Collaborative Filtering of Learning Resources Recommendation Based on Learners' Viewing Behavior

Chong Wang[1], Ziyao Wang[1](✉), Yongzhe Zhao[1], and Yixuan Zhao[2]

[1] School of Business, Guilin University of Electronic Technology, Guilin 450305, China
wchong@guet.edu.cn, wjyo2691@163.com

[2] School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin 450305, China

Abstract. With the popularity of online education and the development of information technology, the number of learning resources on the Internet has increased geometrically, making it often impossible for learners to obtain more accurate learning resource recommendations. The main form of online learning for learners is watching online video courses, making good use of learners' behavioral data can improve the accuracy of recommended resources for them. For this purpose, this research proposes a collaborative filtering learning resource recommendation method based on learners' viewing behavior; Firstly, the learning resource attributes and learners' viewing behavior are mined to build their interest preference model. Secondly, the model is incorporated into the collaborative filtering recommendation algorithm using an improved Pearson similarity calculation method; Finally, the personalized recommendation of learning resources is completed. The experimental results show that the method improves the accuracy and recall rate of personalized learning resource recommendations to a certain extent compared to the traditional collaborative filtering recommendation algorithm.

Keywords: viewing behavior · learning resources · collaborative filtering · personalized recommendation

1 Introduction

Entering the era of Web 2.0, information technology continues to develop, and online education has dramatically changed the way people learn, especially in the epidemic era, when offline education cannot be carried out usually, online education has become the choice of more learners. Data [3] shows that as of June 2021, the scale of online education users in China reached 325 million, accounting for 32.1% of the overall Internet users. Faced with the massive demand for online education and the geometric growth of online shared learning resources, it is difficult for learners to quickly find the resources they need, resulting in 'information overload' and 'information disorientation.'

Y. Zhao and Y. Zhao—Contribute equally.

Z. Zhan et al. (Eds.): EIMSS 2022, AHCS 7, pp. 743–754, 2023.
https://doi.org/10.2991/978-94-6463-024-4_78

Personalized recommendation is an information technology that recommends resources of interest to users based on their preferences [17]. The collaborative filtering algorithm is one of the more mainstream recommendation algorithms currently in use. In the learning resource recommendation scenario, the types of resources can be as diverse as books, articles, audio, and video. The core of the collaborative filtering algorithm is to recommend resources based on the interests of similar users without analyzing the resources' attributes, so it has been widely studied in the field of learning resource recommendation [20].

In the field of learning resource recommendation, researchers have continued to improve the recommendation effectiveness of collaborative filtering algorithms through various methods: Pei-Chann proposed a two-stage user based collaborative filtering approach using an artificial immune system to predict student grades and a professor rating filter in course recommendation for college students [11]. Salehi used attributes of resources and learners and sequential patterns of learner access to resources in the recommendation process, introducing a learner tree (LT) while considering explicit multiple attributes of resources, time-varying multiple preferences of learners, and learner evaluation matrix [12]. Hao introduced relevant knowledge information into the user similarity calculation and rating prediction of traditional collaborative filtering algorithm, which can make the recommendation results match the learning needs of learners [5]. Sun completed the construction of the learner-learning resource evaluation matrix by defining scores for learners' collecting, sharing, and downloading behaviors and their composite behaviors [14]. Ding combined inter-learner trust with a collaborative filtering algorithm to solve the missing data on the scoring matrix for new incoming user data [4]. Wang used dynamic k-nearest neighbor and Slope One algorithm fused with collaborative filtering algorithm to improve personalized recommendation for resources pushing using two-way self-balancing with stage evolution [15]. Hu combines social tags from users' external social networks and labels about learning resources from a self-collaborative filtering system to provide recommendation to users by suggesting learning resources, tutors, or other learners with common interests [6]. Xie applied active and passive ratings to a collaborative filtering algorithm while incorporating social trust mechanisms for learning resource recommendation [19].

However, in the online learning scenario, instructional videos are the core learning content presentation for online courses [16]. At present, research on collaborative filtering algorithm for learning resource recommendation mainly extracts learners' likes, favorites, and comments, compared to which learners' viewing behaviors can better reflect the dynamic changes of learners' interests in learning resources. Based on actual data from online learning platforms, this study analyzes learners' viewing behavior, models learners' interest preferences, and integrates collaborative filtering recommendation algorithm to improve the degree of personalization of resource recommendation for online education.

2 Collaborative Filtering Algorithm Incorporating Learners' Viewing Behavior (CF-VB)

In this research, the interest preference model based on learners' viewing behavior is incorporated into the collaborative filtering recommendation algorithm to establish a personalized learning resource recommendation method, which is improved by introducing a learner behavior log for the cold start problem [18] in the traditional collaborative filtering recommendation algorithm, with the following procedure:

Step1: Based on learner behavior logs and learning resource data, mining the characteristics of the interaction between learners and learning resources.
Step2: Modeling learner interest preferences based on the characteristics of learner interaction with learning resources.
Step3: Based on the learner interest preference model, the weights of each factor are calculated to generate a learner-resource scoring matrix.
Step4: Based on the learner-resource scoring matrix, calculate the inter-learner similarity and find the top k sets $U_k = \{u_1, u_2, \ldots, u_k\}$ of users with the highest similarity to the target user u_i.
Step5: The candidate recommendation set $E_n = \{e_1, e_2, \ldots, e_n\}$ is obtained based on the learning resources learned by users in the set U_k. Then the target user u_i is predicted to prefer the resources in the candidate set.
Step6: Based on the prediction results, Top-N personalized recommendation list are generated.

The algorithm model architecture of this paper is shown in Fig. 1.

2.1 Modeling Learner Interest Preferences

This research's learner interest preference modeling is mainly based on the behavioral data generated during users' viewing of online video courses and the primary attributes of learning resources for the corresponding calculations.

Specifically, suppose there exist M learners, whose constitutive set of learners is $U = \{u_1, u_2, \ldots, u_M\}$; There exist N learning course resources which constitute the set of learning resources $E = \{e_1, e_2, \ldots, e_N\}$; For each learner u_i, there is corresponding historical learning record $B_i^j = \{u_i, e_j, n_i^j, \sigma_i^j, total\_count_i^j, total\_time_i^j, total\_local\_time_i^j\}$, where

B_i^j denotes the historical learning record of learner u_i for learning resource e_j. n_i^j denotes the number of chapters studied by the learner u_i for learning resource e_j. σ_i^j indicates whether learner u_i is subscribed to learning resource e_j. $total\_count_i^j$ denotes the total number of views of the chapter studied in learning resource e_j by learner u_i. $total\_time_i^j$ denotes the total duration of the course that the learner u_i studied for the learning resource e_j. $total\_local\_time_i^j$ denotes the actual local learning duration of learner u_i for learning resource e_j; For each learning resource e_j, there is corresponding learning resource primary attribute $P_j = \{e_j, n_j, t_j\}$, where n_j denotes the total number of chapters of learning resource e_j. t_j denotes the total duration of the course for learning

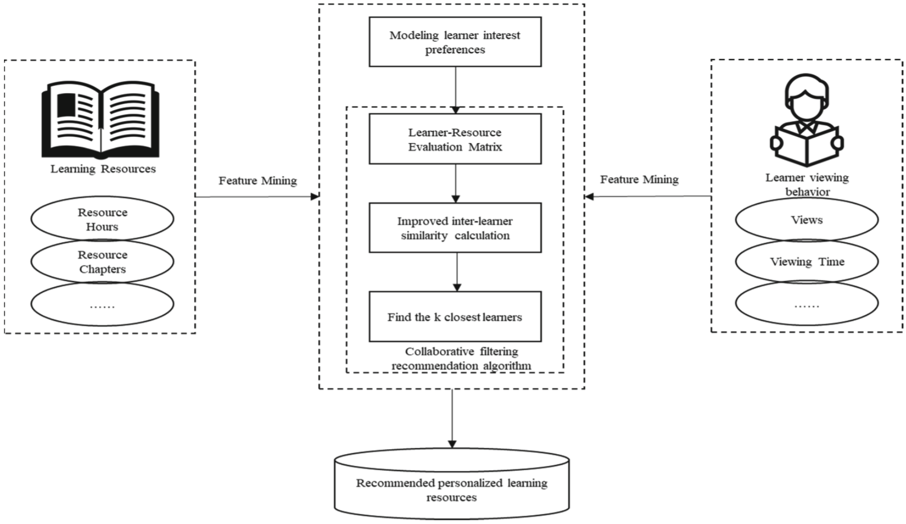

Fig. 1. Collaborative filtering recommendation algorithm model architecture incorporating learners' viewing behavior.

resource e_j. This study uses learners' historical learning records and the primary attributes of learning resources to mine learners' learning interest preferences and build a model of learners' interest preferences.

The traditional approach to modeling user interest preferences considers that user interest preferences for video resources are proportional to the number and length of time users watch the resources [7], and this approach, which only considers a single user behavior, does not make full use of user behavior data and is insufficient to provide feedback on users' valid dynamic interest preferences. Therefore, this research defines five influencing factors that adequately describe learner interest preferences.

Define the subscription factor α based on whether a learner subscribes to a learning resource:

$$\alpha = \begin{cases} 0, \ \sigma_i^j = false \\ 1, \ \sigma_i^j = true \end{cases} \tag{1}$$

where α indicates whether the learner u_i has subscribed to the learning resource e_j, and takes the value of 1 if it has, and 0 if it has not.

Define a frequency impact factor β based on the average number of times a learner views a chapter studied by a learning resource:

$$\beta = \frac{total\_count_i^j}{n_i^j} \tag{2}$$

where β denotes the average number of times learner u_i watched each chapter studied in learning resource e_j. The larger β is, the more interested learner u_i is in learning resource e_j.

Define a time adjustment factor γ based on the average duration a learner spends on a given learning resource:

$$\gamma = \frac{total\_time_i^j}{t_j} \tag{3}$$

where γ denotes the ratio of the total course duration studied by learner u_i for learning resource e_j to the total resource duration of learning resource e_j. The larger γ is, the more interested learner u_i is in learning resource e_j.

Define the frequency impact factor δ based on the number of chapters studied by the learner for a given learning resource:

$$\delta = \frac{n_i^j}{n_j} \tag{4}$$

where δ denotes the ratio of the number of all chapters studied by learner u_i to the total number of chapters of learning resource e_j. The larger δ is, the more interested learner u_i is in learning resource e_j.

Define a rate impact factor θ based on the average viewing rate of learners for a given learning resource:

$$\theta = \frac{total\_local\_time_i^j}{total\_time_i^j} \tag{5}$$

where θ denotes the inverse of the overall viewing rate of learner u_i for learning resource e_j. The larger θ indicates that learner u_i is more interested in learning resource e_j.

After calculating the above five interest preference factors based on learner u_i*'s* viewing behavior for learning resource e_j, the entropy weighting method [2] was introduced to describe the relative importance of each factor due to the different degrees of learner preference for learning resources that each factor can show, resulting in different degrees of contribution to the final rating items. By calculation, the weights corresponding to each factor are shown in Table 1, where ω_x denotes the weight assignment of the corresponding factor.

Based on the above analysis, the final learner interest preference, the learner's rating of the learning resource r_{ij}, can be obtained with the following formula.

$$r_{ij} = \alpha \times \omega_\alpha + \beta \times \omega_\beta + \gamma \times \omega_\gamma + \delta \times \omega_\delta + \theta \times \omega_\theta \tag{6}$$

Table 1. Learner interest preference factor weights.

Influencing Factor	Weight Distribution	Value
α	ω_α	0.211
β	ω_β	0.222
γ	ω_γ	0.160
δ	ω_δ	0.202
θ	ω_θ	0.205

where r_{ij} denotes the rating of learner u_i on learning resource e_j, α, β, γ, δ, and θ fully consider the real learning viewing behavior of learners, which corrects the traditional user interest preference modeling from different perspectives and improves the accuracy of the interest preference model.

2.2 Learner-Resources Rating Matrix

An $M \times N$ order learner-resources rating matrix is created based on the learner's interest preference ratings for the learning resources:

$$\begin{array}{c c} & \begin{array}{cccc} e_1 & e_2 & \cdots & e_n \end{array} \\ \begin{array}{c} u_1 \\ u_2 \\ \vdots \\ u_m \end{array} & \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{pmatrix} \end{array}$$

where the rows denote M learners, the columns denote N learning resources, and the r_{ij} in the matrix represents the rating of resource e_j by learner u_i.

2.3 Similarity Calculation

2.3.1 Traditional Similarity Calculation

In collaborative filtering recommendation algorithm, the choice of the similarity formula has a significant impact on the accuracy of the recommendation results [10]. Cosine similarity uses the cosine of the angle between two vectors to measure similarity, the smaller the angle between two vectors, the higher the similarity [13], this approach is not sensitive to scoring data, it focuses too much on the angle between vectors and ignores the length of the vectors; Jaccard similarity is obtained using the common rating term between users divided by the concurrent set of inter-user rating terms [8], which considers only the number of common ratings of two users, without considering the absolute ratings, and is only applicable to the set represented by Boolean vectors; The Pearson correlation coefficient expresses the similarity of users in terms of linear correlation between vectors and is centralized.

$$Sim(u, v) = \frac{\sum_{k \in I_u \cap I_v} (r_{uk} - \mu_u) * (r_{vk} - \mu_v)}{\sqrt{\sum_{k \in I_u \cap I_v} (r_{uk} - \mu_u)^2} * \sqrt{\sum_{k \in I_u \cap I_v} (r_{vk} - \mu_v)^2}} \tag{7}$$

where r_{uk} denotes learner u's rating of learning resource k, and μ_u denotes the mean of learner u's rating of all resources. Compared with cosine similarity and Jaccard similarity, Pearson similarity is more accurate [22]. Therefore, this research chooses the Pearson correlation coefficient as the basis of similarity calculation.

2.3.2 Improved Pearson Similarity Calculation

It can be seen from Eq. (7), that due to the large sparsity of the scoring matrix in the recommendation system, the existence of learner u and learner v with fewer common scoring items and exactly satisfying the similarity of 1 will lead to fewer common scoring items among the neighborhood learners who are entirely positively correlated with each learner, which eventually leads to inaccurate recommendation results. Therefore, this research will introduce a penalty factor ε for the popular term and a penalty factor η for the common scoring term for correction.

For example, in news and information recommendation systems, almost everyone tunes in when a popular news item appears, and this is also true in the online education field [9]. To reduce the problem of the small contribution of popular resources to the Pearson similarity calculation, a penalty factor ε for popular terms is introduced as follows.

$$\varepsilon = \frac{1}{log(1 + C(e))} \tag{8}$$

where *C(e)* denotes the number of times learning resource *e* is scored in the scoring matrix.

To reduce the impact of fewer learner co-rated items [1] on the Pearson similarity calculation, the introduction of a co-rated item penalty factor η as follows.

$$\eta = \frac{min(|c(u) \cap c(v)|, \varphi)}{\varphi} \tag{9}$$

where $|c(u) \cap c(v)|$ represents the intersection of the learning resources that learner *u* and learner *v* have studied.

The improved Pearson similarity formula with introducing a penalty factor ε for the popular term and a penalty factor η for the common scoring term is as follows.

$$Sim(u, v) = \eta \frac{\sum_{k \in I_u \cap I_v} \varepsilon (r_{uk} - \mu_u)(r_{vk} - \mu_v)}{\sqrt{\sum_{k \in I_u \cap I_v} (r_{uk} - \mu_u)^2} \sqrt{\sum_{k \in I_u \cap I_v} (r_{vk} - \mu_v)^2}} \tag{10}$$

2.4 Recommendation

Based on the set of K-nearest neighborhood learners $S_{uk} = \{u_1, u_2, \ldots, u_k\}$ with the highest similarity to the target learner *u*, the candidate recommendation resource set $E_n = \{e_1, e_2, \ldots, e_n\}$ is obtained. The formula for predicting the target user *u*'s rating of each resource in the candidate recommended resource set E_n is as follows.

$$\widehat{r_{ui}} = \overline{r_u} + \frac{\sum_{v \in S_{uk} \cap N_u(i)} sim(u, v)(r_{vi} - \overline{r_v})}{\sum_{v \in S_{uk} \cap N_u(i)} sim(u, v)} \tag{11}$$

where $\widehat{r_{ui}}$ denotes learner *u*'s predicted rating of learning resource *i*, $\overline{r_u}$ denotes the mean of all resources rated by learner *u*, and $N_{u(i)}$ denotes the set of learners who have ratings for learning resource *i*. Finally, based on the predicted scores of the target learner *u*, the Top-N learning resource recommendation can be completed.

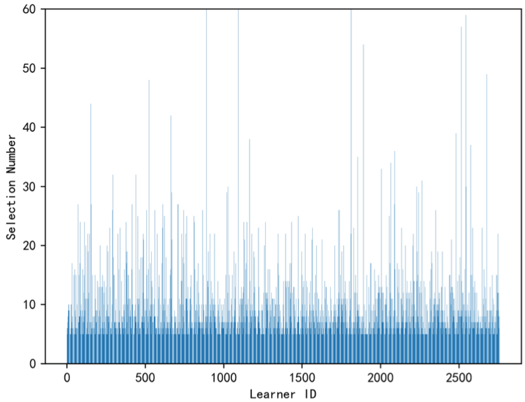

Fig. 2. Distribution of the number of resources selected by learners.

3 Experiments

3.1 Experimental Data

The experimental data used MOOC Cube, an open data repository published by ACL2020 [21], containing 706 real online courses, 38,181 instructional videos, 114,563 concepts, and hundreds of thousands of course selection and video viewing records from 199,199 MOOC users on the XuetangX. From this study, 26,292 courses selection and behavioral information of 2,761 learners were randomly selected as experimental data, which included 480 courses. The corresponding number of course selections by learners is shown in Fig. 2. The experimental data was divided into a training set and a test set in a ratio of 4:1.

3.2 Evaluation Indicators

This research adopts the method of offline measurement. In order to verify the recommendation effect, this paper chooses *Precision* and *Recall* to test the data in the test set. The calculation formula is as follows.

$$Precision = \frac{\sum_{u_i \in U} |R(u_i) \cap T(u_i)|}{\sum_{u \in U} R(u_i)} \tag{12}$$

$$Precision = \frac{\sum_{u_i \in U} |R(u_i) \cap T(u_i)|}{\sum_{u \in U} T(u_i)} \tag{13}$$

where $R(u_i)$ denotes the recommended list of learning resources recommended for learner u_i, and $T(u_i)$ denotes the resources actually learned by learner u_i.

3.3 Experimental Results

This experiment conducts simulation experiments on the algorithm, during which the values of φ in the penalty factor η for the common scoring term are first experimented

Table 2. Influence when φ values change.

φ	Precision	Recall
3	0.212	0.347
4	0.232	0.412
5	0.237	0.431
6	0.220	0.282
7	0.217	0.311
8	0.215	0.288
9	0.209	0.251
10	0.205	0.261

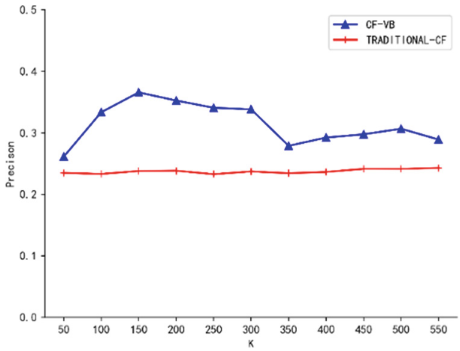

Fig. 3. Comparison of Precision with different numbers of neighbouring users.

with, setting the initial φ to 3, the step size to 1, and the maximum φ value to 10. Table 2 reflects the effect of φ values on the recommendation results.

It can be seen that the experiment achieved better-recommended results when $\varphi = 5$, so in the subsequent experiments of this study, φ will be set to 5.

During the experiments, the Top-N algorithm was used with a recommendation list length of 5 to explore the Precision and Recall of Top K numbers of neighboring users. The collaborative filtering algorithm that incorporates learners' viewing behavior (CF-VB) was compared with the traditional collaborative filtering algorithm (TRADITIONAL-CF) in terms of Precision and Recall in a comprehensive manner. The experimental results are shown in Fig. 3 and Fig. 4.

Precision describes the number of samples whose recommendations are true as a proportion of the length of the recommendation list for predicted results. From Fig. 3, we can see that the Precision of the traditional collaborative filtering recommendation algorithm fluctuates less with the increase of K value, and the optimal value is not obvious. The precision of CF-VB rises and then falls, reaching a maximum around $K = 150$, and is overall higher than that of traditional collaborative filtering algorithms; Recall describes the number of samples for which the recommendation turned out to be

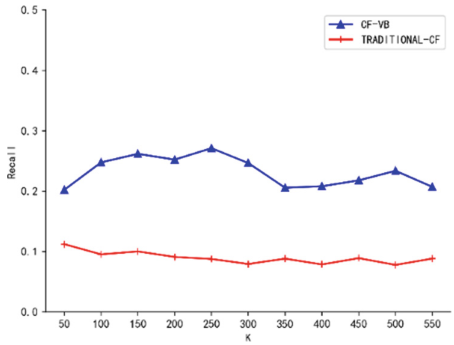

Fig. 4. Comparison of Recall with different numbers of neighbouring users.

true as a proportion of the sample length of the test set for the original sample. As can be seen from Fig. 4, the overall Recall of the algorithm in this paper is higher than that of the traditional collaborative filtering algorithm. Overall, the collaborative filtering recommendation algorithm that incorporates learners' viewing behavior outperforms traditional collaborative filtering recommendation algorithm.

4 Conclusions

Currently, in the field of learning resource recommendation, there is a lack of explicit ratings similar to movie ratings and product reviews, and teaching videos are the core medium for learners to learn online, so it is even more important to use learners' viewing behavior data to make personalized recommendations for this situation. The collaborative filtering recommendation algorithm based on learners' viewing behavior mainly uses behavioral log files to extract implicit expressions of learners' interests through their viewing behavior, and the construction of a learner-learning resource rating matrix is key to the recommendation. The effectiveness of the algorithm has been demonstrated through experimental comparisons, and the accuracy of learning resource recommendation has been improved. However, the algorithm in this study also has certain limitations. In practical applications, the behavioral data must first be obtained before the calculation can be carried out, and further research is needed for the real-time of the recommendation.

Acknowledgements. This research was supported by the following projects: National Natural Science Foundation of China (72061008), Guangxi Natural Science Foundation of China (2018GXNSFAA294123), Guangxi Key Laboratory of Trusted Software (KX201923), and Innovation Project of GUET Graduate Education (2021YCXS093, 2019YCXS066, 2021YCXS068).

References

1. Chen GP, Wang H (2016) A personalized recommendation algorithm on improving Pearson correlation coefficient. J Shandong Agric Univ 47(06):940–944
2. Cheng QY (2010) Structural entropy method for determining the weights of evaluation indicators. Syst Eng Theor Pract 30(07):1225–1228
3. China Internet Network Information Center (CINIC) (2021). The 48th Statistical Report on the Current Status of Internet Development in China
4. Ding YG, Zhang X, Sang QX et al (2016) Recommended collaborative filtering learning resources that integrate learners' social networks. Mod Educ Technol 26(02):108–114
5. Li H, Du F, Zhang M, Wang L, Yu X (2018) Research on collaborative filtering recommendation of learning resource based on knowledge association. In: Ting-Ting Wu, Huang Y-M, Shadieva R, Lin L, Starčič AI (eds) ICITL 2018, vol 11003. LNCS. Springer, Cham, pp 561–567. https://doi.org/10.1007/978-3-319-99737-7_59
6. Hu L, Zhao D, Tong Q, et al (2013) Context-aware recommendation of learning resources using rules engine. In: 2013 IEEE 13th international conference on advanced learning technologies (ICALT). IEEE
7. Kim E, Pyo S, Park E et al (2011) An automatic recommendation scheme of TV program contents for (IP)TV personalization. IEEE Trans Broadcast 57(3):674–684
8. Li B, Zhang B, Liu XJ et al (2016) Collaborative filtering recommendation algorithm based on Jaccard similarity and locational behaviors. Comput Sci 43(12):200–205
9. Li R, Li MQ, Guo WQ (2016) Research on collaborative filtering algorithm based on improved similarity. Comput Sci 43(12):206–208+240
10. Liu F, Tian F, Lin L et al (2021) A collaborative filtering recommendation method for online learning resources incorporating the learner model. CAAI Trans Intell Syst 16(06):1117–1125
11. Pei-Chann C, Cheng-Hui L, Meng-Hui C (2016) A hybrid course recommendation system by integrating collaborative filtering and artificial immune systems. Algorithms 9(3):1–18
12. Salehi M (2013) Application of implicit and explicit attribute based collaborative filtering and BIDE for learning resource recommendation. Data Knowl Eng 87:130–145
13. Schwarz M, Lobur M, Stekh Y (2017). Analysis of the effectiveness of similarity measures for recommender systems. In: Experience of designing & application of cad systems in microelectronics. IEEE
14. Sun X, Wang YG, Qiu FY (2012) Online learning resources based on collaborative filtering technology research on personalized recommendation system. Distance Educ China 08:78–82
15. Wang H, Fu W (2020) Personalized learning resource recommendation method based on dynamic collaborative filtering. Mob Netw Appl 26(1):473–487
16. Wang Y, Zhang JL, Zhang HB (2013) Typical projects of massive open online courses (MOOC) characteristic analysis and insights. J Dist Educ 31(04):67–75
17. Wu ZY, Tang Y, Liu H (2021) Survey of personalized learning recommendation. J Front Comput Sci Technol 16(01):21–40
18. Xiao J, Wang M, Jiang B, et al (2017) A personalized recommendation system with combinational algorithm for online learning. J Ambient
19. Xie XJ, Li XJ, Chen Y et al (2018) Collaborative recommendation algorithm of online learning based on trust-combined similarity model with variable weight. J Chinese Comput Syst 39(03):525–528
20. Yu H, Li JH (2015) Algorithm to solve the cold-start problem in new item recommendations. J Softw 26(6):1395–1408

21. Yu J, Luo G, Xiao T, et al (2020) MOOCCube: a large-scale data repository for NLP applications in MOOCs. In: Proceedings of the 58th annual meeting of the association for computational linguistics
22. Zhuang H, Savage EM (2009) Variation and Pearson correlation coefficients of Warner-Bratzler shear force measurements within broiler breast fillets. Poult 88(1):214–220