Course Selection Recommendation Based on Hybrid Recommendation Algorithms

Luxiao Zhu and Ben Wang(✉)

College of Information Science and Technology, Hangzhou Normal University, Hangzhou 310000, China
20170056@hznu.edu.cn

Abstract. In the era of national 'double first-class' and 'high-level' university construction, the course selection has become an important part of the curriculum system of many colleges and universities in China. Due to the shortcomings of the traditional course selection method, the need for a machine learning-based course selection recommendation system has become more and more urgent. The focus of this paper is to apply a hybrid recommendation algorithm to a course selection system for university students to achieve intelligent recommendations through a correlation between students and students and between students and courses. The experimental results show that the hybrid recommendation algorithm is very good at recommending courses for students in the university course selection recommendation system, and the recommended courses have good accuracy and reasonableness.

Keywords: Collaborative filtering algorithms · Content-based recommendation algorithms · Hybrid recommendation algorithms · Elective courses

1 Introduction

In recent years, universities have added additional elective courses to basic courses to cultivate knowledgeable talents with innovative abilities that meet the requirements of the times. [1] The traditional course selection method currently adopted has numerous shortcomings and deficiencies [2], with an inadequate guidance system for course selection, insufficient understanding of the courses by students, and blindness in course selection. In the context of the era of big data and machine learning, relying on machine learning technology to achieve personalized course selection for students has become a new trend in the development of universities.

This paper proposes a hybrid recommendation algorithm as the overall architecture, with a collaborative filtering algorithm as the main and a content-based recommendation algorithm as a supplement. The experiments use student data from S University as the test set to analyze the results and recommendation quality of the personalized course selection system based on the hybrid recommendation algorithm.

V. Balakrishnan et al. (Eds.): ICMEIM 2022, AHSSEH 6, pp. 476–482, 2022.
https://doi.org/10.2991/978-94-6463-044-2_60

2 The Basic Idea of the Three Recommendation Algorithms

2.1 Algorithms for Collaborative Filtering

Collaborative filtering algorithms are by far the most widely used and successful recommendation algorithms. There are currently two main types of collaborative filtering algorithms, user-based and item-based collaborative filtering. The core idea is the same: firstly, based on the rating data in the system, the similarity between users (or items) is used to find the set of nearest neighbors to the target user (or item), and the rating of the target user (or item) is predicted based on the rating of the user (or item) in the set, and this is used to make recommendations. In this application, collaborative filtering based on items is only used to recommend courses based on student performance, resulting in unsatisfactory results, so this study will not be expanded to describe the explanation [3].

The operation steps of user-based collaborative filtering algorithm are as follows.

1) Historical data about user behavior are captured.
2) Similar nearest neighbors are constructed and many users are selected based on greater similarity.
3) Exclude the user's selected courses from the collection of nearest neighbor courses, the item ratings are ranked and then the top N as recommendations are selected.

The key to a user-based collaborative filtering algorithm is the selection of nearest neighbors. To obtain the set of nearest neighbors a measure of similarity between users is required, and there are various methods, the most popular of which is cosine similarity [4]. In the cosine similarity approach, the user rating information is considered as a vector on an n-dimensional item space, and the rating vectors of users i and j on a common set of rating items are expressed as vector sums, respectively, then the cosine of the angle between users i and j is defined as

$$\text{simil(i, j)} = \cos\left(\vec{\text{i}}, \vec{\text{j}}\right) = \frac{\vec{i} \cdot \vec{j}}{\| \vec{i} \| \times \| \vec{j} \|} = \frac{\sum_{a \in I_{ij}} r_{i,a} r_{j,a}}{\sqrt{\sum_{a \in I_i} {r_{i,a}}^2} \sqrt{\sum_{a \in I_j} {r_{j,a}}^2}} \quad (1)$$

In the above equation, it can be seen that the cosine of the angle between the two users is less than or equal to 1. Its magnitude reflects the similarity between the two users, with close to 1 indicating high similarity and close to 0 indicating low similarity.

2.2 Content-Based Recommendation Algorithms

[4–6] The content-based recommendation method is derived from information retrieval and information filtering, which first models the data based on the attributes of the user and the recommended object, followed by updating the user model by the data model of the object that has been selected by the user. Thereafter, by comparing the similarity of user preferences to the object model, resource items with high similarity are selected for recommendation. Commonly used techniques are decision trees, Bayesian classification algorithms, neural networks, vector-based representations, etc.

Content-based recommendation algorithms are only applicable to textual recommendations, and the analysis of non-textual data, such as images, audio, video and other multimedia data, is still a difficult task.

2.3 Hybrid Recommendation Algorithms

The hybrid recommendation has become a hot topic of research in recommendation systems at this stage. The most important principle of this algorithm is to mix several algorithms according to the actual situation, to complement each other's strengths and weaknesses, and to obtain more accurate recommendation results. Theoretically, there are various recommendation combinations, but the common hybrid recommendation algorithms are weighted, transformed, merged, feature combination, waterfall, feature incremental and meta-hierarchical [4].

The overall architecture of the course selection recommendation system is a 'transformation' hybrid recommendation, which avoids the problems of coarse content-based recommendations and the cold-start problems of collaborative filtering algorithms due to the complexity and variability of course data and student behavior data.

3 Implementation Steps

This paper combines course selection functionality with hybrid recommendation technology to enable intelligent course selection. Using explicit and invisible data from students, trends in students' interests are analyzed and course lists are provided to them. With a focus on the application of the push process, collaborative filtering algorithms and content-based recommendations are applied to the push process, with students and courses as the main objects of study.

3.1 Data Collection

The data generated by students in universities are divided into static and dynamic data [8], which is obtained through a mixture of explicit and implicit acquisition methods [9]. The system obtains static data from students by actively filling in their registration forms, and uses techniques such as web data mining with buried 'invisible' probes or web data mining of the SDK to obtain implicit feedback from students. The data collected through this method contains a lot of 'dirty data' and requires a lot of computational processing to form structured and labeled data, outlining the overall characteristics of the student and building a model. As the dynamic student data captured is subject to change over time, the student model has to be continuously updated and revised based on the underlying data.

The student data set used in this experiment is from the S College academic system and contains 352 items. These data are divided into students, and courses and each table contains the following detailed information content.

Students: student ID, gender, age, year of study, faculty, number of canteen swipes, number of access card swipes, number of library borrowings, number of online course visits, average marks for each category of course.

Table 1. STUDENT COURSE TYPE GRADES [SELF-PAINTED]

Student ID	***Professional Foundations***	***Education***	***Software Programming***	***Design***	***Finance***
001	5	1	0	4	2
002	2	0	2	3	5
…	…	…	…	…	
352	4	0	5	2	3

Course: Course ID, course name, course category, instructor.

This paper is based on the elective courses at S College, which are divided into five categories: Professional Foundations, Education, Software Programming, Design, and Finance. The above information was collected and the students' mean scores for each type of elective course were analyzed and expressed on a 5-point scale. The following two-dimensional matrix is used to represent the student's scores for the courses, with higher value in the table indicating a better mastery of the category (Table 1).

3.2 Data Collation

Student attributes and student behavioral characteristics were extracted from the student data collected above, and then analyzed in terms of students' mastery of each type of course in terms of their logical thinking, creative, operational and application skills, as represented below.

Student (ID, gender, age, year of study, faculty, logical thinking skills, creative skills, operational skills, application skills), e.g. Student (001, male, 22, Computer Science and Technology 14, School of Computer Science, 40, 23, 20, 17).

The course attributes are defined as follows.

Course (course ID, course name, course duration, course type, instructor).

3.3 Research Procedures

The basic framework of the personalized course selection recommendation system is shown in Fig. 1.

A hybrid recommendation algorithm is used as the overall framework, with a collaborative filtering algorithm as the main body and a content-based recommendation algorithm as a supplement to the push process of the course selection system as shown in Fig. 2.

The algorithm is first trained offline to find a set of courses similar to a course. When a student sends a request online, the student model is constructed, similar student models are found, and if there is one, the Top-1 is selected for a recommendation, if not, a candidate set is obtained based on the courses taken by students with high ratings and similar courses in similar student models, and then the Top-N courses are selected as the recommended result set by sorting.

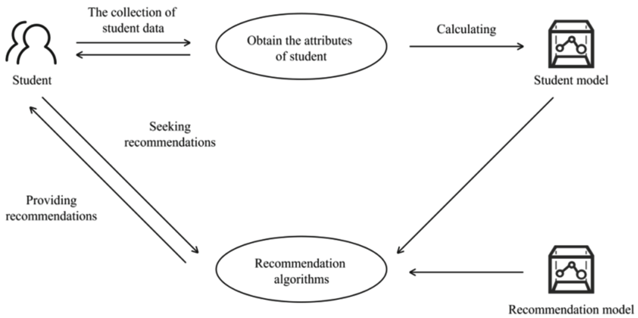

Fig. 1. Recommendation model of the hybrid recommendation algorithm-based course selection system [self-painted]

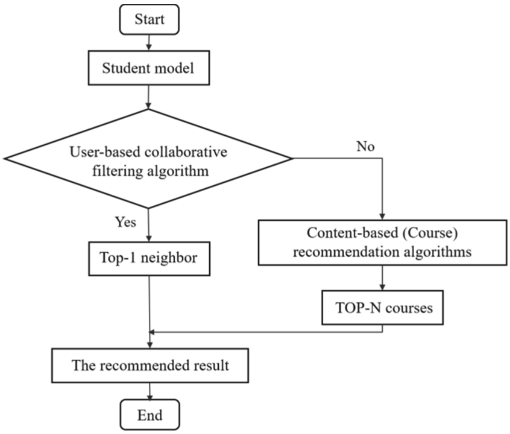

Fig. 2. Flow chart of the hybrid recommendation algorithm-based course selection system [self-painted]

As can be seen from the above description, the hybrid recommendation algorithm in this paper consists of similarity calculation, the course selection and generation of recommendations.

1) **Student similarity calculation**
To calculate student similarity, the combined scores of the four abilities of students i and j are considered as two vectors with the following formula.

$$\mathrm{sim}(\mathrm{i},\mathrm{j}) = \frac{\sum_{\mathrm{a}\in \mathrm{I}_{\mathrm{i},\mathrm{j}}}\left(\mathrm{r}_{\mathrm{i},\mathrm{a}} - \overline{\mathrm{r}_{\mathrm{i}}}\right)\left(\mathrm{r}_{\mathrm{j},\mathrm{a}} - \overline{\mathrm{r}_{\mathrm{j}}}\right)}{\sqrt{\sum_{\mathrm{a}\in \mathrm{I}_{\mathrm{i},\mathrm{j}}}\left(\mathrm{r}_{\mathrm{i},\mathrm{a}} - \overline{\mathrm{r}_{\mathrm{i}}}\right)^2}\sqrt{\sum_{\mathrm{a}\in \mathrm{I}_{\mathrm{i},\mathrm{j}}}\left(\mathrm{r}_{\mathrm{j},\mathrm{a}} - \overline{\mathrm{r}_{\mathrm{j}}}\right)^2}} \tag{2}$$

$\mathrm{sim}(\mathrm{i},\mathrm{j})$ denotes the similarity between the currently specified student *i* and student *j*. *r* denotes the combined rating of the four competencies of *i* and *j*. $I_{i,j}$ denotes the four competencies, logical thinking, innovation, operation and application. *a* denotes one of the four competencies, e.g. logical thinking. $r_{i,a}$ and $r_{j,a}$ denote the ratings

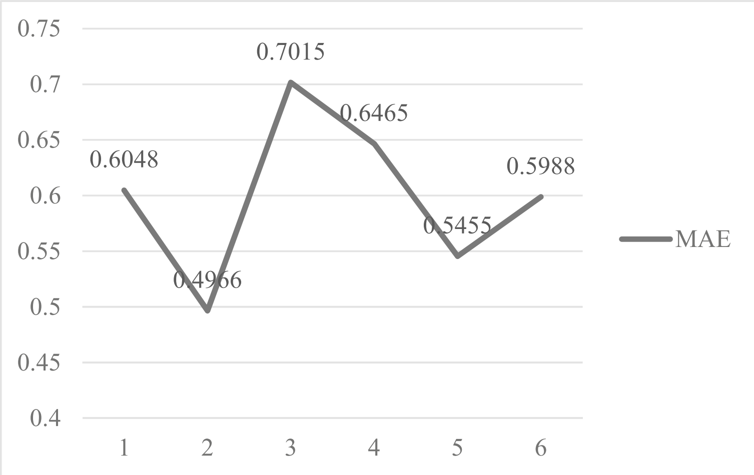

Fig. 3. Predictive accuracy of recommended courses [self-painted]

of students i and j on competency of a respectively. Based on this, similar users of the target students are identified, the passing courses taken by similar users are extracted, and the courses already taken by the target users are excluded to generate the array C.

2) **Course selection**
The target student's performance is standardized, and for the 5 categories of course classification the sum $\{F_1, F_2, F_3, F_4, F_5\}$, F_{Max} is the best category, and the courses are ranked M according to the Top-N rule.

3) **Generating recommendations**
Do the intersection of arrays C and courses in M to generate a list of recommendations; if there are no similar users then recommend M.

4 Experimental Results

In this paper, the mean absolute error (*MAE*) was chosen as the evaluation criterion for the algorithm, and Fig. 3 shows the prediction accuracy of the recommended courses for students with student numbers 1–6. A smaller MAE indicates that the algorithm's recommendation performance is better and that the courses recommended by the system do not differ significantly from those appropriate for the students in the test set.

5 Conclusion

This paper uses a 'transformation' hybrid recommendation algorithm to study and analyze students' logical thinking ability, innovation ability, operational ability and application ability based on their explicit and implicit data and constructs a course selection recommendation system for university students based on a collaborative filtering-based algorithm supplemented by content-based recommendations. The experimental results show that the algorithm can solve the randomness and blindness of students' course selection and help to achieve the best match between students' learning ability and course requirements.

References

1. Zhao Shanshan. Ten reasons why college students find it hard to find jobs were revealed [EB/OL]. China.com.cn. 2014–09–16.
2. Ministry of Education. The decision of the Central Committee of the Communist Party of China on the reform of the education system [R],1985.
3. Sun Huijun. Research on University Course Selection Push Based on Machine Learning [D]. Inner Mongolia Normal University, 2017.
4. Liu Peng. Research and Implementation of doctor Recommendation System Based on Combination Model [D]. Shanghai: Donghua University, 2013.
5. Pazzani M J, Billsus D. Content-based recommendation systems[M]//The adaptive web. Springer, Berlin, Heidelberg, 2007: 325–341.
6. Liu Wei. Research on Information Recommendation Method in E-commerce System [J]. Information Science, 2006, 24(2): 300–303.
7. Shardanand U, Maes P. Social information filtering: Algorithms for automating "word of mouth" [C]//Proceedings of the SIGCHI conference on Human factors in computing systems. 1995: 210–217.
8. Zhao Guoliang, Chen Xiaojun, Li Siqi, et al. A preliminary study of college students' self-portrait based on data analysis [J]. Digital Technology and Application, 2017 (8): 233–234.
9. Zhang Yuhang, Yao Wenjuan, Jiang Shan. Overview of personalized recommendation systems [J]. Value Engineering, 2020.