

Association Rule Model of On-demand Lending Recommendation for University Library

Shixin Xu

Huaiyin Institute of Technology, Huaian, Jiangsu 223003, China

E-mail: xusx@hyit.edu.cn

Student paper

Keywords: library, recommendation, association rules, Bayes

Received: August 31, 2020

University library that is connected to the Internet is more convenient to search, but the huge amount of data is not convenient for users who lack a precise target. In this study, the traditional association rule algorithm was improved by a Bayesian algorithm, and then simulation experiment was carried out taking borrowing records of 1000 students as examples. In order to verify the effectiveness of the improved algorithm, it was compared with the traditional association rule algorithm and collaborative filtering algorithm. The results showed that the recommendation results of the improved association rule recommendation algorithm were more relevant to students' majors, and the coincidence degree of different students was low. In the objective evaluation of the performance of the algorithm, the accuracy, recall rate and F value showed that the personalized recommendation performance of the improved association rule algorithm was better and the improved association rule algorithm could recommend users with the book type that they need.

Povzetek: Opisan je asociativni algoritem z dodanim Bayesovim klasifikatorjem za iskanje po univerzitetni knjižnici.

1 Introduction

The arrival of the Internet era has made great changes in our lives, the most intuitive expression of which is that the amount of information that can be obtained far exceeds the era before the emergence of the Internet [1]. Although the Internet with a large amount of information greatly facilitates people's lives, the huge amount of information also greatly increases the difficulty of people's retrieval of effective information. The same is true in the university libraries. Generally, the number of books in a university library is very large. In order to meet the needs of university teachers and students, the selection of books is often very rich [2]. According to the method of field search, it takes time and effort to browse the bookshelves one by one. When the Internet is combined with the university library, the book information in the university library is uploaded to the Internet, and the university teachers and students can simply retrieve the desired book information by using the Internet [3]. However, similar to the Internet described before, although the amount of book information in university library cannot be compared with the amount of data in the whole Internet, it is still a huge amount of data for university teachers and students. If there is a clear goal, it can be accurately retrieved, but if there is only a vague demand range, it is difficult to accurately retrieve the required information. Zhang [4] proposed a personalized book recommendation algorithm based on time series collaborative filtering recommendation and found through experiment that the book recommendation algorithm met the professional

learning needs of college students. Sohail et al. [5] put forward an opinion mining based recommendation technology which provided college students with promising books in the syllabus and found through experiment that the accuracy of this method improved by 55% and it could be applied to the recommendation of other products. Chahinez et al. [6] proposed a book recommendation method based on complex user query, and the experimental results showed that the combination with retrieval model could significantly improve the standard ranked retrieval metrics. In this study, the traditional association rule algorithm was improved by a Bayesian algorithm, and then simulation experiment was carried out taking borrowing records of 1000 students as examples. In order to verify the effectiveness of the improved algorithm, it was compared with the traditional association rule algorithm and collaborative filtering algorithm.

2 Book recommendation algorithm based on association rules

2.1 Association rule algorithm

Association rule recommendation algorithm [7] is to find the connection between different project elements from a large data set and regard the connection whose degree exceeds the set threshold as a strong association rule to guide the recommendation of books. The key point of the

association rule recommendation algorithm is to find the strong rule in the database. The algorithm is generally divided into two steps: ① search frequent sets in the database; ② search strong rules in the frequent set.

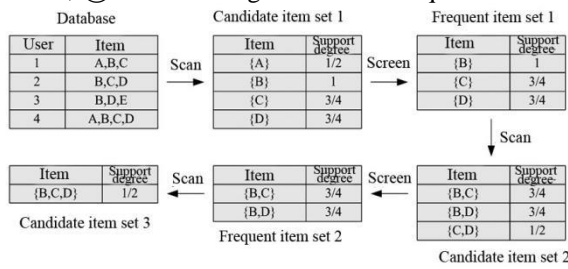


Figure 1: The basic diagram of association rule algorithm flow.

For convenience, as shown in Figure 1, numbers represent users, letters represent the names of books borrowed by users, and the number of records is reduced to 4 users and 4 books. Firstly, a book is taken as the candidate item, and then the support degree of each item in candidate project set 1 is calculated [8] using the following formula:

$$SUP = \frac{n}{N}, \quad (1)$$

where SUP is the support degree of item, N is the total number of records in the database, for example, there are 4 borrowing records of users in Figure 1, and n is the number of records containing the item. According to the set support degree threshold, frequent set 1 is filtered out, then frequent set items are combined to form a new candidate set, and the new frequent set is filtered out; the above operation is repeated until no candidate set can be obtained.

In addition to calculating the support degree, the confidence degree should also be calculated for searching the strong rule in frequent item set. Taking $\{B, C\}$ item in frequent item set 2 in Figure 1 as an example, a strong association rule may produce between its non-empty subset and the set of remaining elements, then the possible strong rule is $\{B\} \Rightarrow \{C\}$ and $\{C\} \Rightarrow \{B\}$. For association item set $X \Rightarrow Y$, the calculation formula of confidence degree [9] is:

$$CON = \frac{|X \cap Y|}{|X|}, \quad (2)$$

where CON stands for the confidence level of the association term set, $|X \cap Y|$ stands for the number of records containing two items at the same time, and $|X|$ is the number of records containing the item. If the confidence level of the association item set exceeds the

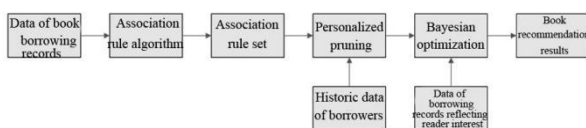


Figure 2: The process of the association rule recommendation algorithm improved by Bayesian network.

threshold, the association item set is considered as a strong rule. The confidence degree of all the items in the frequent item set are calculated as above, and the strong association rule is selected as the reference of book recommendation.

2.2 Improvement of association rule algorithm by Bayesian network

In order to make up for the shortcomings of the association rule algorithm, the association rule algorithm was improved by Bayesian algorithm [10]. The basic steps are as follows. Firstly, a training set is established, and the conditional probability estimation of different characteristic attributes of items to be classified in every classification is counted. Secondly, the probability of belonging to a classification is calculated according to characteristic attributes of the item to be classified [11]:

$$P(Y_i|X) = \frac{P(X|Y_i)P(Y_i)}{P(X)}, \quad (3)$$

where $P(Y_i|X)$ stands for the probability of item X to be classified belonging Y_i , X represents a set of some borrowed books in the historic record of borrows, Y_i indicates the set of some kind of recommended books obtained according to X , i.e., the probability of book X being classified to book Y_i or the establishment probability of association item set $X \Rightarrow Y_i$ after Bayesian calibration, and $P(X|Y_i)$ stands for the distribution probability of X in Y_i , whose value is obtained by estimating the conditional probability of X by the training set.

Thirdly, the probability of X belonging to Y_i is calculated using equation (3), and the set with the largest probability is the most possible association item set.

The association rule set which is calculated by the association rule algorithm is optimized by Bayesian algorithm, and the book recommendation result is obtained according to its probability. The basic flow is shown in Figure 2.

Firstly, the data of book borrowing records in the library are input, and then the association rule set is summarized using the association rule algorithm described above from the borrowing records.

② After obtaining the association rule set, in order to obtain the personalized book recommendation, the association rule set is pruned based on the historical data of the borrower [12]: the items in the association rule set are compared with the items in the historical data, and the record is deleted if the difference is smaller than the set threshold value. The calculation formula of the threshold value is:

$$N = \frac{\text{count}(S_i)}{\text{count}(H_{user})}, \quad (4)$$

where N is the set threshold and $\text{count}(S_i)$ and $\text{count}(H_{user})$ are the number of items in the frequent item set and the number of borrowing records.

③ Through practical investigation, the interest tendency of borrowers to different books in the borrowing records are confirmed, so as to build the borrowing record database [13] which reflects the interest of readers, i.e., the training set of Bayesian algorithm. After the training of

Bayesian algorithm, the association rule set is calibrated after personalized pruning. Finally, the book is recommended according to the probability obtained after the calibration of Bayesian algorithm.

3 Simulation experiment

3.1 Experimental environment

In this study, the above recommended algorithm was simulated using MATLAB software [14]. The experiment was carried out in a laboratory server. The configuration of the server was Windows 7 operating system, 16 G memory and Core i7 processor.

3.2 Experimental setup

First of all, the experimental data used for the simulation experiment came from the book borrowing management system of a university library. Taking 1000 students as subjects, the borrowing records of them from freshmen year to senior year were collected, and then the preliminary processing was carried out, including deleting the records with less than 7 books borrowed (it will reduce the amount of samples, leading to a large contingency in the association rule summarized by the algorithm, deleting the useless fields in the records, such as name and gender of borrows, book author, etc., deleting the invalid data records. There were 26525 borrowing records after final processing, and some records after pretreatment are shown in Table 1.

Library card No.	Borrowing grade	Major disciplines	Book type
A12045	Freshman	Law	D90; D92; D923; D924;
B21541	Sophomore	Economics	F03; F05; G114; G411;
B22548	Junior	Mathematics	O1; O4; P3; Q2;.....
A12365	Senior	Medicine	R4; R75; Q3;.....
.....

Table 1: Some book borrowing records after pretreatment.

Some of the borrowing records after pre-processing are shown in Table 1. Only the library card number which represents the identity of the borrower, the borrowing grade which represents the borrowing time, the major of the borrower and the type of books borrowed by the borrower were left in the borrowing records. Taking the record in the first row of Table 1 as an example, a student whose library card number was A12045 and who was major in law borrowed “D90;D92;D923;D924” books, and most of the books was about law.

In the process of iterative induction of frequent itemsets, the support and confidence degrees of the traditional and improved association rule recommendation algorithms were set as 0.1 and 0.5 respectively, and the

final number of recommended books was set as 5; the data set which was used for training Bayesian algorithm the improved association rule algorithm was the borrowing record which was constructed after investigation and could reflect the interest of readers.

3.3 Evaluating indicator

In this study, the recommendation effect of the recommendation algorithm was evaluated by the accuracy, recall and F value, and their formulas are:

$$\begin{cases} P = \frac{\sum_{i=1}^M L_i}{M \cdot N} \\ R = \frac{\sum_{i=1}^M \frac{L_i}{M \cdot P_i}}{R+P} \\ F = \frac{2RP}{R+P} \end{cases} \quad (5)$$

where L_i is the number of recommended books in line with readers’ interests, M is the number of readers, N is the total number of recommended books, and P_i is the number of books that the reader is interested in.

3.4 Experimental results

The borrowing records obtained after pre-processing were calculated using three algorithms, and finally the book recommendation results of different people were obtained. Limited by the length, this paper only shows some recommendation results, as shown in Table 2. It was seen from Table 2 that the recommendation results of the three algorithms were different for the same person. According to the book classification number, it was found that the books recommended by the collaborative filtering recommendation algorithm were mostly irrelevant, although there were books related to the major; only one or two books recommended by the traditional association rule algorithm were irrelevant; the books recommended by the improved association rule algorithm were basically relevant to the major. The vertical comparison of the recommendation results of different people under the same algorithm showed that the result types under the collaborative filtering algorithm were messy and nearly involved all the majors; the results under the traditional association rule algorithm had overlapping, i.e., high similarity; the results under the improved association rule algorithm involved different types, but different from the collaborative filtering algorithm, they were relevant to the major of the borrower.

The recommendation results of the three recommended algorithms were counted and checked with the corresponding borrower to see if the book was what he was interested in or needed. The final results of the performance of the algorithms are shown in Figure 3. The accuracy of the collaborative filtering algorithm was 67.3%, the recall rate was 72.1%, and the F value was 69.6%; the accuracy of the traditional association rule algorithm was 89.6%, the recall rate was 89.1%, and the F value was 89.3%; the accuracy of the improved association rule algorithm was 98.2%, the recall was 98.3%, and the F value was 98.2%. It was seen from Figure 3 that the improved association rule algorithm had the highest accuracy rate and recall rate, followed by the

traditional association rule algorithm and the collaborative filtering algorithm, indicating that the improved association rule algorithm could provide users with more accurate recommended books; the improved association rule algorithm also had the largest F value, followed by the traditional association rule algorithm and collaborative filtering algorithm. F value is the combination of accuracy and recall rate, which can reflect the personalized recommendation level of the algorithm to different users. The traditional association rule algorithm started from the connection between different book items and used the connection to speculate users' needs. Although personalized pruning was applied, the traditional association rule algorithm was also based on the whole borrowing record, and the strong rule still reflected the overall trend; the improved association rule algorithm used the trained Bayesian algorithm for calibration and optimization to further reflect the demand tendency of different people, therefore the accuracy, recall rate and F value of its personalized recommendation results were larger.

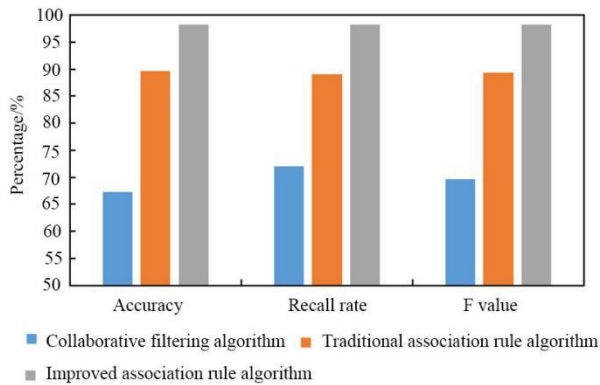


Figure 3: The performance of three recommended algorithms.

4 Conclusion

This paper introduced a recommendation algorithm which mined association rules in borrowing records and improved it with a Bayesian algorithm. Then, borrowing records of 1000 students in the library management system of a university were simulated using MATLAB software. The results are as follows. (1) For the same person, the recommendation results of three algorithms were different: there were many kinds of recommendation results under the collaborative filtering algorithm, only one or two of which were related to the major; there were many kinds of recommendation results under the traditional association rule algorithm, but most of them were related to the major; the recommendation results under the improved association rule algorithm were basically related to the major. (2) Under the same algorithm, the recommendation results for different people were also different: under the collaborative filtering algorithm, the types of recommendation books for different people were diverse; under the traditional association rule algorithm, the types of recommendation books for different people overlapped to a certain extent; under the improved association rule algorithm, the

Library card No.	Recommendation results of collaborative filtering	Results of traditional association rule	Results of improved association rule
A12045	D92;D923;D924;O1;I253.1	D92;D923;D924;I253.1;H1	D923;D924;D923.6;D99;D90
B21541	F05;D99;G20;F12;G114	F05;G114;D923;D924;F12	F03;F05;G114;G411;F12
B22548	O1;O4;P3;F12;H1	O4;P3;G114;D923;D92	O1;O4;P3;Q2;P2
A12365	G20;F12;D923;G114;I253.1	R4;R75;D92;D923;G114	R4;R75;Q3;R8;Q5

Table 2: Some recommendation results of three algorithms.

recommendation books for different people were related to their respective majors, with a low degree of overlap. (3) The results of the objective evaluation showed that the improved association rule algorithm had the largest accuracy, recall rate and F value, followed by the traditional association rule algorithm.

5 References

- [1] Zhou Y (2020). Design and Implementation of Book Recommendation Management System Based on Improved Apriori Algorithm. *Intelligent Information Management*, 12(3), pp. 75-87. <https://doi.org/10.4236/iim.2020.123006>
- [2] Zhang FL (2016). A Personalized Time-Sequence-based Book Recommendation Algorithm for Digital Libraries. *IEEE Access*, pp. 1-1. <https://doi.org/10.1109/ACCESS.2016.2564997>
- [3] Kim JY (2015). A Comparative Study of Pre-service Teachers with Korean Language Education Majors in Book Recommendation Criteria for the Middle and High School Students. *journal of research in reading*, 36, pp. 201-234.
- [4] Zhang F (2016). A Personalized Time-Sequence-Based Book Recommendation Algorithm for Digital Libraries. *IEEE Access*, 4, pp. 1-1. <https://doi.org/10.1109/ACCESS.2016.2564997>
- [5] Sohail SS, Siddiqui J, Ali R (2018). Feature-Based Opinion Mining Approach (FOMA) for Improved Book Recommendation. *Arabian Journal for Science & Engineering*, (2), pp. 1-20. <https://doi.org/10.1007/s13369-018-3282-3>
- [6] Chahinez B, Patrice B (2015). Information Retrieval and Graph Analysis Approaches for Book Recommendation. *Scientific World Journal*, 2015, pp. 1-8. <https://doi.org/10.1155/2015/926418>
- [7] Joa JH, Bangb SW, Parka GD (2016). Implementation of a Recommendation System Using Association Rules and Collaborative Filtering. *Procedia Computer Science*, 91, pp. 944-952. <https://doi.org/10.1016/j.procs.2016.07.115>

- [8] Ping H (2015). The Research on Personalized Recommendation Algorithm of Library Based on Big Data and Association Rules. *Open Cybernetics & Systemics Journal*, 9(1), pp. 2554-2558. <https://doi.org/10.2174/1874110X01509012554>
- [9] Gabroveanu M (2015). Recommendation System Based On Association Rules For Distributed E-Learning Management Systems. *Acta Universitatis Cibiniensis*, 67(1). <https://doi.org/10.1515/aucts-2015-0072>
- [10] dos Santos FF, Domingues MA, Sundermann CV, de Carvalho VO, Moura MF, Rezende SO (2018). Latent association rule cluster based model to extract topics for classification and recommendation applications. *Expert Systems with Application*, 112(DEC.), pp. 34-60. <https://doi.org/10.1016/j.eswa.2018.06.021>
- [11] Gao Y, Xu A, Hu JH, Cheng TH (2017). Incorporating association rule networks in feature category-weighted naive Bayes model to support weaning decision making. *Decision Support Systems*, 96, pp. 27-38.
- [12] Xiao S, Hu Y, Han J, Zhou R, Wen JQ (2016). Bayesian Networks-based Association Rules and Knowledge Reuse in Maintenance Decision-Making of Industrial Product-Service Systems. *Procedia CIRP*, 47, pp. 198-203. <https://doi.org/10.1016/j.procir.2016.03.046>
- [13] Rao W, Zhu L, Pan S, Yang P, Qiao J (2019). Bayesian Network and association rules-based transformer oil temperature prediction. *Journal of Physics Conference*, 1314, pp. 012066. <https://doi.org/10.1088/1742-6596/1314/1/012066>
- [14] Siddiquee MR, Rahman S, Chowdhury SUI, Rahman MR (2016). Association rule mining and audio signal processing for music discovery and recommendation. *International Journal of Software Innovation*, 4(2), pp. 71-87. <https://doi.org/10.4018/IJSI.2016040105>

