

Research of Personalized Course Recommended Algorithm based on the Hybrid Recommendation

Qing Yang¹, Peiling Yuan^{1,*}, and Xi Zhu²

¹Department of Computer, Central China Normal University, Wuhan, China

²TCL Yunchuang Technology Co., Ltd., Guangdong, China

Abstract. This paper presents a personalized course recommended algorithm based on the hybrid recommendation. The recommendation algorithm uses the improved NewApriori algorithm to implements the association rule recommendation, and the user-based collaborative filtering algorithm is the main part of the algorithm. The hybrid algorithm adds the weight to the recommendation result of the user-based collaborative filtering and association rule recommendation, implementing a hybrid recommendation algorithm based on both of them. It has solved the problem of data sparsity and cold-start partially and provides a academic reference for the design of high performance elective system. The experiment uses the student scores data of a college as the test set and analyzes results and recommended quality of personalized elective course. According to the results of the experimental results, the quality of the improved hybrid recommendation algorithm is better.

1 Overview

In recent years, the research of education data mining in school teaching has arisen gradually, such as clustering analysis, genetic algorithm, association rule mining, collaborative filtering recommendation algorithm, etc. How to use the existing large amount of educational administration data and combine the related technology of data mining to provide students with personalized teaching services is a very important topic in the current university teaching management application.

The personalized recommendation technology mainly includes the recommendation based on the content, the recommendation based on collaborative filtering and the combination recommendation. Content-based filtering algorithm is an important aspect in the field of information retrieval. The key of this method is to describe the content features of users and projects, but the extraction of text content will affect the content characteristics of the project, which is a difficult problem to be solved for discovering some new interests of users¹. For content-based filtering defects, in 1992, Goldberg proposed the concept of collaborative filtering in the research report² and applied collaborative filtering technology to establish an e-mail filtering system, which is equivalent to a later collaborative filtering recommendation system. Collaborative filtering recommendation includes two kinds; one is user-based collaborative filtering, the other is item-based collaborative filtering³. The user-based collaborative filtering algorithm is user-centric and recommends preferences of similar users to the user. If the number of users continues to increase, it is difficult to get similar

neighbors of the current user among a large number of users. In 2001, Karypis proposed a project-based collaborative filtering algorithm⁴. Instead of using similarity between computing users, the algorithm uses the similarity between the calculated items. Since the number of items is relatively small, this solves the problem that it is difficult to find the similar neighbors of the specified users due to too many users. The algorithm is based on the similarity of the items to recommend similar items, which can solve the problem of some missing rating data. The traditional project-based collaborative filtering recommendation algorithm still can't find the users' similarity of different evaluation items, and it is not very effective in seeking multiple content items. Therefore, some scholars have proposed a collaborative filtering algorithm based on project attributes⁵. This algorithm calculates the similarity of project attributes to recommend for users, which has solved the problem of cold-start of the project and the sparsity of rating data, but the algorithm still has the problem of cold-start for new users.

Domestic research on personalized recommendation technology is relatively late, and related research has only begun in the past decade. With the continuous development of informational technology, personalized recommendation systems have been widely used in many fields and provide users with real-time recommendations. But the system's recommended accuracy and cold-start problems are still difficult to solve. At present, personalized information services are also gradually applied in the field of education, such as personalized libraries, online teaching, educational management systems and personalized resource

* Corresponding author: 1327509019@qq.com

recommendations, etc. University of Science and Technology of China has a personalized course system for undergraduates. The system obtains the relationship between different students and different courses through the strong association rules discovered by association rules. It can recommend related subjects for students and achieve personalized recommendations for individuals⁶. The system uses the student's course data as the test data set for mining. However, if the amount of data are too large, the running time of association rule mining will increase, and the system also has problems of cold-start for new users and new courses. The hybrid recommendation algorithm studied in this paper can solve the problem of missing data to a certain extent. For example, it can predict the students' achievement data for a missing course and solve the cold-start problem of new projects.

2 The Basic Idea of Association Rule Recommendation Algorithm and Collaborative Filtering Algorithm

2.1 Association Rule Recommendation based on improved NewApriori Algorithm

The basic idea of association rule mining algorithm is to find the items that are often selected simultaneously through the mining of association rules. When these association rules are mined, the user is recommended based on these rules⁷.

Apriori algorithm is one of the most commonly used algorithms in association rule mining. It generates candidate sets and frequent item sets through continuous iteration and generates association rules until the end of the iteration. Apriori algorithm is easy to understand and convenient to implement, but it has some defects in efficiency. In this paper, an improved NewApriori algorithm is used by adopting the improved strategy of transaction compression and candidate set pruning. During the process of generating candidate sets, the algorithm has removed the candidate sets with minimal support, reduced the size of the data sets that need to be scanned, saved the time of the algorithm scanning error database and improves the efficiency of the algorithm. The implementation steps and the comparison of experimental quality of the improved NewApriori algorithm have been described in detail in the published papers⁸.

The recommendation algorithm based on association rules is to obtain useful strong association rules through association rule mining, analyze the obtained rules and recommend it to users. The algorithm used in this paper for association rule mining is the improved NewApriori algorithm proposed above.

2.2 User-based Collaborative Filtering Recommendation Algorithm

The user-based collaborative filtering recommendation algorithm is the most widely used method in personalized recommendation. There are two main work of the algorithm, the first is to search for the user's nearest neighbor set, and the second is to generate a recommendation. The key point of the first step is to calculate the similarity between users. The similarity can be obtained through the Jaccard formula or cosine similarity, that is, the proportion of common ownership behavior. First, the similarity calculation is performed on the specified user and other users, the nearest neighbor set of the specified user is found out according to the degree of similarity, then the preference of the specified user is obtained according to the preference of the similar neighbor. Therefore, the user-based recommendation algorithm mainly predicts the user's rating of the item by seeking the similarity between the users and generating the recommendation process, and selects the item with the largest rating value to be recommended to the specified user.

This algorithm is more suitable for a system with a small number of users. If the number of users is large, the user's preference similarity matrix will be very large, which will increase the time complexity of the matrix calculation. And if the user has a new behavior, the recommendation results from this algorithm cannot be updated accordingly.

3 Hybrid Algorithm based on improved Association Rules and Collaborative Filtering

In each university's course selection system, thousands of new student data are imported into the course selection system each year, and there will be new courses. When importing new users and new courses, these new users have not selected courses and have not produced results. And no user elects to take this new course and produces a grade, then new users and the new courses cannot be scored. This is a missing data problem. For these existing defects, this paper proposes a hybrid recommendation algorithm based on the improved NewApriori algorithm for association rule mining and collaborative filtering course selection algorithm, which is recommended for students to select classes.

3.1 Principle and Model of Hybrid Recommendation Algorithm

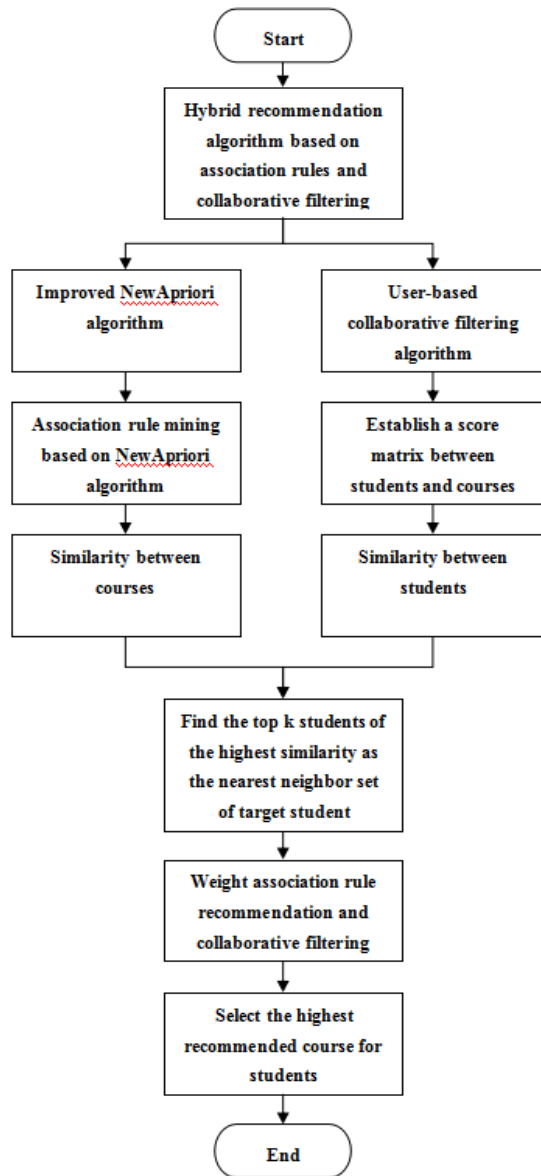


Fig. 1. Flowchart of hybrid recommendation algorithm based on association rules and Collaborative filtering.

This paper introduces the idea of generating strong association rules by mining frequent item sets, establishes the association matrix between curriculum items, obtains the similarity between courses, then generates frequent item sets and produces association rules, which are recommended by the calculated rules to students' recommendation courses. Because the frequent course item set mined by the association rule recommendation has a certain correlation relationship with the course that the student has chosen, the degree of course association is introduced, which is more comprehensive when considering the student interest course. In the part of user-based collaborative filtering recommendation algorithm, the similarity between students is first sought, similar student neighbors of the specified student are searched, then the recommended value recommended by the association rule and the recommended value obtained by the collaborative filtering recommendation are weighted according to a

certain weight, and the courses recommended to the highest value are recommended to students to achieve the hybrid recommendation of association rules and collaborative filtering.

3.2 Steps for implementing a Hybrid Recommendation Algorithm

This paper's hybrid recommendation algorithm includes the similarity calculation, the establishment of a score matrix between students and courses, the search for students' nearest neighbors and the generation of recommendations.

(1) Similarity between students

In the process of calculating similarity, the scores of the courses selected by the students i and j can be regarded as two vectors respectively. The students' average scores for the elective courses are subtracted to achieve the effect of correcting the similarity, because the values of different students' course scores will affect the recommendation of the selected course. The formula is as follows:

$$sim(i, j) = \frac{\sum_{a \in I_{i,j}} (r_{i,a} - \bar{r}_i)(r_{j,a} - \bar{r}_j)}{\sqrt{\sum_{a \in I_{i,j}} (r_{i,a} - \bar{r}_i)^2} \sqrt{\sum_{a \in I_{i,j}} (r_{j,a} - \bar{r}_j)^2}}$$

$sim(i, j)$ denotes the similarity between the current specified student i and student j , r denotes the set of results of i and j for an elective course, and $I_{i,j}$ denotes the course selected by i and j together and produces the achievement, $r_{i,a}$ and $r_{j,a}$ respectively represent the common scores of the students i and j for the course a . Sort according to the similarity size, and select the top k students with the highest similarity to form the nearest neighbor set of the specified student.

(2) Establish a score matrix between students and courses

First, the data need to be cleaned and processed. The student's course performance data are obtained based on the student's majors, gender, course selection records and course results. The conversion value of the course performance is used as the value of the matrix element to form score matrix between the students and the courses. As for the missing scores in the students' course performance data, the similarity between the courses can be calculated according to the association rules, so as to predict the results of the course.

For the missing grades in the student's course achievement data, the similarity between the courses can be calculated according to the association rule recommendation, so as to predict the score of the user on the courses that do not result in the achievement. Assume that student i is specified, and the specified course for predicting the achievement is a , the similarity of a is obtained and the nearest neighbor set of a is obtained, then the score on a is predicted by the result of nearest neighbor set of i .

(3) Search the nearest neighbor of the student

Compare the similarity between all the students in the course score matrix and the specified student. Sort the similarity according to the size of the similarity. Select the top k of student with the highest degree of similarity with the specified student to get the nearest neighbor set of the specified student.

(4) Generate recommended courses

The recommendation value generated by the association rule recommendation is weighted with the recommendation value generated by the user-based collaborative filtering recommendation.

$$P_{i,a} = \bar{r}_i + \lambda \times \frac{\sum_{j \in S} \text{sim}(i, j) \times (r_{j,a} - \bar{r}_j)}{\sum_{j \in S} |\text{sim}(i, j)|} + (1 - \lambda) \times \frac{\sum_{b \in N_a} R_{a,b} \times (r_{i,b} - \bar{r}_b)}{\sum_{b \in N_a} R_{a,b}}$$

$P_{i,a}$ represents the predicted score of student i on course a in the collaborative filtering recommendation algorithm, $R_{a,b}$ represents the strength of the association of course b on course a , and $r_{i,b}$ represents the score of course b of student i , N_a refers to the nearest neighbor set of course a . S is the set of k similar students for user i , that is, the nearest k neighbor set. $r_{j,a}$ is the score of student j for the course a . Sort according to the value of $P_{i,a}$ from big to small, select the course with the largest recommended value to recommend to the specified student.

4 Experiment of Personalized Course Recommendation

Experimental method: Run hybrid recommendation algorithm of improved association rule mining and collaborative filtering. Divide the student achievement data into test data sets and training data sets. Predict the courses and scores in the test data sets through the courses and scores in the training data sets, and then compare the courses and scores actually selected by students in the test data set with the courses and scores predicted to compare the quality of the recommended algorithms.

4.1 Experimental Data Set

The experimental data set adapts the elective information and course results of three computer courses for all professional students of the 2013, 2014 and 2015 levels of the Central China Normal University. There are a total of 13,140 student information. The three computer courses are multimedia technology and applications, database technology and applications, and high-level language programming. In order to facilitate the operation of the recommended algorithm, the experimental data set was processed. The value of student computer-related courses was numerically converted. The percentage system was divided into 0(score=60)to indicate that the score was empty, 1(score<60) indicates unsatisfactory results; 2(60<score<70) indicates that the score is acceptable;

3(70<score<80) indicates that the score is intermediate; 4(80<score<90) indicates that the score is good; 5(90<score<100) indicates a total of six outstanding results. Therefore, these 6 values respectively represent 6 behaviors of student achievement: 0:missing results; 1:unsatisfactory results; 2:qualified results; 3:moderate results; 4:good results; 5:excellent results. The student's course performance data set is a good response to the student's performance of the course. At the same time, the student's course performance data set is divided into training data set and test data set, which respectively account for 70% and 30% of the total data set.

4.2 Personalized Course Selection Recommended Experimental Results and Recommended Quality Analysis

The course recommendation results are as follows:

The course recommended for user 0 is:Database Technology Capabilities,the predicted score:3.29

The course recommended for user 1 is:Database Technical Knowledge,the predicted score:3.00

The course recommended for user 2 is:Database Technical Knowledge,the predicted score:5.00

The course recommended for user 3 is:Database Technology Capabilities,the predicted score:4.29

The course recommended for user 4 is:Database Technical Knowledge,the predicted score:4.00

The course recommended for user 5 is:Database Technology Capabilities,the predicted score:4.41

Fig. 2. Improved course recommendation results based on association rules and collaborative filtering recommendations.

Figure 2 is a screenshot of a partial selection of a student's recommended course. As shown in Figure 2 for the recommended courses of 6 students, the personalized recommendation algorithm predicts the students' courses and corresponding scores. It can be seen that the prediction scores of recommended course for the students are all above the middle score, which can provide students with a selected items for reference.

Among the three recommendation algorithm quality comparison analysis experiments of the collaborative filtering selection recommendation algorithm, the improved association rule recommendation algorithm and the hybrid recommendation algorithm based on improved association rules and collaborative filtering, compare the recall rates and accuracy of the three different recommended algorithms respectively in the cases of 10, 20, 30, 40, 50, 60 and 70, then compare their recommended quality to each other. When the value of k is small, the operation time of the algorithm may be greatly increased. However, if the value of k is too large, the accurate recommendation result may not be obtained. Finally, the value of k is selected as the final value when the recommended quality is optimal. The corresponding accuracy rate is compared with the line graph shown in Figure 3:

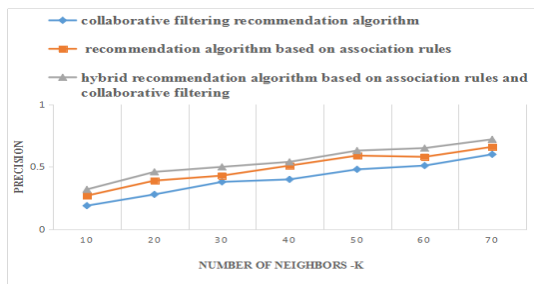


Fig. 3. The comparison of the accuracy rate of the course recommendation algorithm.

From Figure 3, we can see that when k is gradually increased, the accuracy of the three algorithms has been gradually improved, and the accuracy of the improved hybrid recommendation algorithm is the highest, indicating that there is not much difference between the recommended courses for students and the courses selected by students on the test set.

Comparison experiments were conducted on the three algorithms, and the corresponding recall ratio versus the line graph are shown in Figure 4 respectively:

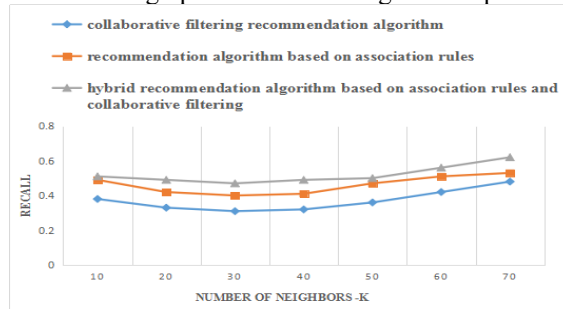


Fig. 4. The comparison of recall rate of the course recommendation algorithm.

From Figure 4, we can see that when k increases from 10 to 30, the recall rate of the three algorithms all show a slight decrease and gradually increase when k rises from 30 to 70. The recall rate of the improved hybrid recommendation algorithm is higher than that of the original collaborative filtering recommendation algorithm and the association rule recommendation algorithm.

Under normal circumstances, the recall rate will decrease when the accuracy rate is high, the accuracy rate will decrease when the recall rate is high, and the ideal condition for both increases is the best. However, this state cannot always be achieved. From the experimental results of the accuracy rate and recall rate of this paper, we can see that when k is between 10 and 30, the recall rate has a slight decrease with the improvement of the accuracy rate. When k is between 30 and 70, the recall rate increases, the recall rate has also gradually increased. According to the analysis of the experimental results, we can get more accurate recommendation results when the value is about 70. At this time, the hybrid recommendation algorithm achieves an ideal state in which both the accuracy rate and the recall rate increase. Therefore, the improved mixed selection algorithm has more accurate recommendation results and better recommendation quality.

5 Conclusion

This paper presents a hybrid recommendation algorithm based on the improved NewApriori association rules and collaborative filtering. The recommendation algorithm uses the improved NewApriori algorithm to mine the association rules and implements the association rule recommendation, and the user-based collaborative filtering algorithm is the main part of the algorithm, and a hybrid recommendation algorithm based on both is implemented. The algorithm obtains the target student's similarity, generates a nearest neighbor set, and provides a recommended course for the target student. At the same time, this paper designs the contrast experiment of the hybrid recommendation algorithm and compares it with the selection of the original collaborative filtering algorithm and the recommendation algorithm based on association rules, and then analyzes the experimental results. The recall and accuracy of collaborative filtering recommendation algorithm and hybrid recommendation algorithm based on association rules and collaborative filtering were also compared. The experimental results show that the improved hybrid selection algorithm has a smaller average absolute error and higher accuracy. Therefore, the recommended quality of the proposed hybrid selection algorithm is better.

The next research will apply the personalized recommendation to the college elective system, store the students' majors, interests, selected courses and other information in the elective system, then get the recommended results and recommend the students by the calculation of the related algorithms of data mining. This will play a very good role in promoting the informationization and intellectualization of elective courses in colleges and universities. In the near future, with the development of personalized recommendation technology, the traditional elective system of pure credit system has the possibility of being replaced by the individualized elective system gradually.

References

1. Zeng Chun, Xing Chunxiao, Zhou Lizhu. Journal of Software, **14**(5)(2003)
2. Goldberg D, Nichols D, Oki B M, et al. Communications of the ACM, **35**(12):61-70.
3. Wang Guoxia, Liu Heping. Computer Engineering and Applications, **48**(7)(2012)
4. Sarwar B, Karypis G, Konstan J, et al. Proceedings of the 10th international conference On World Wide Web. ACM, 285-295(2001).
5. Sun Jingang, Ai Lirong. Journal of Computer Applications, **32**(03):658-660(2012).
6. Xiao Jianqiong, Song Guoqin. Computer and Applications, **2**(05):73-74(2012).
7. Zhang Yuntao, Yu Zhilou, Zhang Huaxiang. Computer Engineering and Applications(2011).
8. Zhu Xi, Yang Qing and Du Fei Yang. ETMI(2016).