Eng Int Syst (2023) 5: 389–394

Engineering
Intelligent Systems

Online Educational Resources of International Chinese Education on MOOC Platform Based on Personalized Learning

Yi Sun*

School of International Culture and Education, Nanjing Normal University, Nanjing, Jiangsu 210097, China
Basic Teaching Department, Chuzhou Polytechnic, Chuzhou, Anhui 239000, China

In this paper, the international Chinese education courses on the Massive Open Online Courses (MOOC) platform are examined, and two recommendation algorithms are analyzed: the user-based collaborative filtering (UserCF) algorithm and the item-based collaborative filtering (ItemCF) algorithm. The UserCF algorithm was improved by the K-means algorithm to obtain a Kmeans-UserCF algorithm. The similarity of users was calculated according to their Chinese level and course preferences. Then, resources were recommended by the K-means-UserCF algorithm. The experimental results showed that compared with UserCF and ItemCF algorithms, the K-means-UserCF algorithm had a higher recall rate (56.64%), accuracy (46.79%), and coverage rate (52.76%) when the number of nearest neighbors was 20. The experimental results verify the reliability of the K-means-UserCF algorithm in recommending online educational resources on the MOOC platform, which is conducive to realizing personalized learning.

Keywords: personalized learning, massive open online courses, international Chinese education, online education resources

1. INTRODUCTION

Due to the open and convenient Internet, online education has broken the restrictions of traditional education in terms of time and space, making it possible for people to learn the content of courses at anytime and anywhere. Online education is rich in resources and extensive in content, which can meet people's individual learning needs better and encourage lifelong learning (Wang et al., 2020). Extensive open online courses (MOOC) and microlectures are becoming increasingly popular (Aparicio et al., 2019). However, online education also has some limitations. In order to find the resources that learners need among the vast amount of online educational materials, studies can be carried out based on personalized recommendation technology. Recommendation technology (Wang, 2018) can find user preferences based on user behavior data, which has been widely used in product recommendation (Putriany et al., 2019) and social media (Rao and Mounika, 2018). Subramaniyaswamy et al. (2018) designed a recommendation method that combined user interests and preferences. This method makes food recommendations based on users' personnel choices and nutritional values. Experimental results showed that this method performed well. Sun et al. (2018) proposed trust similarity for video recommendation, built user interest feature vectors based on users' history and tags, and experiments on Tencent, Weibo, and Youku found that the method effectively improved the recommendation accuracy. Klasnja-Milicevic

*Corresponding address: No. 2188, Fengle Road, Nanqiao District, Chuzhou, Anhui 239000, China. Email: yi24129@163.com

Tab1

Table 1 International Chinese language education resources on the MOOC platform.

Course name	Institution	Course overview
Charming Chinese	Huazhong Agricultural University	Students learn about the characteristics and charms of the Chinese language from the perspective of phonetics and vocabulary; students improve Chinese literacy, and learn about traditional culture.
Chinese UP UP	Wuhan University	Aimed at students with an intermediate or higher level of Chinese; further expands the vocabulary commonly used in the written language, and improves students' ability to use the language.
Functional Chinese crash course	Beijing Language and Culture University	Explains daily Chinese in a sitcom style and helps learners get out of the classroom and adapt to the real language environment.
Comprehensive Chinese for beginners	Zhejiang Institute of Science and Technology	Aimed at international students or ethnic Chinese; helps students consolidate their basic knowledge of Chinese and pass the HSK Level 3 exam successfully.
Chinese language-express to HSK	Shandong University	This course combines the learning of language structure and function with the introduction of Chinese culture, i.e., integrating learning, examination, and use to improve learners' Chinese language ability.
Intermediate Chinese	Shanghai Jiao Tong University	If you have an elementary level of Chinese, join this course and continue to improve your ability to use the world's most spoken language.
Elementary spoken Chinese	Beijing Normal University	Suitable for students learning Chinese from scratch, and learners can master about 150 common vocabulary words and acquire a preliminary knowledge of Chinese characters and culture.

et al. (2018) studied personalized recommendations of e-learning based on learners' needs and knowledge level, and designed a label-based recommendation model that improved execution time while reducing memory requirements. Yang et al. (2019) analyzed privacy protection in personalized recommendations and proposed a PrivRank method that mitigated users' privacy breaches by continuously obfuscating user activity data. The experiments on synthetic and real datasets found that the method effectively protected users' private data while preserving the role of data in recommendations. Zhang (2020) proposed a sparse bilinear convolution-based collaborative filtering recommendation algorithm for recommending personalized dynamic web pages and found, through simulation experiments, that the method performed well in terms of classification and recognition, suggesting good recommendation ability. In this paper, recommendation algorithms for online educational resources were studied, and validated them on a real dataset to demonstrate their effectiveness. This work makes several contributions to achieving reliable recommendations of online educational resources and promoting personalized learning for learners.

2. ONLINE EDUCATION RESOURCE RECOMMENDATION ALGORITHM

Compared with traditional education methods, the MOOC platform has a wider range of educational resources and can be freely chosen by learners, which is conducive to personalized learning, lifelong learning, and independent learning (Xiang, 2021). This paper focuses on the study of international Chinese language education available on the MOOC platform. With the development of globalization, Chinese language education has been attracting increasing attention, and it is convenient for learners to learn Chinese through the MOOC platform. Currently, there are many courses teaching the Chinese language on the MOOC platform, some examples of which are shown in Table 1.

The essence of a recommendation algorithm is to count and analyze users' browsing data, add tags, and recommend content with similar tags to users by calculating the similarity. In online education, in order to achieve personalized learning, recommendation algorithms can be used to find suitable educational resources. Collaborative filtering (CF) is a common recommendation algorithm (Pan et al., 2019), which can be divided into two types.

(1) User-based collaborative filtering algorithm (UserCF) (Wang et al., 2021)

The UserCF algorithm models the user profile, clusters similar users based on their behavioral information, predicts their preferences for items, and builds a user-recommendation resource matrix:

$$R(m,n)=\begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ R_{m1} & R_{m2} & \cdots & R_{mn} \end{bmatrix}, \tag{1}$$

where R_{mn} is the score given to recommended resource n by user m. After obtaining the matrix, the user similarity id calculated. The calculation methods are as follows.

- Jaccard coefficient (Qin et al., 2020): this is obtained by dividing the number of elements in the intersection set by

Table 2 Quantification of user preference courses.

User behavior	Score
Click	1
Favorite	2
Download	2
Learn	3
Favorite + learn	4
Download + learn	5
Favorite+download+learn	6

the number of elements in the union set. Its calculation formula is:

$$w_{xy} = \left| \frac{N_x \cap N_y}{N_x \cup N_y} \right|, \quad (2)$$

where N_x and N_y refer to the sets of users who like x and y.

- Pearson similarity (Zeng et al., 2021): the degree of linear correlation between two variables. It is calculated with:

$$sim(X, Y) = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X})^2 \times \sum_{i=1}^{n} (Y_i - \overline{Y})^2}}, \quad (3)$$

where X_i refers to the score given to resource i by user X and $\overline{X}$ is the average value of scores.

- Cosine similarity (Liu et al., 2018): based on the cosine value of the vector angle. Its calculation formula is:

$$sim(X, Y) = \cos\theta = \frac{\vec{X} \cdot \vec{Y}}{\left\| \vec{X} \right\| \cdot \left\| \vec{Y} \right\|}, \quad (4)$$

where $\overrightarrow{X}$ and $\overrightarrow{Y}$ are vectors of scores given to all resources by users X and Y.

- Euclidean distance (Rahman and Oldford, 2018): based on the Euclidean distance between arbitrary users. Its c formula is:

$$D(x, y) = \sqrt{\left(x^2 - y^2\right)}. \quad (5)$$

For target user u, k users that are the most similar to user u are selected and written as set S(u, k). Items preferred by all the users are selected from S. After deleting items that have been preferred by u, the remaining items are analyzed before recommending to u. For every item that may be recommended, the level of interest of user u in it is calculated with:

$$P_{u,m} = \overline{R_u} + \frac{\sum_{u \in U} (sim(u, v)(R_{um} - \overline{R_v}))}{\sum_{u \in U} sim(u, v)}. \quad (6)$$

Finally, the obtained results are sorted in descending order, and the $top - k$ item is recommended.

(2) Item-based collaborative filtering algorithm (ItemCF) (Kusumawardhani et al., 2019).

The principle of the ItemCF algorithm is to take an item as the subject to find the most similar item and recommend items according to the preferences of similar users. This process is similar to the UserCF algorithm, but the biggest difference is that it calculates the similarity between items.

In this paper, the UserCF algorithm is used as an educational resource recommendation algorithm. In order to reduce the computation and improve the recommendation performance of the algorithm, this paper improves the UserCF algorithm with the K-means algorithm.

The K-means algorithm is a method of clustering (Liu et al., 2021) that can classify users with similar characteristics into one class. The process of the K-means-UserCF algorithm is:

- Users are divided by the K-means algorithm, and the similarity of users in every cluster is calculated.
- The top K users with the greatest similarity are taken as the neighbor set of target user i, and scores given to resources by user j are retrieved.
- Users' scores are predicted, and the top N resources with the highest score are recommended to i.

3. PERSONALIZED LEARNING ANALYSIS FOR INTERNATIONAL CHINESE LANGUAGE EDUCATION LEARNERS

3.1 Data Acquisition

The Python language crawler was used to collect various educational resources relevant to international Chinese education from the Chinese university MOOC platform (https://www.icourse163.org/); the collected content included user information, score information, etc. Finally, 1508 educational resources and 92,345 users were collected. The Chinese language level of users was expressed by HSK level (1-6 levels). Users' preferred courses were expressed by users' attitudes towards educational resources, as shown in Table 2.

3.2 Similarity Calculation

- Chinese level similarity.

  The similarity of the Chinese language level of users X and Y is expressed by $H(X, Y)$, and its formula is written as:

Table 3 Examples of experimental data.

User number	Chinese language level	Course number	Course preference	Resource score
1	1	1	6	5
2	2	2	5	4
3	4	1	6	5
4	3	3	4	3
5	6	1	4	4

$$H(X, Y) = \begin{cases} 1, S_x = S_y \\ 0, S_x \neq S_y \end{cases}, \tag{7}$$

where S_x and S_y denote the Chinese language levels of users X and Y. $H(X, Y) = 1$ when the Chinese language level of two users was the same. $H(X, Y) = 0$ when the Chinese language level of the two users was different.

- Similarity of preferred courses

  The similarity of the preferred courses of users X and Y was expressed by $J(X, Y)$. The set of the preferred courses of users X and Y was $J_X = \{J_{x1}, J_{x2}, \cdots, J_{xm}\}$ and $J_Y = \{J_{y1}, J_{y2}, \cdots, J_{yn}\}$, respectively. In the two sets, the number of the same courses was h, then the calculation formula of $J(X, Y)$ is written as:

$$J(X, Y) = \frac{h}{m + n - h}. \tag{8}$$

The formula for calculating the user similarity considering the Chinese language level and course preference is:

$$sim(X, Y)_{user} = \alpha H(X, Y) + \beta J(X, Y), \tag{9}$$

where α and β are weights, $\alpha + \beta = 1$.

Then, the score similarity was calculated. The user-score matrix ($V(m, n)$) and course-user matrix $U(m, n)$ are written as:

$$V(m, n) = \begin{bmatrix} V_{1,1} & V_{1,2} & \cdots & V_{1,n} \\ \cdots & \cdots & \cdots & \cdots \\ V_{m,1} & V_{m,2} & \cdots & V_{m,n} \end{bmatrix}, \tag{10}$$

$$U(m, n) = \begin{bmatrix} User_1 & User_2 & \cdots & User_n \\ \cdots & \cdots & \cdots & \cdots \\ User_1 & User_2 & \cdots & User_n \end{bmatrix}. \tag{11}$$

Based on the cosine similarity, the score similarity of users ($sim(X, Y)_{score}$) was calculated. The final similarity is written as:

$$sim(X, Y) = sim(X, Y)_{user} + sim(X, Y)_{\text{score}}. \tag{12}$$

Finally, the educational resources were recommended by the K-means-UserCF algorithm.

3.3 Analysis of Results

Eighty percent of the collected dataset was used for training and 20% for testing. Some of the experimental data are shown in Table 3.

Based on the experimental data, user similarity and score similarity were calculated using the above method, and then the recommendation algorithm was used to recommend online educational resources. Suppose that the set of resources recommended by the algorithm was $R(u)$ and the actual data set contained in the test set was $A(u)$. The performance of the algorithm was evaluated using the following indicators:

(1) recall rate: $Recall = \frac{\sum_{u \in U} |R(u) \cap A(u)|}{\sum_{u \in U} |A(u)|}$,

(2) precision: $Precision = \frac{\sum_{u \in U} |R(u) \cap A(u)|}{\sum_{u \in U} |R(u)|}$,

(3) coverage rate: $Coverage = \frac{U_{u \in U} R(u)}{|I|}$.

The performance of UserCF, ItemCF, and K-means-UserCF algorithms was compared. The comparison of the recall rate between different algorithms under different numbers of nearest neighbors is shown in Figure 1.

Figure 1 shows that the recall rate of different algorithms increased gradually with the increase of the number of nearest neighbors. Taking the UserCF algorithm as an example, its recall rate was 26.77% when K=5 and 41.26% when K=20, showing an increase of 14.49%. The comparison of different algorithms suggested that the recall rate of the ItemCF algorithm was the lowest, followed by UserCF and K-means-UserCF algorithms. When $K = 5$, the recall rate of the three algorithms was 26.77%, 25.36%, and 34.78%, respectively, the recall rate of the K-means-UserCF algorithm was 9.42% higher than the ItemCF algorithm and 8.01% higher than the UserCF algorithm. These results verified the reliability of the K-means-UserCF algorithm in online resource recommendations.

A comparison of the accuracy rate between different algorithms is shown in Figure 2.

It was seen from Figure 2 that, similar to the recall rate, the accuracy of different algorithms all showed an increase with the increase of the number of nearest neighbors. Taking the UserCF algorithm as an example, its accuracy was 35.77% when $K = 5$ and 41.12% when $K = 20$, indicating an increase of 5.35%. The comparison of different algorithms demonstrated that the accuracy of the ItemCF algorithm was the lowest and the accuracy of the K-means-UserCF algorithm was the highest. When $K = 20$, the accuracy of the K-means-User algorithm was 46.79%, which was 5.67% higher than the UserCF algorithm and 7.12% higher than the ItemCF algorithm.

Finally, a comparison of the coverage rate between different algorithms is shown in Figure 3.

It was seen from Figure 3 that when $K = 5$, the coverage rates of the three algorithms were 36.77%, 35.78%, and

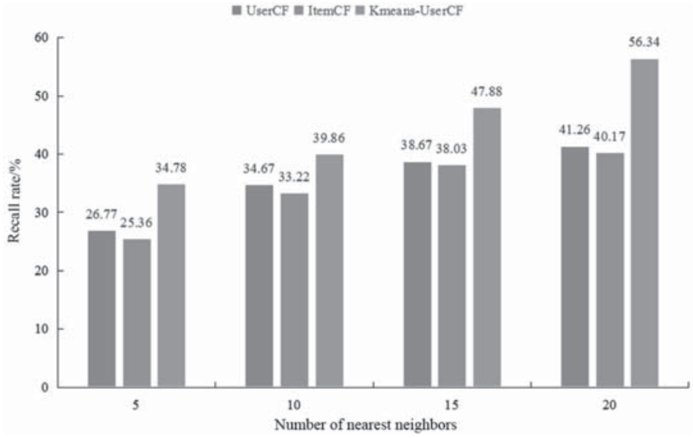

Figure 1 Comparison of the recall rate between different algorithms.

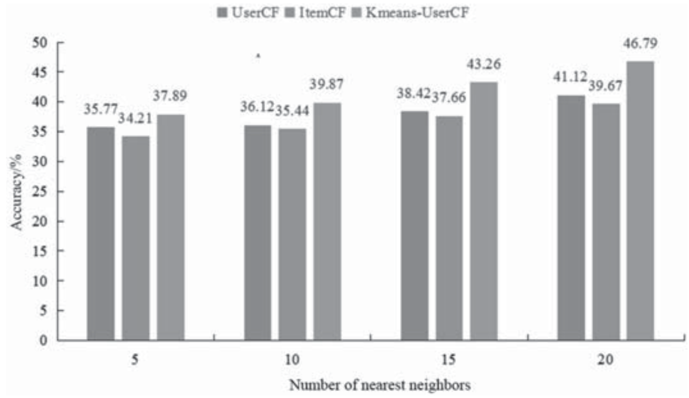

Figure 2 Comparison of the accuracy rate between different algorithms.

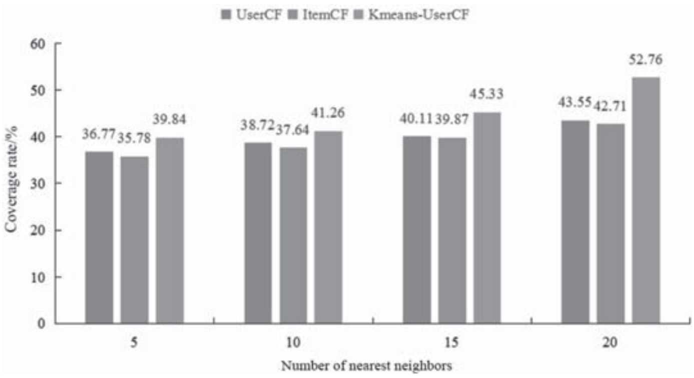

Figure 3 Comparison of the coverage rate between different algorithms.

39.84%, respectively, and the coverage rate of the K-means-UserCF algorithm was 3.07% and 4.06% higher than UserCF and ItemCF algorithms, respectively; when $K = 20$, the coverage rate of the K-means-UserCF algorithm was 52.76%, which showed an improvement of 12.92% compared to that when $K = 5$.

4. CONCLUSION

In this paper, an algorithm was designed for the purpose of recommending online educational resources on the MOOC platform. Based on learners' Chinese language level and course preferences, a collaborative filtering algorithm combined with the K-means algorithm was used to recommend these resources. The experiment found that the K-means-UserCF algorithm had better recommendation performance than the traditional UserCF and ItemCF algorithms. The K-means-UserCF algorithm can be further promoted and applied in practical personalized learning.

REFERENCES

1. Aparicio, M., Oliveira, T., Bação, F., & Painho, M. (2019). Gamification: A Key Determinant of Massive Open Online

Course (MOOC) Success. *Information & Management*, *56*(1), 39–54.

2. Klašnja-MIlićević, A., Ivanović, M., Vesin, B., & Budimac, Z. (2018). Enhancing e-learning systems with personalized recommendation based on collaborative tagging techniques. *Applied Intelligence*, *48*, 1519–1535.
3. Kusumawardhani, N.K., Nasrun, M., & Setianingsih, C. (2019). Web Recommended System Library Book Selection Using Item Based Collaborative Filtering Method. *2019 IEEE International Conference on Engineering, Technology and Education (TALE)*, 1–8.
4. Liu, D., Chen, X., & Dan, P. (2018). The Intuitionistic Fuzzy Linguistic Cosine Similarity Measure and Its Application in Pattern Recognition. *Complexity*, *2018*, 1–11.
5. Liu, X., Sang, X., Chang, J., Zheng, Y., & Han, Y. (2021). The water supply association analysis method in Shenzhen based on k-means clustering discretization and a priori algorithm. *PLoS ONE*, *16*(8), 1–21.
6. Pan, W., Chen, L., & Ming, Z. (2019). Personalized recommendation with implicit feedback via learning pairwise preferences over item-sets. *Knowledge & Information Systems*, *58*, 295–318.
7. Putriany, V., Jauhari, J., & Heroza, R.I. (2019). Item Clustering as An Input for Skin Care Product Recommended System using Content Based Filtering. *Journal of Physics: Conference Series*, *1196*, 1–6.
8. Qin, X., Jiang, J., Yuan, C.A., Qiao, S.J., & Fan, W. (2020). Arbitrary shape natural scene text detection method based on soft attention mechanism and dilated convolution. *IEEE Access*, *8*, 122685–122694.
9. Rahman, A., & Oldford, R.W. (2018). Euclidean distance matrix completion and point configurations from the minimal spanning tree. *Siam Journal on Optimization*, *28*(1), 528–550.
10. Rao, P., & Mounika, A. (2018). Recommended system of collaborative filtering based in social networks. *Neural, Parallel & Scientific Computations*, *26*(3–4), 247–256.
11. Subramaniyaswamy, V., Manogaran, G., Logesh, R., Vijayakumar, V., Chilamkurti, N., Malathi, D., & Senthilselvan, N. (2018). An ontology-driven personalized food recommendation in IoT-based healthcare system. *The Journal of Supercomputing*, *75*(2), 3184–3216.
12. Sun, P.C., Yin, S., Zhang, Y.P., & Tan T. (2018). Research on Personalized Recommendation Algorithm Based on Time Weighted and Sparse Space Clustering. *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)*, 153–156.
13. Wang, H., Shen, Z., Jiang, S., Sun, G., & Zhang, R.J. (2021). User-based Collaborative Filtering Algorithm Design and Implementation. *Journal of Physics: Conference Series*, *1757*(1), 1–6.
14. Wang, W. (2018). Recommended system of application and development. *Advances in Materials, zmachinery, Electronics II: Proceedings of the 2nd International Conference on Advances in Materials, Machinery, Electronics (AMME 2018)*, 040180.
15. Wang, W., Zhang, P., & Zeng, C.Z. (2020). Construction of Learning Behaviour Tracking Analysis Model for MOOC Online Education Platform. *International Journal of Continuing Engineering Education and Life-Long Learning*, *30*(1), 1.
16. Xiang, L. (2021). Lifelong Learning: A Study of College English Autonomous Learning. *Engineering Intelligent Systems*, *29*(2), 103–108.
17. Yang, D., Qu, B., & Cudré-Mauroux, P. (2019). Privacy-Preserving Social Media Data Publishing for Personalized Ranking-Based Recommendation. *IEEE Transactions on Knowledge and Data Engineering*, *31*(3), 507–520.
18. Zeng, W., Lu, T., Liu, Z., Xu, Q., Peng, H., Li, C., Yang S., & Yao, F. (2021). Research on a laser ultrasonic visualization detection method for human skin tumors based on Pearson correlation coefficient. *Optics & Laser Technology*, *141*, 107117-.
19. Zhang, X. (2020). Collaborative Filtering Recommendation Algorithm Based on Sparse Bilinear Convolution. *Engineering Intelligent Systems*, *28*(4), 205–214.