comparison of different fusion weight factors $\lambda$. The fusion weight factor $\lambda$ in Eq. (3) controls the weight of user rating similarity and user learning behavior similarity in the final fusion user similarity, which is a key factor of this algorithm. The value range of $\lambda$ is $[0, 1]$. The experiment starts with the value of $\lambda$ equal to 0, increasing by 0.1 every time, a total of 11 times. The number of neighboring users $k$ is set to 30. The experimental results are shown in Figure 5.

According to the experimental results, the RMSE value is the minimum when the fusion weight factor $\lambda$ is 0.7.

3.3.3. Experimental comparison with traditional collaborative filtering algorithm. In order to verify the performance of this hybrid collaborative filtering recommendation algorithm, this hybrid algorithm is compared with the other two traditional collaborative filtering recommendation algorithms. Three algorithms are described as follows:

- $HCF$, Hybrid collaborative filtering recommendation algorithm;
- $T_{-}UCF$, Traditional collaborative filtering recommendation algorithm based on users;
- $T_{-}ICF$, Traditional collaborative filtering recommendation algorithm based on items;

The number of neighboring users $k$ of the three algorithms is set to 30, the fusion weight factor $\lambda$ of $HCF$ is set to 0.7, and other parameters are consistent. The experimental results are shown in Figure 6.
By comparing the experimental results, the RMSE value of \textit{HCF} is smaller than that of two traditional collaborative filtering recommendation algorithms \textit{T_UCF} and \textit{T_ICF}. It is proved that the hybrid collaborative filtering recommendation algorithm with user learning behavior has better performance than the traditional collaborative filtering recommendation algorithm.

3.3.4. Experimental comparison with other improved algorithms. In order to further verify the performance of \textit{HCF}, \textit{HCF} is compared with the collaborative filtering recommendation algorithm based on implicit rating and similarity transfer (IRST-CF) in reference [6], and the collaborative filtering recommendation algorithm which integrates the trust between learners in social networks (Social-CF) in reference [7]. The fusion weight factor $\lambda$ of \textit{HCF} is set to 7. The transfer path threshold of IRST-CF is set to 2. The trust calculation weight parameter of Social-CF is set to 0.4. The number of neighboring users $k$ of the three algorithms is increased from 10 to 50, and each step is 5. There are 9 groups of experiments. RMSE value was used as the experimental comparison result. The results are shown in Figure 7.

The experimental results show that \textit{HCF} can keep smaller RMSE value than the other two improved algorithms IRST-CF and Social-CF under different number of neighboring users, which further proves the performance of this algorithm.

4. Conclusions
To solve the sparse matrix problem of learning resource recommendation algorithm based on collaborative filtering, an improved hybrid collaborative filtering learning resource personalized recommendation algorithm is proposed, which integrates user learning behavior and rating. The
experimental results show that the algorithm is effective. But the algorithm in this paper also has some shortcomings. On the one hand, it does not consider the gradualness of the learning process, so the learning resources have a gradual relationship; on the other hand, the algorithm does not consider the change of user interest, because the user interest is likely to change with time, so the historical data has timeliness. These problems need further study.

Acknowledgement
Fund: NSFC funded project (Project No.: 71461012); Science and technology project of Jiangxi Provincial Department of Education (GJJ181550).

References
[1] Wang X, Zhu Z, Yu J, et al. A learning resource recommendation algorithm based on online learning sequential behavior[J]. International Journal of Wavelets Multiresolution & Information Processing, 2019,17(1):1-17.
[2] Ding J H,Liu H Z. Accurate Recommendation of Learning Resources Based on Multi-dimensional Correlation Analysis in Age of Big Data[J]. e-Education Research,2018,39(02):53-59 +66.(In Chinese)
[3] Shi Y , Larson M , Hanjalic A. Collaborative Filtering beyond the User-Item Matrix: A Survey of the State of the Art and Future Challenges[J]. ACM Computing Surveys (CSUR), 2014, 47(1):1-45.
[4] Guo L J, Liang J Y, Zhao X W. Collaborative Filtering Recommendation Algorithm Incorporating Social Network Information[J]. Pattern Recognition and Artificial Intelligence, 2016,29(03):281-288. (In Chinese)
[5] Shu L C , Yu X P . Matrix Factorization for Recommendation with Explicit and Implicit Feedback[J]. Knowledge-Based Systems, 2018,158:109-117.
[6] Fu F, Dou Y S, Han P, et al. Learning resource recommendation based on implicit scoring and similarity propagation[J]. Application Research of Computers,2017,34(12):3725-3729. (In Chinese)
[7] Ding Y G, Zhang X, Sang Q X,et al. The Collaborative Filtering Recommendation of Learning Resources Combined with Learners' Social Network[J]. Modern Educational Technology,2016,26(02):108-114. (In Chinese)
[8] Lei W , Qing F , Zhou J . Improved Personalized Recommendation based on Causal Association Rule and Collaborative Filtering[J].International Journal of Distance Education Technologies, 2016,14(3):21-33.
[9] Shu J, Shen X, Liu H, et al. A content-based recommendation algorithm for learning resources[J]. Multimedia Systems, 2017(1): 1-11.