



# Relevance Analysis of Online Course Development on Education Platform and Significance of Teaching Reform Research Based on Collaborative Filtering Algorithm

Yuan-Nong Ye<sup>1,2</sup>(✉), Min Wang<sup>2</sup>, Shuai Jin<sup>2</sup>, and Meng-Ya Huang<sup>2</sup>

<sup>1</sup> Key Laboratory of Infectious Immune and Antibody Engineering of Guizhou Province, School of Big Health, Guizhou Medical University, Guiyang, China

yyn@gmc.edu.cn

<sup>2</sup> Department of Medical Informatics, School of Big Health, Guizhou Medical University, Guiyang, China

{jssss, huangmengya}@gmc.edu.cn

**Abstract.** Based on the rapid development of internet and education, online education platforms have also followed the pace of the times and achieved great success, among which Rain Classroom, Tencent Classroom, Learning Pass, MOOC and other platforms are almost mandatory for educational institutions to use during the epidemic prevention and control period, combined with QQ and WeChat for data preparation and schedule notification arrangement, achieving the effect of both epidemic prevention and teaching. With the prevention and control of the epidemic in China, the normal offline schooling has now started nationwide, and the active time of the users of the education platform has decreased. Data mining, data analysis, and the calculation of the user attrition rate of the education platform and making course correlation recommendations have played a decision-making role in promoting the development of the education platform model and the reform of the education model at this stage. Method: The collaborative filtering algorithm of the base model is based on the information of user preferences, trained into a complete model, and then based on the information of user preferences to make predictive recommendations and define the user churn rate calculation method. Results: We obtained the thermal distribution of user logins in each region calculated the percentage of user churn, the binary relationship between users and courses, and the ten most searched and welcomed courses. Conclusion: Based on the similarity of behavior between users and courses, there is a potential connection between different courses.

**Keywords:** Medical Big Data Mining · Scientific Research Training · Undergraduates · Collaborative Filtering Algorithm

## 1 Introduction

With the development of Internet + mode, the demand for quality learning resources has increased [2]. The construction and sharing of learning resources present a new

development trend based on the online education platform of teaching mode [5]. Various online courses, MOOCs and live streaming courses have emerged in an endless stream, and various online education platforms and learning applications have emerged [7]. In particular, in the spring semester of 2020, due to the impact of COVID-19, the online platform became an important platform to display the achievements. Therefore, how to provide accurate course recommendation services for education platforms and users through data analysis according to online user information and learning information of education platforms has become a hot issue in online education [4].

At present, foreign countries are trying a more scientific online quality-oriented education platform model. Online preschool education has unique strategic value. Although the market value in the short term is not clear, it is of great significance in cultivating children's habit of using the upper education system and improving parents' confidence in online education [1].

There is significant significance of online courses: 1) Let high-quality teacher resources not restricted by region; 2) Let students acquire knowledge without geographical restrictions. In the big data mode, fast development is its outstanding feature that cannot be ignored. Mobile Internet is surging in, the rapid development of mobile Internet is changing the world, online education and learning platform to adapt to the development of The Times. Become the current urgent to solve the problem, but also a great opportunity for enterprise development. In the face of opportunities and challenges, failing to make corresponding strategies to promote the development of online education system will surely fall under the tide of The Times. It is a problem that enterprises must pay attention to make relevant recommendations by analysing the courses favoured by users, and to make personalized development plans according to users' needs and the proportion of users' loss [6].

## 2 Methods and Materials

### 2.1 Analyse Material

This paper selects online course operation data of an education platform from September 2018 to June 2020, and provides data for its subject. According to the platform data, Python language and related package analysis tools are selected to complete data analysis and write a report.

### 2.2 Method-The Cooperative Filtering Algorithm

Neighbourhood-based algorithm is the most basic algorithm in recommendation system, which has been studied deeply in academia and widely used in the industry [8]. Neighbourhood based algorithms can be divided into two categories: - category is user-based collaborative filtering algorithm, and alternative is item-based collaborative filtering algorithm. When we talk about collaborative filtering, it basically means user-based or item-based collaborative filtering algorithms. The degree of user  $U$ 's interest in item  $J$  is computed by the following equal.

$$p(u, j) = \sum_{i \in N(u) \cap S(j, k)} w_{ij} r_{ui} \quad (1)$$

where  $N(u)$  is the set of items favoured by user  $U$ ,  $S(j, k)$  is the set of  $k$  items most similar to item  $J$ .  $w_{ij}$  is the similarity between item  $i$  and  $j$ .  $r_{ui}$  is user  $U$ 's interest in item  $i$ . The basic idea is that if you want to calculate the interest degree of user  $U$  and item  $J$ , first find  $k$  items most similar to  $j$ , then look at the interest degree of user  $U$  and these items, and then get the interest degree of user  $U$  in item  $J$  by weighting.

Based model of collaborative filtering algorithm is based on the user's preference information, training into a complete model, and then the foundation information of user preferences to predict recommendation [3], make full use of the collective opinion, namely in the decision-making behaviour of large crowds and choose the answer to collect key information, in order to help us to get the general population statistical significance on the judgment of the conclusion, The recommendation has two characteristics: (1) users with similar interests may be interested in things with the same characteristics; (2) Users may prefer things they have already purchased, such as the same brand or similar manufacturer, considering their historical relationship. It is recommended that people's behaviour is influenced by their subjective consciousness. User-based collaborative filtering algorithm (UserCF). The basic idea of user-based collaborative filtering algorithm is as follows: in the recommendation system of an online personalized education platform, when user  $A$  needs personalized recommendation, he can first find other users with similar interests. Then recommend the courses that user  $A$  likes but has never heard of USER  $A$ . As can be seen from the above description, the user-based collaborative filtering algorithm mainly includes two steps. Step 1: Find a user set with similar interests to the target user. Step 2: Find the courses that the users in this collection like and that the target users have not heard of and recommend to the target users. Here, the key of Step 1 is to calculate the similarity of interest of two users, and the collaborative filtering algorithm mainly uses the similarity of behaviour to calculate the similarity of interest. Given user  $A$  and user  $V$ , let  $N(A)$  represent the course set that user  $A$  has ever had positive feedback, and let  $N(v)$  be the item set that user  $V$  has ever had positive feedback.

## 2.3 Data Pre-processing and Data Analysis

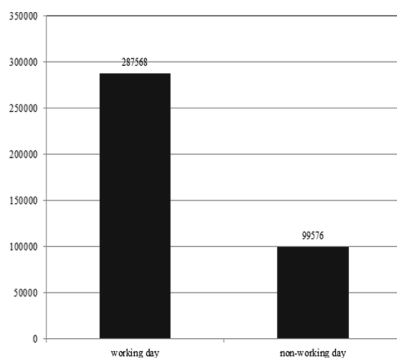
### 2.3.1 Data Pre-processing

This paper mainly filters duplicate values, deletes or fills missing values, and deletes "-" fields. For the data of low importance, we directly delete them. If the data is of high value, we use interpolation or mean complement. Understand the meaning of each field and carry out necessary processing of missing value, repeated value and other aspects. Process the "-" value of the recently logged field in the user information table wedge is non-objective, which makes it very boring to learn. Therefore, students' enthusiasm for this course is weakened, leading to the weak willingness and ability of independent learning.

Compared with the original data, understand the meaning of each field, conduct necessary processing of missing values, duplicate values and other aspects, and understand and sort out the data. The data pre-treatment results are shown in Table 1.

**Table 1.** This caption has one line so it is centered.

Number	City	Login Times
0	None	99518
1	GUANGZHOU	27626
2	SHANTOU	10146
3	SHENZHEN	9098
4	HUIZHOU	6557
5	WUHAN	6534
...	...	...



**Fig. 1.** Frame chart of working day and non-working day.

### 2.3.2 Data Analysis

Platform user activity analysis is divided into the following four steps:

- (1) Draw the thermal map of platform login times in each province.
- (2) Draw the bar chart of user login times of working days and non-working days respectively.
- (3) Remember Tend as the cut-off time of data observation window
- (4) Analyse the activity of platform users and provide suggestions for online management decisions of the education platform.

After data parsing and pre-processing, we used Pyecharts to draw the user login times thermal map of 34 provincial administrative regions in China, the following regions and an unnamed location (the login location is China). First, we imported login.csv file. Shengfen. CSV, chengshi. CSV. Next, we import the Pyecharts library and read the data to draw the required related classes.

Histogram charts of user login times in working day and non-working day are drawn according to the user active time period, and the main user active time period is analysed. Data source is login.csv. After importing the corresponding package, open the file

login.CSV and generate the results directly through matplotlib. The result generated by Matplotlib is shown in Fig. 1.

### 3 Results and Discussion

#### 3.1 Churn Rate of Platform Users

Platform user churn rate tends to indicate the end time of the data observation window (for example, the data collection deadline is June 18, 2020),  $T_i$  indicates the latest access time of user  $i$ , and if the time is greater than 90 days, user  $i$  is lost. The churn rate of platform users is calculated based on the following definition.

$$\Theta_i = T_{\text{end}} - T_i \quad (2)$$

After opening the login.CSV file in pandas, the user.CSV file is calculated by the code. The user loss rate is 56.97%.

#### 3.2 The Popularity of the Course

Record of user participation in learning, count the number of participants of each course, list the top 10 most popular courses, and draw the corresponding bar chart. According to the situation of course selection by users, construct the relationship table between users and courses (binary matrix),

The collaborative filtering algorithm is used to calculate the similarity between courses, and recommend a course for the five users with the highest total learning progress based on the records of selected courses. According to the records of user participation in learning, count the number of participants in each course, calculate the popularity of each course, list the top 10 most popular courses, and draw the corresponding bar chart. Popularity is defined as follows:

$$r_i = \frac{Q_i - Q_{\min}}{Q_{\max} - Q_{\min}} \quad (3)$$

where,  $r_i$  is the popularity of the  $r_i$  course,  $Q_i$  is the number of participants in the study of the  $Q_i$  course,  $Q_{\max}$  and  $Q_{\min}$  are the number of participants in the course with the most and least participants in all courses respectively. The popularity of the top 10 courses is obtained through data processing and visualization of the study\_information.CSV file. The following figure shows the top 10 most popular courses, as shown in Fig. 2.

#### 3.3 Build a Relational Table of Users and Courses

According to the courses selected by users, the relationship table between users and courses (binary matrix) is constructed, the similarity between courses is calculated by using object-based collaborative filtering algorithm, and a course is recommended for the 5 users with the highest total learning progress combined with the records of the courses selected by users. The relationship table between each user and the course (binary matrix) is shown in the file task3\_2.csv, the results are shown in Fig. 3. The result with the highest similarity between the user and the course is 1. Then the course is recommended by the algorithm for the user.

### 3.4 The Comprehensive Recommendation Strategy for the Course

The comprehensive recommendation strategy of the course combines the user’s learning progress data on the basis of the popularity of the course search and the construction of the relationship table between users and the course. Based on the similarity between users and courses, there is a potential connection between different courses. Different users tend to choose or purchase the same course in a large probability. Therefore, two recommendation strategies can be made according to the current situation. Several courses that are often watched at the same time can be pushed by message, or they can be directly associated under a certain course ranking. Bundle and recommend some highly viewed free courses and paid courses, which can increase the sales of paid courses through users’ interest in learning and discount sale.

Based on the above research, it can be concluded that the course selection behaviours of some users are similar. In order to embark on a journey in this era of big data, enterprises need more accurate recommendation algorithms to study the attention and relationship between users and make recommendations based on their subjective consumption psychology.

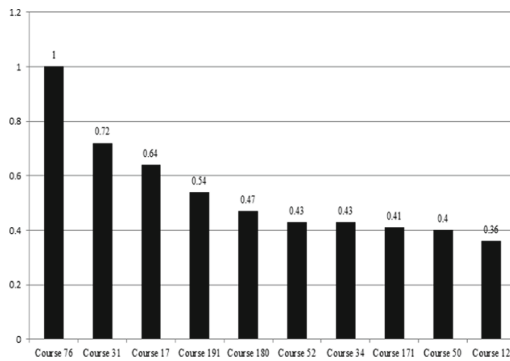


Fig. 2. Chart of the top 10 most popular courses.

	USER a	USER b	USER c	USER d	.....
Course A	0	0	0	0	0
Course B	0	0	0	0	0
Course C	0	0	1	0	0
Course D	0	0	0	0	0
Course E	0	0	0	0	0
.....	0	0	0	0	0

Fig. 3. Recommended course chart.

## 4 Conclusions

The characteristics of online teaching determine that the class system established on weekdays is facing impact. The lack of supervision and communication in the online atmosphere makes it difficult to accurately implement class management. At this time, the class teacher needs to continue the class management system, not only to carefully follow the classroom discipline, but also adhere to the class representative responsibility system and other offline systems, to ensure that the class order is consistent. In addition, besides classroom teaching, mental health content should be added to eliminate students' tension and anxiety by popularizing the knowledge of PREVENTION and treatment of COVID-19, so as to ensure that students can carry out online learning in a positive and relaxed state of mind.

This research is an information project for online education services under the influence of the Internet, aiming to develop a comprehensive online course recommendation strategy through data analysis of user information and learning information of online education platform, so as to provide accurate course recommendation service for education platform and users. Through the source data provided by the education platform, three parts are analysed, namely, data pre-processing, user activity analysis and online course recommendation. At the same time, according to the requirements of each analysis through Python data processing and data visualization to get the corresponding analysis results; each part is treated as described above. Through a series of processing, we get good results, and build an intelligent recommendation model of courses, which provides a basis for the development strategy of online recommendation service of education platform.

**Acknowledgements.** The authors are grateful to anonymous reviewers for their valuable suggestions and comments, which have led to the improvement of this paper. And we thank the funds of Teaching Reform Research Project of Education Steering Committee of Electronic and Information Majors of Higher Education, Ministry of Education (2021-JG-21); Association of Fundamental Computing Education in Chinese Universities Research Project (2021-AFCEC-188); Guizhou Medical University Teaching Content and Curriculum System Reform Project (0620302050136); Ministry of Education industry-University Cooperative Education Program (202101018013); Curriculum Ideology and Politics Construction Project of Guizhou Medical University (Biostatistical analysis, No. 39).

## References

1. Eichhorn, S., and G.W. Matkin. 2016. Massive open online courses, big data, and education research. *New Directions for Institutional Research* 2015(167):27–40.
2. Ellaway, R.H., M.V. Pusic, R.M. Galbraith, et al. 2014. Developing the role of big data and analytics in health professional education. *Medical Teacher* 36 (3): 216–222.
3. Jing, W., D. Tong, Y. Wang, et al. 2017. MaMR: High performance MapReduce programming model for material cloud application. *Computer Physics Communications* 211: 79–87.
4. Lane, J. 2016. Big data: The role of education and training. *Journal of Policy Analysis and Management* 35 (3): 722–724.

5. Li, X. 2019. History and status quo of the reform of the multicultural curriculum in American primary and secondary schools. *Elementary & Secondary Schooling Abroad* 313(01): 62+73–81.
6. Song, Y., Wang, H., Li, J. and H. Gao. 2016. MapReduce for big data analysis: Benefits, limitations and extensions. *ICYCSEE* 2016(1): 453–457.
7. Wang, Y.H. 2020. Design-based research on integrating learning technology tools into higher education classes to achieve active learning. *Computers & Education* 2020.
8. Yin, C., L. Shi, R. Sun, and J. Wang. 2019. Improved collaborative filtering recommendation algorithm based on differential privacy protection. *The Journal of Supercomputing* 76 (7): 5161–5174. <https://doi.org/10.1007/s11227-019-02751-7>.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

