ISSN 2522-6398 Vol. 4, Issue 16: 24-30, DOI: 10.25236/FER.2021.041605

An Intelligent Recommendation Strategy for Online Courses Based on Collaborative Filtering Algorithm for Educational Platforms

Wenjun Li[1, *], Yating Peng[2]

[1]School of Computer Science and Engineering, Changchun University of Technology, Changchun, 130102, China
[2]School of Information and Computer Engineering, Northeast Forestry University, Harbin, 150006, China
**Corresponding author: liwenjun1129@163.com.*

Abstract: *In the context of the new crown epidemic, online education courses have become mainstream. Along with the rapid development of the Internet and the rich reserves of educational resources, how to accurately understand the state of students to provide the required teaching courses, and how to better adapt students to online education to meet the differentiation between each other has become a mainstream problem and challenge at present. Analyze the user data, obtain the decision attributes therein, and explore the classification of online education courses; based on collaborative filtering algorithm, realize the accurate matching between students and courses. The intelligent recommendation strategy of online courses not only helps to improve the usage rate of online courses, but also helps to improve students' individuality, responding to the general policy of "stopping classes and not studying" in major universities, and carrying out more flexible teaching methods.*

Keywords: *Internet+Education, Collaborative Filtering Algorithm, Intelligent Recommendation*

1. Introduction

With the development of technology, the diversity of education has taken a new trend. Various educational platforms and online courses have flooded into the market as well as among various applications. Especially during the epidemic, offline teaching has been affected to a considerable extent not only at home but even overseas. However, the use of online courses was not satisfactory from country to country. Under the national policy of "no classes, no school", universities have launched online courses, and the online platform has become an important support for "Internet + education".

According to the user information of various education platforms, how to grasp the user's preference for courses and accurately provide customers with the required online courses has become an imminent topic. At present, most schools use large classes and open classes for collective teaching, which has the problems of single form, poor targeting, lack of synergistic effect, and inability to form personalized collaborative education mechanism, etc. Based on the above shortcomings, the course recommendation system of collaborative filtering algorithm[1] is promoted in combination with the demand of education informatization under the new situation, and by using collaborative filtering algorithm and introducing changes based on user characteristics, it better It solves the disadvantages of low efficiency, weak adaptability and novelty of course recommendation system[2]. It is of positive significance to create a collaborative, diverse and innovative educational atmosphere for teachers and students[3]. Scholars have also launched a series of research on online teaching, Ke Xiuwen[5] has built an intelligent strategy for online course recommendation from the perspective of intelligent algorithm; Chen Yangxue[6] has analyzed the data of mu class software and studied the recommendation algorithm; Zhao Quan[7] has analyzed the systematic research of online intelligent course in the context of big data.

Therefore, referring to the existing literature to find the right approach to respond, this article uses big data analysis techniques to filter and integrate online information and user information from educational platforms and uses collaborative filtering algorithms to find the best recommendation strategy for online courses.

ISSN 2522-6398 Vol. 4, Issue 16: 24-30, DOI: 10.25236/FER.2021.041605

2. Data pre-processing

2.1. Data Introduction

Based on the last two years of operation data of the education platform, the statistics are synthesized into three csv tables, namely users.csv (user information table), study_information.csv (study information table) and login.csv (login information table). The decision making data attributes are learn_time and learn_process these two data attribute columns are the most direct reflection of the user's satisfaction with the online course. The data column attributes in the table are shown below.

Table 1: Users.csv description

Field Name	description
user_id	User id
registration_time	Registration time
recently_logged	Time of last visit
learn_time	Length of study
number_of_classes_join	Number of classes joined
number_of_classes_out	Number of classes withdrawn
school	School to which the user belongs

Table 2: Study_information.csv description

Field Name	description
user_id	User id
course_id	Course id
course_join_time	Join the course time
learn_process	Learning Progress
price	Course Unit Price

Table 3: Login.csv description

Field Name	description
user_id	User id
login_time	Login time
login_price	Login Address

2.2. Data processing

After statistical analysis it can be concluded that there are missing, anomalous and duplicate data in the original three tables. These abnormal data need to be processed before studying the problem. The anomalous data of the three tables are handled through the info function in pandas, for example study_info.info as follows.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 194974 entries, 0 to 194973
Data columns (total 5 columns):
user_id             194974 non-null object
course_id           194974 non-null object
course_join_time    194974 non-null object
learn_process       194974 non-null object
price               190736 non-null float64
dtypes: float64(1), object(4)
memory usage: 7.4+ MB
```

Figure 1: Learning information table processing data results

After the information processing, the original table is processed as follows.

(1) In the users.csv table, delete the records with empty user_id. For user_id duplicate data value delete, only one is kept. Delete the whole column of school. (b) Convert the "_" value in recently_time to registration time.

(2) In the study_information.csv table, PRICE has a null value related to free courses known in reality,

ISSN 2522-6398 Vol. 4, Issue 16: 24-30, DOI: 10.25236/FER.2021.041605

so the record is 0.

3. The platform user activity analysis

3.1. Mapping of provinces, cities heat map

Construct variables for each province, construct variables for each city and replace the province names with city names for municipalities and administrative regions.

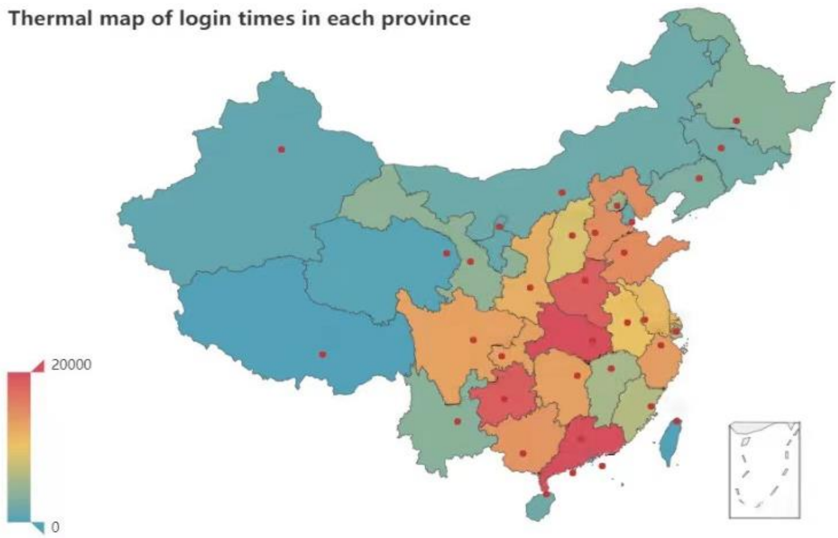

Figure 2: Heat map of the number of landings by province

The color reflects the frequency of the number of logins in each province, and the color tends to be red, indicating that the more logins in the province. Figure 2 shows that the provinces with the highest number of logins are Guangdong Province, followed by Hubei Province, and Central and Southern China, where the level of online education is much greater than that of the Northeast and Western regions.

Figure 3: Heat map of the number of logins by city

The color reflects the frequency of landings in each city, with the color tending to be red indicating that the city has logged in more often. Figure 3 shows that the city with the highest number of logins is Guangzhou, followed by Chongqing, and four of the top five cities are from Guangdong Province. This proves that Guangdong is the leading city in the country in terms of education level of users.

ISSN 2522-6398 Vol. 4, Issue 16: 24-30, DOI: 10.25236/FER.2021.041605

3.2. Histogram of weekday vs. non-weekday user logins

3.2.1. Workday view results and analysis

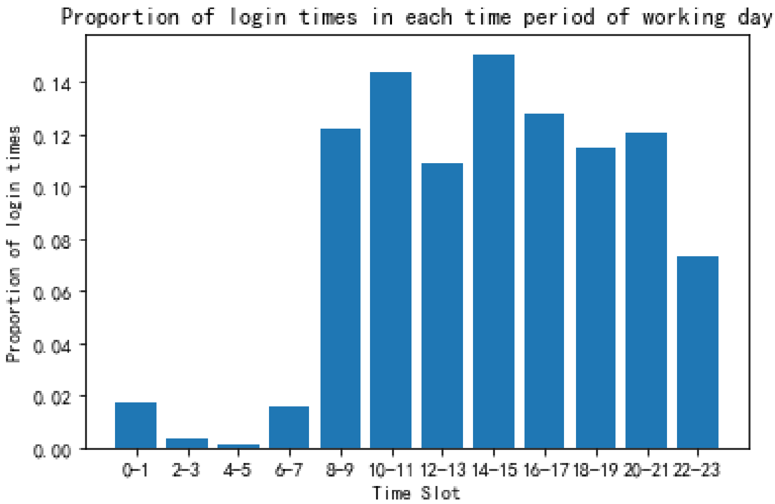

Figure 4: Visualization results for logging in at various times during the working day

From the above graph, it can be seen that during weekdays, users log in between 8:00 and 23:00, and peak at 14:00-15:00 as the main active time period for users. The distribution probabilities are not very different from each other. Before 0:00-7:00, users are basically asleep and few people choose this time to log into the education platform.

3.2.2. Non-working day view results and analysis

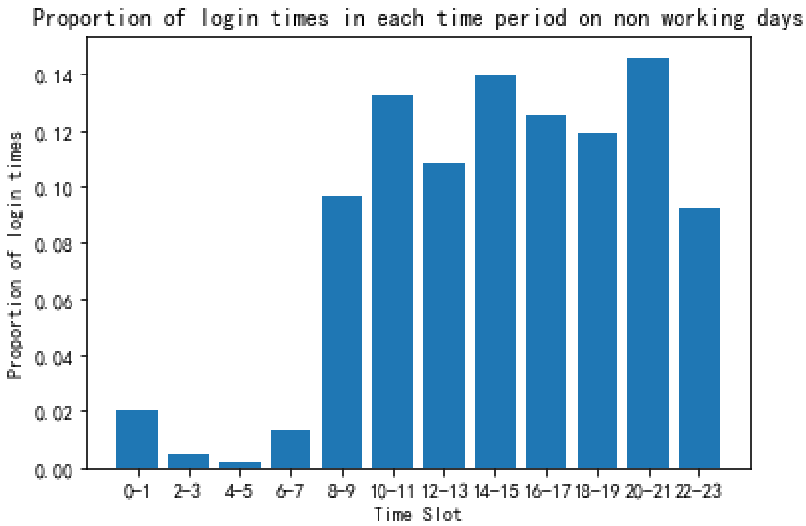

Figure 5: Visualization results for logging in at various times during non-working days

From the above graph, it can be seen that during non-working day's users log in between 8:00-23:00, and the most active time period for non-working days is between 20:00-21:00. The difference from working days is the number of logins in the evening. It is clear that during non-working days, users prefer to study in the evening.

3.3. Recommendations for online management decisions

3.3.1. Advocacy recommendations

The results of user distribution show that the online platform's users are concentrated in Guangdong, and there are obvious regional differences. Therefore, for the regions with dense user data, the online platform can maintain the original strength to promote, further open the market of online courses and increase customer credibility and activity. More importantly, for these key areas to increase research

ISSN 2522-6398 Vol. 4, Issue 16: 24-30, DOI: 10.25236/FER.2021.041605

efforts, to provide the data base for future marketing programs.

It is clear from the survey that the number of users of online education platforms is proportional to the economic level of the corresponding city and the level of development of the Internet, so these two points need to be added to the consideration.

3.3.2. Timing recommendations

According to the visual results of the histogram, it can be seen that the log-in frequency of users is basically the same for both weekdays and non-working days, which are in the three hot periods of 9:00-11:00, 14:00-17:00 and 20:00-21:00. And the highest peak value occurs in the morning on weekdays, while the peak occurs in the evening on non-working days, so according to the peak degree of the time interval, the online education platform strengthens the stability of the system and the fluency of the course during that time period to ensure the normal operation of the course. Once again, based on the implantation of the corresponding marketing strategy, the amount of users is further increased through user activities to firmly grasp the users and improve the attractiveness of the courses.

4. Online Course Recommendations

4.1. Collaborative filtering algorithm

Introducing prediction and recommendation algorithm, user-based collaboratlve filtering algorithm (user-based collaboratlve filtering). The algorithm is used to locate and discover users' preferences for online courses by using their historical behavioral data, and a metric is applied to these online courses. The relationship between users is calculated based on their preferences for the same online courses and different online courses are recommended among users with the same preferences.

Thus the binary matrix between the user and the course is constructed as follows.

course_id	course0	course1	course10	course100	course101	course102	course103	course104	course105	course106	...	course90	course91	course92
user_id														
user10	0	0	0	0	0	0	0	0	0	0	...	0	0	0
user100	0	0	0	0	0	0	0	0	0	0	...	0	0	0
user10000	0	0	0	0	0	0	0	0	0	0	...	0	0	0
user10001	0	0	0	0	0	0	0	0	0	0	...	0	0	0
user10002	0	0	0	0	0	0	0	0	0	0	...	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
user9993	0	0	0	0	0	0	0	0	0	0	...	0	0	0
user9994	0	0	0	0	0	0	0	0	0	0	...	0	0	0
user9995	0	0	0	0	0	0	0	0	0	0	...	0	0	0
user9996	0	0	0	0	1	0	0	0	0	0	...	0	0	0
user9999	0	0	0	0	0	0	0	0	0	0	...	0	0	0

Figure 6: Binary matrix of user courses

4.2. Course similarity

Jaccard related coefficient is introduced in the calculation of course similarity. Set A is user_id, set B is course_id, and the number of elements in the intersection of sets A and B is called the Jaccard coefficient of the two sets A and B. The higher the Jaccard coefficient means the higher the similarity. The formula is as follows.

$$J(A,B) = |A \cap B| / |A \cup B| \tag{1}$$

Then the results of similarity between courses are as follows.

ISSN 2522-6398 Vol. 4, Issue 16: 24-30, DOI: 10.25236/FER.2021.041605

course_id	course0	course1	course10	course100	course101	course102	course103	course104	course105	course106	...	course90	course91
course_id													
course0	1	0.5	0.0277778	0.4	0	0.666667	1	1	0.25	0	...	0	0
course1	0.5	1	0.0263158	0.285714	0.00206186	0.4	0.5	0.5	0.166667	0	...	0	0
course10	0.0277778	0.0263158	1	0.0526316	0.00193798	0.027027	0.0277778	0.0277778	0.027027	0.00247525	...	0	0
course100	0.4	0.285714	0.0526316	1	0.00205761	0.6	0.4	0.4	0.142857	0	...	0	0
course101	0	0.00206186	0.00193798	0.00205761	1	0	0	0	0	0	...	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
course95	0.00862069	0.00847458	0.171875	0.0169492	0.00336134	0.00854701	0.00862069	0.00862069	0.00854701	0	...	0.00869565	0
course96	0.00438596	0.00873362	0.0650407	0.0131004	0.0216138	0.00436681	0.00438596	0.00438596	0.00436681	0.00505051	...	0.00440529	0
course97	0.00555556	0.0110497	0.0862944	0.010989	0.0216383	0.00552486	0.00555556	0.00555556	0.00552486	0.00182482	...	0.00558659	0
course98	0.00348432	0.00521739	0.0322034	0.00695652	0.0677452	0.00347826	0.00348432	0.00348432	0.00173611	0.0407938	...	0.00174216	0.00174216
course99	0.00529101	0.0105263	0.0878049	0.0157895	0.0213415	0.010582	0.00529101	0.00529101	0.00526316	0.00179533	...	0.00531915	0

Figure 7: Similarity of some courses

4.3. Course recommendations

Extracts the top user ids and corresponding courses for total learning progress, and recommends the top three courses for users based on collaborative filtering algorithm and course similarity.

	user_id	course_id	rem
0	user1193	[course180, course184, course202, course201, c...	[course162, course97, course191]
1	user13841	[course32, course202, course56, course130, cou...	[course162, course158, course201]
2	user32684	[course202, course201, course56, course141, co...	[course7, course147, course67]
3	user36989	[course19, course202, course201, course56, cou...	[course2, course67, course32]
4	user24985	[course19, course202, course201, course56, cou...	[course67, course32, course43]

Figure 8: Top 5 users' corresponding courses and recommended courses

4.4. Online course recommendation strategy development

The first part is to develop a series of strategies for the course, and the second part is to develop some strategies for the users, where the central idea is to make more users learn more exciting and beneficial online courses.

4.4.1. User perspective

Different age groups have different needs for educational courses. For example, primary and secondary school students are more suitable for courses related to this stage, such as improving writing, study skills, mental arithmetic and oral arithmetic, while working people are more suitable for courses related to improving daily spoken English and work skills. Therefore, in order to sell courses in online education, it is necessary to first have a clear positioning of the user group. In the subsequent positioning operation, a more accurate positioning of users should be done using algorithms with higher accuracy.

4.4.2. Curriculum perspective

(1) Use soft copy marketing to build. Many times the sale of products need to be reflected in quality soft copy, through the use of quality articles can allow more users to understand the selling point of the online education course for sale, as long as the deepening of the user's understanding of the online education course, in order to achieve a deeper understanding of the course[8].

(2) Packaging of courses. Sometimes users often choose multiple courses with strong relevance, these courses will be added to the same section for "packaging", batch placement to the user. One to save the time of the user to find the collection, and the other to increase the flow of the course and user goodwill.

(3) Trial of the course. For some high quality fee courses can be used low price trial, such as "9.9 trial 30 minutes" to allow users to feel the quality of online courses, not only to increase the flow of the course, the amount of consumption can also enable users to understand the content of high-quality courses[9].

ISSN 2522-6398 Vol. 4, Issue 16: 24-30, DOI: 10.25236/FER.2021.041605

5. Conclusion

The Internet has ushered in a new phase of the information age, with online courses in Internet+education promoting diversity and self-improvement for students. Intelligent course recommendations allow students to choose a targeted and monotonous range of knowledge, avoiding unnecessary analysis and exploration time. However, it is difficult to guarantee the accuracy of matching, which limits the learning of other courses to some extent. Therefore, intelligent recommendations for online courses should be learner-centered, consider the differences and similarities between learning and students, avoid high concentration of courses, and help students achieve personalized and focused development based on the completion of general education courses[10].

Biographical notes

Wenjun Li is an undergraduate student of Changchun University of Technology, China. His research interest focuses on Software Engineering.

Yating Peng is an undergraduate student of Northeast Forestry University, China. Her research interest focuses on Big Data.

References

[1] Guo Qiang, Zhang Xuhu. A personalized recommendation system for MOOC based on collaborative filtering algorithm [J]. Information Technology, 2017 (6): 99-103.
[2] Gao Deyi, Zong Aidong. From Civic and Political Science Course to Curriculum Civic and Political Science: Constructing College Ideological and Political Education Curriculum System from Strategic High The construction of ideological and political education curriculum system in colleges and universities from a strategic perspective [J]. China Higher Education, Education in China, 2017 (1): 43-46.
[3] Gao Deyi, Zong Aidong. The Role of Classroom Education as the Main Channel for Education of People The natural choice to effectively play the role of the main channel of classroom education [J]. Journal of Ideological Education, 2017 (1): 33-36. (1): 33-36.
[4] Ke Xiuwen, Research on online course recommendation based on collaborative filtering technology. Intelligent Computers and Applications, 2018. 8(03): pp. 185-187.
[5] Intelligent adaptation of curriculum for "Internet + Education".
[6] Chen Yangxue, Research and application of MOOC hybrid recommendation algorithm based on big data platform, 2017, Hangzhou University of Electronic Science and Technology. Page 82.
[7] Zhao Quan, Research on intelligent course recommendation system in the context of big data. Information and Computer (Theory Edition), 2019(09): pp. 101-103+106.
[8] Li Xiangyong. Design and development of online course platform of Southwest Normal University Press [D]. Southwest University, 2020.
[9] Du R. Research on the application of hybrid recommendation based on ALS and TF-IDF in E-Learning platform [D]. Nanjing Normal University, 2020.
[10] Shang L. An improved course recommendation algorithm based on collaborative filtering [J]. Science and Technology Communication, 2020, 12(05): 132-134.