

Research of Intelligent recommendation system based on the user and association rules mining for books

Wei Ji^{1, a}, Shi Liu^{1, b}, Yannan Song^{1, c} and Ji Qi¹

¹School of Computing Science, Inner Mongolia University, Hohhot 010000, China;

^ajwyg1992@163.com, ^bliushi@imu.edu.cn, ^c512849571@qq.com.

Keywords: Based on the user, collaborative filtering, association rules mining, books recommended.

Abstract. The increasing number of colleges and universities library books makes users' difficulty of choosing interested books becoming much higher. This paper proposed the research of Intelligent recommendation system based on the user and association rules mining for books. The model integrates the advantages of the collaborative filtering algorithm based on user, and uses association rules to produce recommended list. The experimental results show that this model can produce good recommended results.

1. Introduction

With the increasing of universities' library resources, and comparing to the past passive learning, active learning becomes a mainstream trend now. However, facing that so many library resources, how to choose resources becomes a major problem of college students. Effecting of universities' library is not only providing students with books, but also providing realistic books to students, and helping students to rational use the library resources.

The concept of recommendation system in the 90s of last century is proposed by the United States firstly. After several years of researching, the recommendation system gradually began to use in the commercial areas. And now recommendation system is almost everywhere. It is estimated that the recommendation system improved 20% of product sales at least. In the book recommendation system, the most representative of the site is Amazon Books. The recommendation system is an intelligent platform built on the massive data mining to help site's users to solve their personalized decision support and information services. This paper presents a book recommendation system, introduced the factors of college, grades, books' borrowing records and so on. And it realized a book recommendation system based on the user's collaborative filtering algorithm and Association rules.

2. Related work

Comparing with traditional information retrieval systems, the recommendation system is to provide the user with a personalized recommendation service system. It not only can provide the user with better product recommended, but also can provide the information manager with the proposal of products or services. So it has been widely used in many fields. In the field of book recommendation, scholars have proposed different ways to create a personalized book recommendation system. For example, Ding Xue used the association rule of data mining to create books intelligent recommendation system [2], Dong Kun proposed personalized book recommender system of university library based on collaborative filter [3], Li Mo designed hybrid recommendation system model based on tags and association rules mining for books [7]. Now, there are recommendation system based on association rules, based on the user's collaborative filtering algorithm and based on the content's collaborative filtering algorithm.

Recommendation system based on association rules is to analyze the relation between books and others, find their association, and recommend books to the reader. The basic idea is that the reader will choose the book that is related to books they chose. Recommendation system based on

association rules is divided into two steps: first step is to generate frequent item sets, and select books meet requirements. This step determines the overall efficiency of mining; second step is to extract strong association rules, and search books that is interested by the reader. This step is to recommend books to the reader [2]. The disadvantage of this method is that association rules require a lot of time to extract data and if database is big, it will need a lot of time to analyze.

Recommendation system based on the user's collaborative filtering algorithm is to analyze the relation between the reader and others, find other reader favorite books, and recommend the books to the reader. The basic idea is that the reader will choose same book that is chosen by others with similar interests. The core of this algorithm is that select readers who is similar to the reader (it directly affects the accuracy of this algorithm) and recommend books to the reader. The disadvantage of this method is that it exists sparsity and cold-start problem. With the increasing number of readers and books, it is difficult to find readers who is similar to the reader and a new book is hardly recommended [4,5].

Recommendation system based on the content's collaborative filtering algorithm is to analyze the contact and attributes of books, combine the book that the reader scan to produce recommendations. The basic idea is that reader will choose the similar book that he has scan. This algorithm is to run from the point of content of books, need the system to analyze the contents of the book (can also use tags of the book). The disadvantage of this method is that it exists sparsity (tags of books is so little) and the calculation is high complexity [5,6].

This paper draws on the results of the above and proposes intelligent recommendation system based on the user and association rules mining for books. Due to the readers who is in same college would select similar books, this paper pay attention to the influence of around the reader. And finding a similar book recommends the reader through the data mining association rules. Finally, the recommended results are analyzed[7].

3. Intelligent recommendation system based on the user and association rules mining for books

3.1 Model based on user recommendations and generating neighbor nearest the reader.

Model based on user recommendations is that from the reader's perspective, it recommends books to the reader by around people's choice. Here, we have to make the results of data reliable by choosing around people who is in a reader's college, find the same interest readers in these people and generate neighbor nearest the reader. This model can solve cold-start problem in recommendation system. Facing with newly registered users, general recommendation system can't recommend suitable books to readers because they did not produce any data. But this model can get the college where the reader is in and recommend books which is interested by his classmates to the reader. So we can obtain higher quality recommended results when we do not have the reader's behavior records.

So, how to draw interest similarity between the reader and others? We obtain the data by calculating the cosine similarity between two readers. If the coincidence (the two readers like these books' collection) is high, then the two readers have high interest similarity. The formula is as follows [8]:

$$w_{uv} = \frac{|N(u) \cap N(v)|}{\sqrt{|N(u)| \cup |N(v)|}} \quad (1)$$

in the formula, $N(u)$, $N(v)$ refer to a collection of the reader u and the reader v like books. This formula's mean is that if the two readers like more similar books, then they have high interest similarity. By calculating the interest similarity between the reader u and other readers, we get the readers nearest neighbor (these are some readers who have the highest interest with the reader u).

3.2 Model based on based on association rules.

Association rule mining task is to find all meet the requirements of the support and confidence, such as association rules $A \rightarrow B$. The basic method is to find favorite books of readers' students,

then generate a recommendation list to the reader [1]. Algorithm based on association rules (also known as Apriori algorithm). This algorithm is divided into two steps:

First step is to generate all frequent item sets. The frequency of occurrence of frequent item sets is great than minimum support(min_sup). The formula is as follows [8]:

$$\text{support}(A) = \frac{\text{support_count}(A)}{\text{support_count}(U)} \quad (2)$$

in the formula, support_count(U) is all transactions, support_count(A) is all transactions included item A.

Second step is to extract strong association rules. This step requires rules not less than minimum support(min_sup) and minimum confidence(min_conf). The formula is as follows [8]:

$$\text{confidence}(A \rightarrow B) = \frac{\text{support_count}(A|B)}{\text{support_count}(A)} \quad (3)$$

in the formula, support_count(A|B) is transactions included item A and item B, support_count(A) is all transactions included item A.

3.3 The book intelligent recommendation system work process.

The specific processes of recommended procedure are as follows:

- The user login Books inquiry system.
- Interrogator based on the user obtain groups that are similar to the user by personal information of the user.
- Association rule builder combine groups that are similar to the user with the mining results of association rules to produce recommended list.
- We recommend these books of recommended list to the target user.

4. Experiment and analysis.

4.1 The obtain of data and pre-processed.

This paper uses borrowing records of Southern Campus Library of Inner Mongolia University in March 2015 to July 2015. We removed some interference records, such as borrowing records of library interior, borrowing records of non-school students and so on. Then we obtained 27864 records. These records include identification number of readers, units (college), borrowing date, books barcode number, title, call number and other information.

| 1 | 证件号 | 单位 | 借书日期 | (应)还书日期 | 财产号 | 书刊条码 | 题名 | 作者 | 出版社 |
|----|-----------|-----|------------|------------|----------|----------|-------|-----------|------|
| 2 | 012111012 | 法学院 | 2015-05-22 | 2015-06-01 | 01207274 | 01207274 | 法治是什么 | 李贵连著 | 广西师范 |
| 3 | 012111168 | 法学院 | 2015-04-01 | 2015-04-01 | 01212555 | 01212555 | 法老归来 | 马兆锋编著 | 北京工业 |
| 4 | 012111168 | 法学院 | 2015-04-01 | 2015-04-01 | 01087129 | 01087129 | 法律人的 | 洪果著 | 中国民主 |
| 5 | 012111168 | 法学院 | 2015-05-22 | 2015-06-01 | 01095032 | 01095032 | 刑事法律 | 顾永忠主 | 北京大学 |
| 6 | 012111168 | 法学院 | 2015-05-22 | 2015-05-22 | 01028979 | 01028979 | 中国法律 | 沈红卫著 | 湖南人民 |
| 7 | 012111168 | 法学院 | 2015-05-22 | 2015-05-22 | 01053500 | 01053500 | 社区矫正 | 刘志伟, 王 | 中国人民 |
| 8 | 012111168 | 法学院 | 2015-05-22 | 2015-05-22 | 01041165 | 01041165 | 法律移植 | 何勤华, 王 | 北京大学 |
| 9 | 012111168 | 法学院 | 2015-05-22 | 2015-05-22 | 01029895 | 01029895 | 苏格拉底 | (古希腊) | 华夏出版 |
| 10 | 012111168 | 法学院 | 2015-06-01 | 2015-06-01 | 01034723 | 01034723 | 礼与法 | 法马小红著 | 北京大学 |
| 11 | 012111173 | 法学院 | 2015-03-01 | 2015-03-01 | 01057945 | 01057945 | 谁的青春 | 刘同 | 中信出版 |
| 12 | 012111173 | 法学院 | 2015-04-01 | 2015-04-01 | 01209760 | 01209760 | 异世淘宝 | 柳暗花溟 | 花山文艺 |
| 13 | 012111173 | 法学院 | 2015-04-01 | 2015-04-01 | 01209819 | 01209819 | 异世淘宝 | 柳暗花溟 | 花山文艺 |
| 14 | 012111173 | 法学院 | 2015-04-01 | 2015-04-01 | 01073405 | 01073405 | 晨曦 | 黎明海著 | 花山文艺 |
| 15 | 012111176 | 法学院 | 2015-05-22 | 2015-05-22 | 01046765 | 01046765 | 权力与荣 | Graham, (| 上海译文 |
| 16 | 012111177 | 法学院 | 2015-06-01 | 2015-06-01 | 01089887 | 01089887 | 新中国犯 | 康树华编 | 北京大学 |
| 17 | 012111177 | 法学院 | 2015-06-01 | 2015-06-01 | 01060489 | 01060489 | 犯罪心理 | 主编梅传 | 法律出版 |
| 18 | 012111177 | 法学院 | 2015-03-01 | 2015-03-01 | 01060865 | 01060865 | 外国宪法 | 主编胡锦 | 法律出版 |
| 19 | 012111177 | 法学院 | 2015-03-01 | 2015-04-01 | 01094743 | 01094743 | 企业与公 | 甘培忠著 | 北京大学 |

Fig. 1 Borrowing records

4.2 Data processing.

We assume student ID called 0121112016 to login to the library. And we calculated interest similarity between him and the students who is in the same college.

The following are top 10 readers' information of highest interest similarity.

Table 1 The ranking of interest similarity

| Student ID | Similarity |
|------------|------------|
| 0121112016 | 1 |
| 0121112083 | 0.2928 |
| 0141114447 | 0.2928 |
| 31404015 | 0.2108 |
| 31404041 | 0.2070 |
| 0121111685 | 0.1721 |
| 31404107 | 0.1633 |
| 0121130424 | 0.1491 |
| 0131131659 | 0.1491 |
| 31404119 | 0.1432 |
| 31304131 | 0.1380 |

The code in matlab is as follows [9,10]:

for k=1:455

$d(k,1)=(\text{numel}(\text{intersect}(c(1,:),b(k,:)))-1)/((\text{numel}(\text{find}(b(k,:)))*\text{numel}(\text{find}(c(1,:))))^{(1/2)})$

end

When done experiments, we removed these records that borrowing record less than 5. Because these record would interfere with results of association rules and affect the final results.

4.3 Association rule mining results.

In this part, we used weka 3.6 version and selected Apriori algorithm.

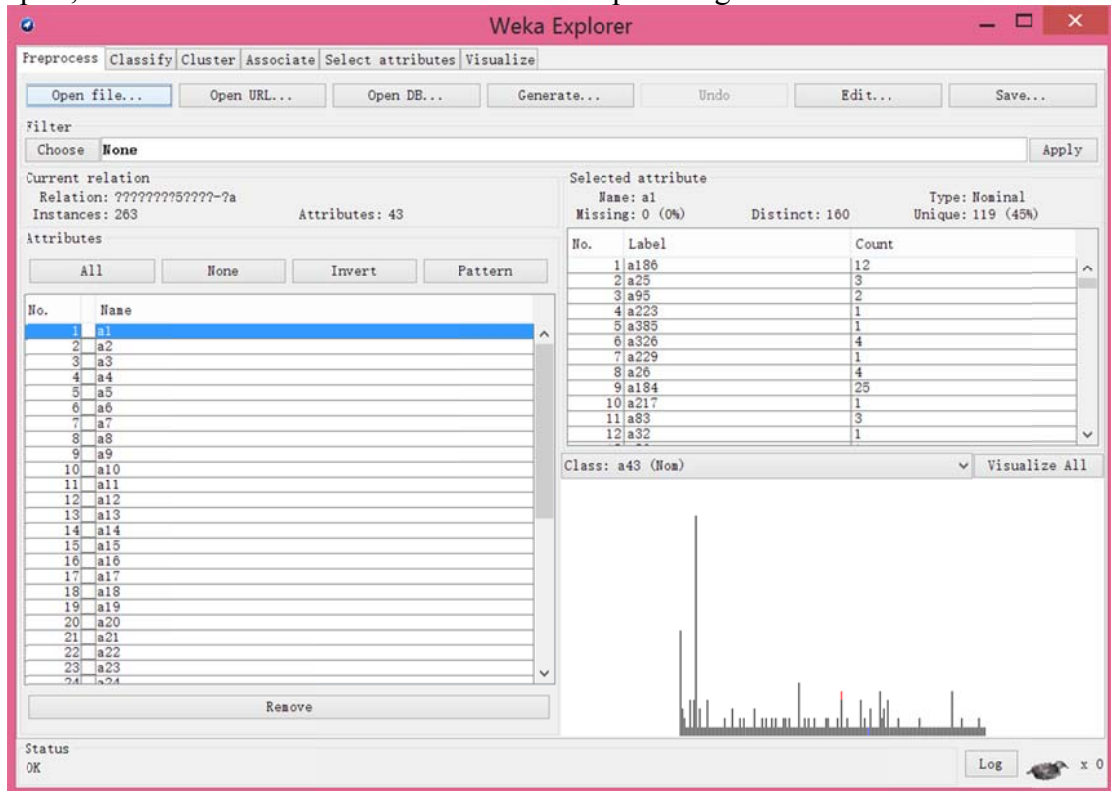


Fig. 2Import data

These data include 43 properties (43 books), 263 borrowing records. Mining results are as follows:

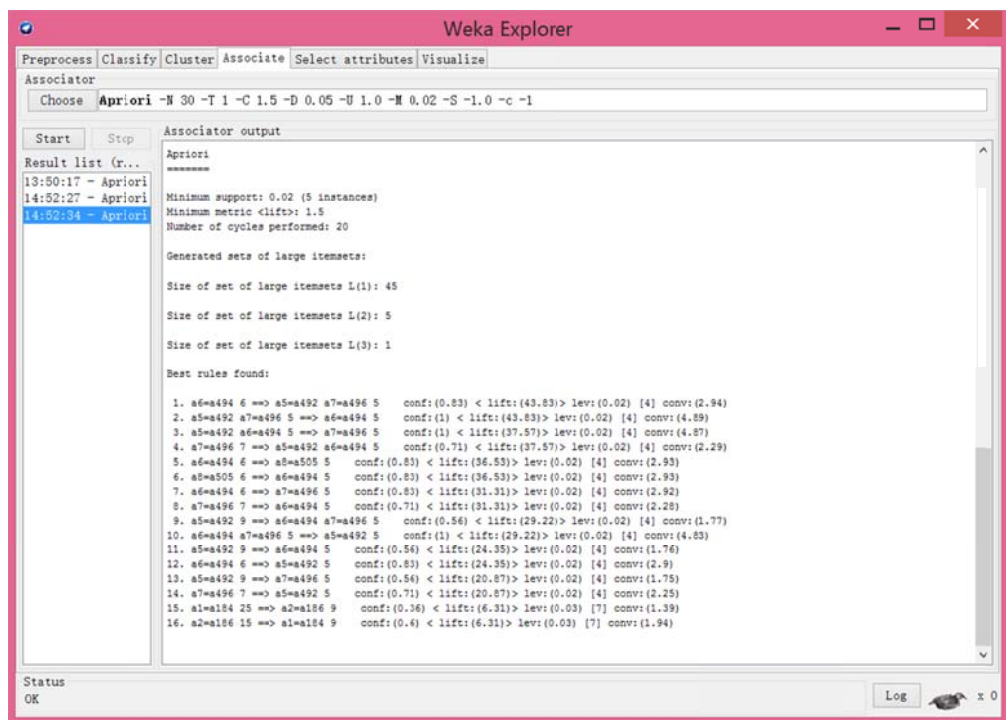


Fig. 3 The results of mining

Experiment got some rules, such as $D923.64 \Rightarrow D923.61, D923.65$ $\text{conf}(0.83)$

$D923.61, D923.65 \Rightarrow D923.64$ $\text{conf}(1)$

$D90-03 \Rightarrow D90$ $\text{conf}(0.6)$ and so on.

These rules show that if students borrow Book Id $D923.64$, they will borrow $D923.61$ and $D923.65$, if students borrow $D923.61$ and $D923.65$, they will borrow $D923.64$, and if students borrow $D90-03$, they will borrow $D90$.

4.4 The mining results of combining Association rule with users.

We combined the mining results of association rule with top 10 readers borrowing records, get the following information:

Table 2 The number of books

| Book ID | Number |
|----------|--------|
| D90 | 7 |
| D90-03 | 3 |
| D923.64 | 2 |
| D923.65 | 2 |
| D923.901 | 1 |

Producing this recommendations list:

Table 3 Book ID and the title

| Book ID | Book Title |
|----------|--|
| D90-03 | 《Legal Methodology》 |
| D90 | 《Legal Transplantation Theory》 |
| D923.65 | 《Contract law judicial interpretation understanding》 |
| D923.64 | 《Insurance contract》 |
| D923.901 | 《Marriage and Family Law》 |

When this paper chooses the nearest neighbor, we only selected these students in the same college, but we did not consider other colleges. Because other colleges could have higher interest similarity's students. We will do the work in this area.

5. Conclusion.

With the acceleration of informatization construction of university library in China, it is an important task of library construction that providing users with high quality personalized books recommended service. This paper proposes a book intelligent recommendation system based on user and association rules, uses MATLAB to data preprocessing and calculation part of the collaborative filtering algorithm, uses Weka to complete the recommendation part of the book, excavates the relationship in borrowing records and generates the recommended list of books for the user. It provides a new thought for research and practice of library personalized service. However, this research is only a small part of the digital library construction. The future research work will set a more reasonable association rule and its threshold, enhance the accuracy of recommendation system, meet the more personalized service of digital library.

References

- [1]. Jie Li, Yong Xu, Yunfeng Wang, et al. Recommended for Personalized strong association rule mining. Systems Engineering Theory and Practice. Vol. 29 (2009) No. 8, p. 146-152.
- [2]. Xue Ding. Books intelligent recommendation system based on association rules. Library and Information Forum. Vol. 1 (2010) No. 4, p. 48-52.
- [3]. Kun Dong. University Library Book recommendation system based on collaborative filtering algorithm study. Modern Information. Vol. 1 (207) No. 11, p. 44-47.
- [4]. Shouyi Sun, Wei Wang. Collaborative filtering based on user clustering personalized book recommendation system. Modern Library and Information Technology. Vol. 1 (2011) No. 11, p.139-142.
- [5]. J.L.Herlocker, J.A.Konstan, L.G.Terveen, et al. Evaluating collaborative filtering recommender systems [J]. ACM Trans.Information Systems, Vol.1(2004) No.22,p.5- 53.
- [6]. Konstan J, Miller B, MaltzD, et al. GroupLens: applying collaborative filtering to usenet news [J]. Communications of theACM. Vol.40(1997) No. 3,p.77- 87.
- [7]. Mo Li, Yongquan Liang. Model tag recommendation system based on a combination of books and association rule mining. Application Research of Computers. Vol. 31 (2014) No. 8, p.2390-2393.
- [8]. Liang Gong, Yi Chen, Yi Wang. Recommended system Practice. People Post Press, 2012, p.44-47.
- [9]. Lei Zhang, Liangying Guo, BingCong. MATLAB Tutorial. People Post Press, 2014, p. 15-24.
- [10]. Guosheng Zhang, Xiang Yu, Jian Wang. MATLAB complete study manual. Tsinghua University Press, 2015, p. 46-76.