Research on Hybrid Recommendation Algorithm of Educational Courseware Resources Based on Heterogeneous Information Fusion

Sheng-bin Wu[1](✉) and Ying Yang[2]

[1] College of Information Engineering, Changsha Medical University, Changsha 410219, China
[2] School of Computer, Guangdong AIB Polytechnic College, Guangzhou 510507, China

Abstract. Aiming at the poor recommendation effect of the traditional hybrid recommendation algorithm for educational courseware resources, a hybrid recommendation algorithm based on heterogeneous information fusion is proposed. Through the description of the characteristics of educational courseware resources, the attributes are mapped into the rating matrix, and the average values of the attributes of all the evaluated educational courseware resources are calculated. After the similar items are merged, the double attribute rating matrix is obtained. The modular matrix is used to process the sub factor sequence, and the mean value of the correlation coefficient corresponding to each sub factor sequence is calculated. This paper studies the coupling relationship between educational courseware resources, completes the modular processing of educational courseware resources, and realizes the recommendation of network hybrid information combined with the design of network hybrid information recommendation algorithm. The experimental results show that the hybrid recommendation algorithm based on heterogeneous information fusion can better solve the problem of low recommendation efficiency caused by sparse score matrix and "cold start". The recommendation effect is better than the traditional collaborative recommendation algorithm, and the quality is higher.

Keywords: Heterogeneous information fusion · Educational courseware resources · Hybrid recommendation · Recommendation algorithm

1 Introduction

With the continuous development of the Internet and information technology, the popularity of intelligent terminals, and the large-scale coverage of 4G network and wireless network, people's life and work have undergone great changes. People take the network as the medium, and the learning mode of distance learning course resources is also developing continuously. Learners can learn through network courseware or other learning resources, and are no longer limited by time and space [1]. However, if the online education platform simply presents the offline course content and teaching resources,

W. Fu et al. (Eds.): eLEOT 2021, LNICST 390, pp. 313–324, 2021.
https://doi.org/10.1007/978-3-030-84386-1_26

it will lead to the poor interaction between teachers and students and the learning efficiency of students due to the time and space barrier of distance education; at the same time, there is no difference in one-way teaching for all students, which leads to poor personalization. The explosive growth in the number and scale of educational resources makes it difficult for ordinary learners to choose learning resources [2]. The resources obtained through traditional search engines are usually complex and inaccurate, which can not satisfy them. At the same time, although the current organizational structure of learning resources can meet the requirements of online learning, due to the proposal of personalized learning, learning resources also need to meet the needs of learners for the structural, dynamic and retrievable resources [3].

The study of learning resources in online education has a long history abroad. As early as 1998, the United States Department of education and the National Library of Education launched the gem project, which proposed to use metadata coding to describe, organize and manage network resources, so as to facilitate people to retrieve and obtain learning resources. Luo J et al. [4] because the traditional keyword search method is difficult for learners to find and obtain the most appropriate resources from the vast amount of educational resources, context aware resource recommendation service has become an important part of pervasive learning environment. To solve this problem, a context aware resource recommendation model and related recommendation algorithm are proposed in this paper The first part combines the content-based and collaboration based recommendation mechanism, and introduces the individual preference tree to consider the multi-dimensional attributes of resources, the rating matrix of learners and the energy of access preferences In the second part, in order to enhance the effectiveness of the recommendation, the connection type correlation and time satisfaction are calculated according to other relevant contexts. Then, the candidate resources can be filtered and sorted by combining the two parts to generate the recommendation results. The simulation results show that the proposed method is effective It is better than other most advanced algorithms in traditional and newly proposed metrics, and may be more suitable for pervasive learning environment H et al. [5] in view of the fact that the existing course video recommendation system is generally limited to one course and ignores the knowledge correlation between courses, in this work, a two-stage cross course video recommendation algorithm is proposed. The algorithm considers both the implicit feedback of learners and the knowledge association between course videos. Firstly, collaborative filtering is used to generate video seed set, which seeds Secondly, a cross course video association knowledge map is constructed, and the random walk algorithm is used to measure the relevance of the course video. The relevance is based on each video seed as the starting node, and is extended to the video subgraph. Then, several cross course video oriented subgraphs are recommended to the learners. The experimental results show that the cross course video association knowledge map can be used to measure the relevance of the course video. The recommendation algorithm is superior to the traditional recommendation algorithm based on collaborative filtering in accuracy, recall rate and knowledge relevance.

Scholars in the field of educational technology in China have also carried out a lot of research on learning resources in online education. Ren Lei [6] aimed at the problem of excessive granularity of user interest descriptions in traditional collaborative filtering algorithms, and the problem of inaccurate calculation of similarity caused by sparse score matrix., A hybrid recommendation algorithm based on incremental learning WHHR is proposed. This algorithm constructs a content-based user model through Widrow-Hoff incremental learning, and combines the collaborative filtering recommendation mechanism to achieve score prediction. The experiment verifies that the WHHR algorithm is in convergence speed and Compared with similar recommendation algorithms, the accuracy of recommendation has been greatly improved; Yangfengrui et al. [7] has relatively low accuracy and reliability for the traditional recommendation algorithm. In view of the problem of cold start of users and projects, this paper proposes a hybrid recommendation algorithm based on probability matrix decomposition. Firstly, the trust relationship of users is mined from the perspective of user rating, and then the relevance between items is measured by label context according to user characteristics, and then integrated into probability matrix model for recommendation. The experiment shows that, the proposed algorithm has achieved good results in the accuracy of recommendation compared with the conventional method.

2 The Design of Algorithms for Mixed Recommendation of Educational Courseware Resources

2.1 Build a Dual Attribute Score Matrix

All features of educational courseware resources are represented by a set, such as a. $c = \{c_1, c_2, c_3, \ldots, c_n\}$ vector composed of features of educational courseware resources is used to represent an educational courseware resource

$$p = \left[p_1, p_2, p_3, p_4, \ldots, p_n\right] \tag{1}$$

Among them, p_j describes a certain attribute characteristic of the educational courseware resource. When the educational courseware resource has this attribute characteristic, its value is 1; when it does not have this characteristic, its value is 0.

First, map the attributes of the educational courseware resources in the scoring matrix through the characteristic description of the educational courseware resources. Assuming that a_{ij} represents user u's rating of the j-th educational courseware resource attribute of the educational courseware resource i, then user u's rating of the j-th educational courseware resource attribute the score $R(u, j)$ is expressed as the average value of the j th educational courseware resource attribute of all the evaluated educational courseware resources by user u:

$$R(u, j) = \frac{\sum_{i=1}^{N} a_{ij}}{N} = \frac{\sum_{i=1}^{N} p_j \times R(u, i)}{N} \tag{2}$$

Among them, $R(u, i)$ is the score of user u on educational courseware resources i, and N is the total number of educational courseware resources evaluated by N.

Similarly, all features of a user can be represented by a set, such as $b = \{b_1, b_2, b_3, \ldots, b_n\}$. A user can be represented by a vector composed of user features

$$q = \left[q_1, q_2, q_3, q_4, \ldots, q_n\right] \tag{3}$$

Among them, q_k is used to describe a certain user attribute characteristic of the user. When the user has this attribute characteristic, its value is 1; when it does not have this characteristic, its value is 0.

Map the user's attribute characteristics to the newly established user-education courseware resource attribute scoring matrix. The score of user attribute k on the j-th education courseware resource attribute of $R(k,j)$ is expressed as the mean value of user u's evaluation of the j-th education courseware resource attribute:

$$R(k,j) = \frac{q_k}{H} \times R(u,j) = \frac{\sum_{i=1}^{N} q_k \times p_j \times R(u,i)}{N \times H} \tag{4}$$

Where H is the number of user attributes owned by user u.

After the same category items are merged, the double attribute score matrix is obtained as follows:

$$\text{Two attribute scoring matrix} = \begin{pmatrix} R(1,1), R(1,2), R(1,3), \ldots, R(1,j) \\ R(2,1), R(2,2), R(2,3), \ldots, R(2,j) \\ R(3,1), R(3,2), R(3,3), \ldots, R(3,j) \\ \ldots\ldots \\ R(k,1), R(k,2), R(k,3), \ldots, R(k,j) \end{pmatrix} \tag{5}$$

In the above description, the attributes of the educational courseware resources are mapped in the scoring matrix, and the average value of the attributes of all the evaluated educational courseware resources by the user is calculated. After the similar items are combined, a dual-attribute scoring matrix is obtained.

2.2 Modular Processing of Educational Courseware Resources

If the recommendation of educational courseware resources is taken as a system, modular processing refers to distinguishing the key factors of the recommendation of educational courseware resources, and rationally analyzing the sub-factors in the module [8]. Suppose u_{ij} represents the j-th factor in Educational Courseware Resource Module i, and its influence is recorded as x_{ij}.

When using the sequence of n sub-factors to deal with the influencing factors of educational courseware resource recommendation, a modular matrix based on personalized adaptive learning will be formed, namely:

$$\left(X_1', X_2' \cdots, X_n'\right) = \begin{pmatrix} x_1'(1) & x_2'(1) & \cdots & x_{nt}'(1) \\ x_1'(2) & x_2'(2) & \cdots & x_{n-1}'(2) \\ \vdots & \vdots & \vdots & \vdots \\ x_1'(m) & x_2'(m) & \cdots & x_n'(m) \end{pmatrix} \tag{6}$$

Using personalized adaptive learning to modularize the sequence of sub-factors, you can get:

$$(X_0, X_1, \cdots, X_n) = \begin{pmatrix} x_0(1) & x_1(1) & \cdots & x_n(1) \\ x_0(2) & x_1(2) & \cdots & x_n(2) \\ \vdots & \vdots & \vdots & \vdots \\ x_0(m) & x_1(m) & \cdots & x_n(m) \end{pmatrix} \tag{7}$$

The mean value of the correlation coefficient corresponding to each sub-factor sequence is calculated separately to reflect the correlation between the educational courseware resources and the sub-factor sequence [9], which is called the correlation sequence, which is recorded as:

$$r_{i0} = \frac{1}{m}\sum_{k=1}^{m} \zeta_i(k) \tag{8}$$

Then the educational resources of u_{ij} can be recommended as follows:

$$\lceil u_{ij} \rceil = \begin{cases} \frac{(\text{arcmax}\, \zeta_i(j) - r_{i0})}{\zeta_i(j)}, u_{ij} \text{ The recommendation effect was negative} \\ \frac{(r_{i0} - arc \max \zeta_i(j))}{\zeta_i(j)}, u_{ij} \text{ The recommended effect is positive} \end{cases} \tag{9}$$

The coupling degree function is used to study the coupling relationship between educational courseware resources. The coupling degree between different educational courseware resources can be expressed as:

$$T_m = \left\{ \frac{(u_1 \times u_2 \times \ldots u_m)}{\left[\prod \left(u_i + u_j\right)\right]} \right\}^{\frac{1}{m}}, i \neq j \tag{10}$$

Among them, the value of T_m is between 0 and 1, and u_i represents the total recommendation effect of educational courseware resources. The calculation formula is:

$$u_i = \sum^{m} \lambda_{ij} u_{ij} \tag{11}$$

Although formula (11) can be used to calculate the coupling degree between educational courseware resources, it can not reflect the status of personalized adaptive level to recommend educational courseware resources. Therefore, a coupling coordination function is constructed to calculate the coupling degree between educational courseware resources

$$\begin{cases} F = \alpha u_1 + \ldots + \beta u_m \\ A = \sqrt{T_m} \cdot F \end{cases} \tag{12}$$

Among them, A represents the coupling coordination degree function, T_m represents the coupling degree of the education courseware resources, F represents the coordination index, and α and β represent the overall recommendation weight of the education courseware resources.

Modularization of the sub-factor sequence is carried out using the modular matrix, and the mean value of the correlation coefficient corresponding to each sub-factor sequence is calculated. The coupling relationship between the educational courseware resources is studied using the coupling degree function, and the modular processing of the educational courseware resources is completed. Next, by calculating the similarity of educational courseware resources, to simplify the recommendation process of educational courseware resources.

2.3 Design a Hybrid Recommendation Algorithm for Educational Courseware Resources

Every piece of information released by users will involve a topic or multiple topics, and the characteristics of user information release match the topic model of educational courseware resources [10–12]. Therefore, the topic model is used to determine the topic distribution of the information posted by the user, so as to initially determine the user's interest orientation. The flow of the hybrid recommendation algorithm for educational courseware resources is shown in Fig. 1.

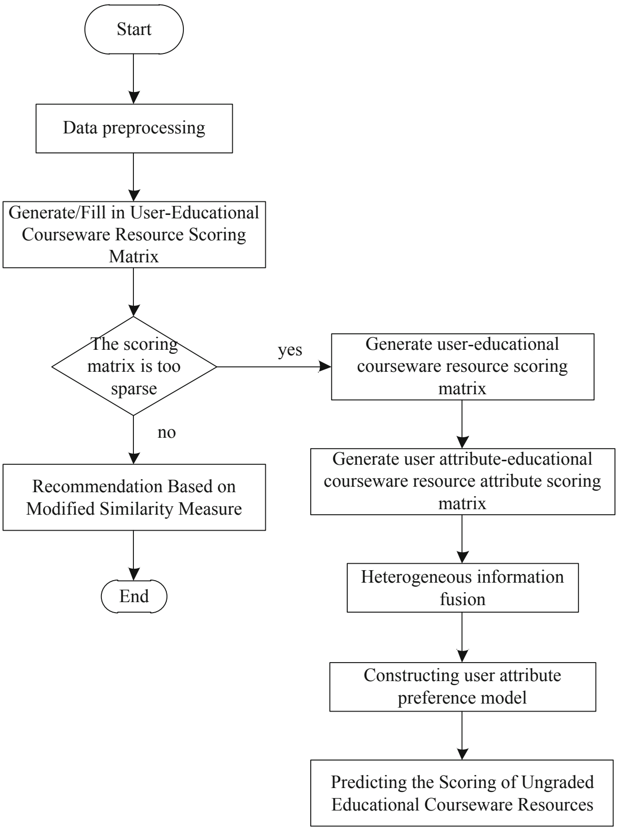

Fig. 1. Flow chart of hybrid recommendation algorithm for educational courseware resources

For the topic distribution of educational courseware resources, definition $C = \{C_1, C_2, \cdots, C_T\}$ represents the theme set of educational courseware resources. For a piece of educational courseware resource t, the posterior probability is $p(C_i \mid t)$. Therefore, the main part vector of the education courseware resource composed of posterior probability is $(p(C_1 \mid t), p(C_2 \mid t), \ldots, p(C_T \mid t))$, T Represents the number of subject collections of educational courseware resources. For the user's interest orientation, suppose $\{t_1, t_2, \cdots, t_d\}$ represents the set of educational courseware resources released by the user, then the T dimensional vector $(v_1, v_2, \cdots, v_T)$ represents the user's interest orientation in the educational courseware resources, and the calculation formula is:

$$v_i = \frac{1}{d}\sum_{j=1}^{d} p(C_i \mid t_j) \tag{13}$$

According to the user's interest weight, v_i is improved by using personalized adaptive learning

$$v_i = \frac{1}{d}\sum_{j=1}^{d} \alpha_j p(C_i \mid t_j) \tag{14}$$

Among them, $\sum_{j=1}^{d} \alpha_j = 1$ and α_j represent the user's interest weight value for educational courseware resources, which can reflect the user's preference set for educational courseware resources. The larger the value of α_j, the higher the user's preference for educational courseware resources.

Assuming that the education courseware resource t is composed of n words, the n-th word is recorded as w_n, and the topic of w_n is defined as z_{w_n}, then the probability of z_{w_n} is calculated as:

$$P\big(z_{u_i} = j \mid Z_{t,\neg i}, t, \varphi, \alpha\big) \propto \frac{P\big(z_{w_i} = j, Z_{t,\neg i}, t \mid \varphi, \alpha\big)}{P\big(Z_{t,\neg i}, t \mid \varphi, \alpha\big)} \tag{15}$$

On the basis of formula (15), if the topic distribution of word w_n is $V_{w_n} = (v_1, v_2, \cdots, v_T)$, then the standardization probability of v_i is:

$$v_j = \frac{P\big(z_{x_i} = j \mid Z_{t,\neg i}, t, \varphi, \alpha\big)}{\sum_{j=1}^{T} P\big(z_{u_i} = j \mid Z_{t,\neg i}, t, \varphi, \alpha\big)} \tag{16}$$

The probability that educational courseware resource t belongs to topic j is as follows:

$$\theta_{t,j} = \frac{n(j,t) + \alpha}{n(t) + T\alpha} \tag{17}$$

According to $\theta_{t,j}$, the average preference similarity of users in M cycles can be defined:

$$\mathrm{sim}_M = \frac{\sum_{\frac{1}{2}}^{M} \mathrm{sim}(u_{k-1}, u_k)}{M - 1} \tag{18}$$

In order to facilitate the comparison, the average preference similarity of users can be processed by personalized adaptive learning, and the user preference similarity can be calculated

$$\mathrm{sim}(U_{k-1}, U_k) = \cos\theta = \frac{\overrightarrow{u_{k-1} \cdot \overrightarrow{u_k}}}{|\mu_{k-1}\| \cdot |\mu_k\|} \tag{19}$$

In conclusion, heterogeneous information fusion is used to improve the recommendation method of educational courseware resources. Through modular processing of educational courseware resources, a network hybrid information recommendation algorithm is designed to realize the recommendation of network hybrid information.

3 Experimental Comparative Analysis

By comparing the difference in recommendation accuracy and the number of recommended educational courseware resources between the education courseware resource hybrid recommendation algorithm based on heterogeneous information fusion and the traditional user-based collaborative recommendation algorithm, it is verified that the recommendation algorithm in this paper can solve the sparse score matrix and cold start. And other issues.

3.1 Experimental Environment and Experimental Data Set

The experimental environment is inter (R) core (TM) i7–4790 CPU @ 3.60 GHz, the memory is 8.0 GB, and the operating system is Windows 7. Pychar 2017.1 and MATLAB are used to implement the data preprocessing and recommendation algorithm_R2017b.

The experimental data set is movielens data set, which is collected and founded by the group lens educational courseware resource group of Minnesota University in the United States. It can receive users' ratings on movies and provide personalized movie recommendation to them.

Among them, each evaluation score ranges from 1 to 5. Each user provides his / her age, gender, occupation and other attribute information when registering, and each movie provides feature information such as film title, release date, theme type, etc.

In this paper, the movielens LM data set is selected and divided into five subsets, which are disjoint. Each subset contains 80000 scoring base data sets and 20000 scoring test data sets. The base data set and the test data set are complementary.

The sparsity of base data set is 95%, which is a typical sparse matrix, which can help to verify the improvement of the algorithm in solving the problem of sparse score matrix.

3.2 Experimental Methods and Evaluation Indicators

Use offline experiments to verify the correctness of the improved algorithm. Each experiment selects a subset, uses the improved algorithm to predict the user ratings on the base data set, and then uses the true score of the test to calculate the error of the predicted score to verify The effectiveness of the algorithm.

Use evaluation indicators based on prediction accuracy to calculate the accuracy of the recommendation algorithm, and use MAE to measure its error:

$$MAE = \frac{\sum_{s \in S} \left| R_S - \hat{R}_s \right|}{|S|} \tag{20}$$

Among them, S represents the collection of all products, R_S represents the real score of products, and $\hat{R}_s$ represents the predicted score of product s by the current user calculated by the recommendation system.

Mae calculates the average error between the predicted score and the real score. The smaller the value, the higher the quality of the recommendation algorithm.

3.3 Experimental Results and Analysis

3.3.1 Experimental Results Recommended by the Algorithm

Select the number of neighbor user sets to be 10, 20, 30, 40, 50, and use the user-based collaborative recommendation algorithm (User-based C) and the hybrid recommendation algorithm of educational courseware resources based on heterogeneous information fusion to conduct comparative experiments, and predict the scoring result is compared with the average absolute error of the test data set, and the result is shown in Fig. 2.

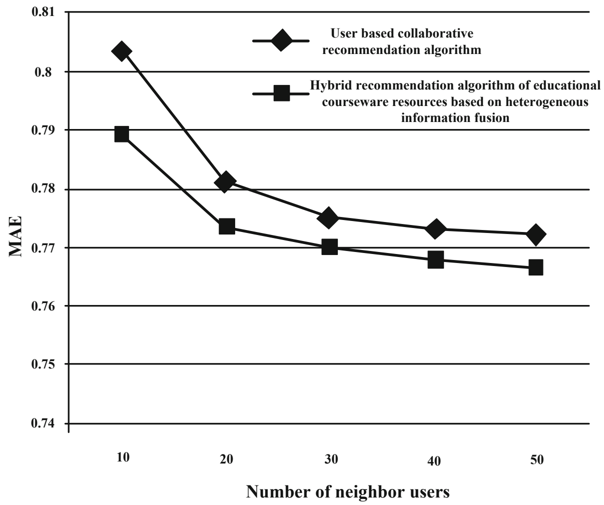

Fig. 2. Comparison of average absolute error of recommended algorithms

As shown in the figure above, the hybrid recommendation algorithm based on heterogeneous information fusion has smaller average absolute error and better recommendation effect than collaborative recommendation algorithm. Based on the double attribute rating matrix, the improved algorithm uses heterogeneous information fusion to simulate the user's attribute preference model, and makes reasonable rating prediction for the non rated items, which alleviates the inaccurate recommendation problem caused by the sparse rating data of traditional algorithms.

3.3.2 Algorithm Cold Start Recommended Experimental Results

(1) Make recommendations for new users
Twenty users were randomly selected and their rating data was set to zero to simulate the situation of new users entering the recommendation system. The number of neighbor user sets are 20, 40, 60, 80, 100, 120, 140, and the user-based collaborative recommendation algorithm (User-based CF) and the hybrid recommendation algorithm of educational courseware resources based on heterogeneous information fusion Users make recommendations and compare the number of items recommended by the two for new users. The experimental results are shown in Fig. 3.

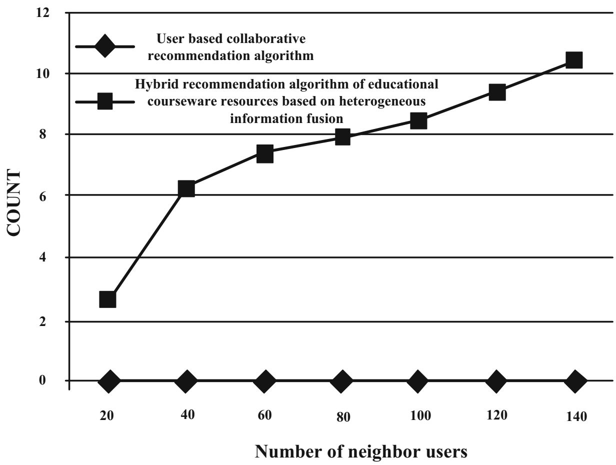

Fig. 3. Comparison of recommendation algorithms for solving user's "cold start" problem

As shown in Fig. 3, when new users enter, the hybrid recommendation algorithm based on heterogeneous information fusion can recommend more items for new users than the collaborative recommendation algorithm. With the increase of the number of neighbor users, the recommended items also increase, which effectively alleviates the impact of user cold start on the recommendation quality.

(2) Recommend new projects
Ten items were randomly selected and their score data were set to zero to simulate the situation of new items entering the recommendation system. The number of neighbor users is 20, 40, 60, 80, 100, 120, 140, respectively. The user based CF algorithm and the hybrid recommendation algorithm based on heterogeneous information fusion are used to recommend new projects. The number of potential users is compared. The experimental results are shown in Fig. 4.
As shown in the figure above, when a new project enters the system, the hybrid recommendation algorithm of educational courseware resources based on heterogeneous information fusion can recommend this new project to more users

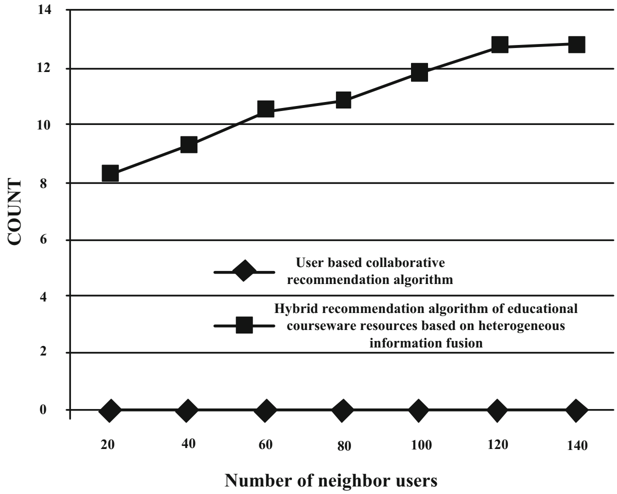

Fig. 4. Comparison of recommended algorithms for solving the "cold start" problem of items

than the collaborative recommendation algorithm, and as the number of neighbor projects increases, Recommending new items to more potential preference users has alleviated the impact of the cold start of items on the recommendation quality.

4 Conclusion

This paper proposes a hybrid recommendation algorithm for educational courseware resources based on heterogeneous information fusion. Heterogeneous information fusion is used to improve the recommendation method of educational courseware resources. Through modular processing of educational courseware resources, a network hybrid information recommendation algorithm is designed to realize the recommendation of network mixed information. The results show that the algorithm has better recommendation effect. Online education platform can not only help students learn knowledge more effectively, but also can obtain feedback information from users and collect diverse learning behavior data. The mining and analysis of learning behavior data is beneficial to the development of educational resources, and can help educators improve curriculum design and teaching methods, make effective evaluation on learners and promote them to improve their learning methods and improve learning efficiency. This paper mining the association between knowledge points in a course according to the user learning behavior data, and then provide the user with knowledge point recommendation service according to the association between knowledge points, which helps to consolidate the students' knowledge mastery and improve the learning efficiency.

Fund Projects. This paper is a scientific research fund project of Hunan Provincial Department of Education (project number: 20c0218).

References

1. Huayue, C.: The study on interesting mining based on topic model for web educational resource recommendation. Adv. Inf. Sci. Serv. Sci. **4**(2), 275–281 (2012)
2. Manouselis, N., Vuorikari, R., Assche, F.V.: Collaborative recommendation of e-learning resources: an experimental investigation. J. Comput. Assist. Learn. **26**(4), 227–242 (2010)
3. Lau, B.Y., Lee, C.S., Singh, Y.P.: A folksonomy-based lightweight resource annotation metadata schema for personalized hypermedia learning resource delivery. Interact. Learn. Environ. **23**(1), 79–105 (2015)
4. Luo, J., Dong, F., Cao, J., et al.: A context-aware personalized resource recommendation forpervasive learning. Clust. Comput. **13**(2), 213–239 (2010)
5. Zhu, H., Liu, Y., Tian, F., et al.: A cross-curriculum video recommendation algorithm based on a video-associated knowledge map. IEEE Access 1 (2018)
6. Liu, S., Lu, M., Li, H., et al.: Prediction of gene expression patterns with generalized linear regression model. Front. Genet. **10**, 120 (2019)
7. Liu, S., Li, Z., Zhang, Y., et al.: Introduction of key problems in long-distance learning and training. Mob. Netw. Appl. **24**(1), 1–4 (2019)
8. Liu, S., Glowatz, M., Zappatore, M., et al. (eds.): e-Learning, e-Education, and Online Training, pp. 1–374. Springer, Heidelberg (2018)
9. Corbellini, A., Mateos, C., Godoy, D., et al.: An architecture and platform for developing distributed recommendation algorithms on large-scale social networks. J. Inf. Sci. **41**(5), 686–704 (2015)
10. Mahboob, V.A., Jalali, M., Jahan, M.V., et al.: Swallow: resource and tag recommender system based on heat diffusion algorithm in social annotation systems. Comput. Intell. **33**(1), 99–118 (2017)
11. Ferreira, L.G., Barbosa, J., Gluz, J.C., et al.: Using learner group profiles for content recommendation in ubiquitous environments. Int. J. Inf. Commun. Technol. Educ. Official Publ. Inf. Resour. Manag. Assoc. **16**(4), 1–19 (2020)
12. Chen, H., Yin, C., Li, R., et al.: Enhanced learning resource recommendation based on online learning style model. Tsinghua Sci. Technol. **25**(3), 348–356 (2020)