# A Book Recommendation Algorithm Considering User Development

Hu Daiping Antai College of Economics & Management, Shanghai Jiao Tong University

## ABSTRACT

There commonly exists the problem of unsatisfactory recommendations given by book recommendation systems. Aiming at the developing characteristic of library user group in university which reflects the consistency between the book borrowing sequence and the accumulation of knowledge, the paper introduces the book lists borrowed by similar users in user group with different development stage to the application of collaborative filtering algorithm in order to increase effective information contained by the whole recommendation process. Consequently, the consistency of the recommendation results and the actual needs of users is improved while the performance of recommendation system rises. The effectiveness of the introduction of user development stage is shown and proved by experiments.

#### **General Terms**

Book Recommendation, Recommendation Algorithm,

#### **Keywords**

Book Recommendation, Recommendation Algorithm, User Development, Academic Library

### **1. INTRODUCTION**

Book recommendation system is a kind of information filtering mechanism which can reduce the searching cost of users in the field of books. With the rapid increase in the number of modern books, an effective bibliographic recommendation mechanism can help readers to save time and energy in choosing books, so that readers can focus on the content of books.

It's showed based on Li Min's [1] investigation and study that the book recommendation system in colleges and universities of China has been generally established and put into use. However, the recommended results from the recommendation system are not satisfactory. The results show that about third of three readers believe that there is a need for the recommendation system, while the readers generally do not have a high degree of satisfaction with the recommendation results, and there are privacy protection problems in the recommendation system. There are many reasons for this imperfection situation, however, the of library recommendation system is one of the most important reasons [2].

Most of the existing studies are to propose new ways of book recommendation in order to improve the efficiency of algorithms, and carry out research on the current situation of recommendation or research on satisfaction. Xie Linhui [3] put forward a special frame model for the book recommendation system in colleges and universities. Dong Kun [4] made use of the relevant factors that affect readers' demand for information to establish reader characteristic model by means of collaborative filtering in order to better Tang Ming Antai College of Economics & Management, Shanghai Jiao Tong University

excavate readers' demand from many angles and deep levels. Yao Wang [5] has designed a recommendation system that can recommend books to students based on different majors and different learning levels. Liu Kai et al. [6] put forward the H3W theoretical framework of personalized recommendation by taking into account the situation, users and resources. Xu Jiali and Chen Jia [7] improved the Apriori algorithm (Shopping basket algorithm) to better excavate the relationship between books from readers' borrowing records. Dou Lingyuan et al. [8] solved the problem of data sparsity by using tag information to construct the characteristic matrix of users and resources; through the Zhang Wenhua's [9] analysis of library books borrowing records by using K-Means model, it is shown that the recommended service can be improved by using the result of book clustering. Li Shuqing et al. [10] have designed an iterative algorithm for evaluating the recommended quality of books in combination with the binary network formed by the relationship among readers.

This study starts from the growing characteristics of the user group of university library, which is consistent with the borrowing behavior and the accumulation of knowledge, study and consider the validity of the recommendation algorithm of user growth information, and the performance of the recommendation algorithm is improved by the introduction of user growth information.

### 2. RECOMMENDATION ALGORITHM CONSIDERING USER DEVELOPMENT

This study considers the recommendation algorithm applicable to university libraries. Compared with the recommendation system in the field of electronic commerce, the book recommendation of university library has its own distinct particularity [1]. First of all, the university library resources cover a wide range, and professional books are relatively large[4]; in addition, the user group of the library is dominated by teachers and students, where the user group is stable, the degree of specialization is higher, and the knowledge structure has a high similarity under some classification (such as, the classification by college) [11]. In the end, the object that the system needs to recommend is relatively simple, and taking the book text as the main recommendation result basically does not involve the audio, the video and other complex data object.

As a group of university library collection resources, college students' own user characteristics can not only be used to classify students from different majors and colleges and to promote referral services in the recommendation system, it can also be seen that the borrowing behavior of books and the accumulation of knowledge are inseparable for college students. The professional nature of books, the order of borrowing before and after and the gradual absorption of users in study can be expected. Therefore, it's possible to consider to use collaborative filtering as the main basic algorithm, and introduce user growth to improve the recommended results based on this algorithm.

The recommendation system is improved according to the characteristics of the user group on the basis of collaborative filtering algorithm. For user  $u_i \in U$  in a set of usersU, it can collate lending behavior  $A_i(b_1, b_2, ..., b_s)$  based on historical data over a selected period of time, in which the  $A_i$  a list of borrowing behaviors, and the  $b_k$  is a list of borrowed books to correspond to a collection of books and form user's behavior feature vector  $R_i(x_1, x_2, ..., x_N)$ . The corresponding value  $x_j$  for each book is the reciprocal of the length of borrowing time to user registration time

$$x_j = \frac{1}{TimeSinceRegister_j} \tag{1}$$

The neighbor user list of user e can be calculated by using cosine similarity (formula 1) through the similarity between users and according to the behavior feature vector.

$$\operatorname{sim}(R_i, R_j) = \frac{R_i \times R_j}{\sqrt{R_i^2} \sqrt{R_j^2}}$$
(2)

After obtaining the  $u_i$  user's list of neighbor users, it can select books borrowed by the neighbor user according to the order before and after the list of users, while the useru\_i does not borrow the books as the recommended result. In this study, the effectiveness of the algorithm is compared and analyzed by using the traditional collaborative filtering algorithm and the difference in recommendation effect between the traditional collaborative filtering algorithm and the collaborative filtering algorithm after processing the information of different growing groups.

The commonly used evaluation criteria of recommendation system have their own measurement methods in terms of accuracy, diversity, user satisfaction and so on in which the diversity of recommended results and the evaluation of user satisfaction need to be obtained in the actual online test, while the accuracy and diversity can be obtained in both the actual test and the off-line experiment of historical data[12]. Taking into account that this study uses off-line data experiments, evaluation indicators in terms of accuracy and diversity are used.

Recommended availability: recommendation effectiveness applies to evaluating the accuracy of the recommendation list. The commonly used indicators are Recall), Precision and F value:

$$\operatorname{Recall} = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |T(u)|}$$
(3)  

$$\operatorname{Precision} = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |R(u)|}$$
(4)  

$$\operatorname{F} = \frac{2 \times \operatorname{Recall} \times \operatorname{Precision}}{\operatorname{Recall} + \operatorname{Precision}}$$
(5)

Where u is the user, U is the user set, R is the actual recommendation list, and T is the test recommendation list, that is, the recommended result. Intuitively speaking, the recall is the proportion of the items hit in the recommended results in the set of recommended results list, Precision is the proportion of items that are recommended hit in the list of recommended results, and the F-value synthesis, recall and precision describe the similarity between the two item list

sets, and can be used to represent the overall performance of the recommendation system.

Recommended coverage: the recommended coverage is suitable for evaluating the diversity of recommendation systems. The commonly used indicator is the Coverage, and formula expression as follows:

$$Coverage = \frac{|\cup_{u \in U} R(u)|}{|I|}$$
(6)

Where I is the whole set of items. Coverage describes how much the recommended results can reach the entire range of items, and greater coverage means that the recommended results are more likely to provide those items that the user has not come into contact with.

The primacy accuracy: The primacy accuracy applies to the accuracy of evaluating the accuracy of the first item in the recommended list as a successful recommendation. The standard commonly used is the Top 10 Accuracy, and the formula is as follows:

$$\text{TopNAccuracy} = \frac{\sum_{u \in U} |T(u)_0 \in R(u)_{0 \sim N}|}{|T|} \quad (7)$$

Top 10 Accuracy indicates the likelihood of success of the first recommendation in the recommendation list, and the practical significance is that the recommendation result in the first place of the recommendation list will enter the user's eye even if the user does not carefully come into contact with the recommendation list, which can be used to evaluate the performance of the recommendation system when the user is only at a glance.

Column inch availability: column inch availability applies to evaluating the effectiveness of a recommendation list, which describes the likelihood that at least one item in a fixed-length recommendation list will be recommended successfully and the formula is as follows:

$$\text{ListNAccuracy} = \frac{\sum_{u \in U} ||T(u)_{0 \sim N} \cap R(u)| > 0|}{|T|} (8)$$

Column inch availability describes whether there is a valid recommendation in the entire recommendation list, and the practical significance is whether the user can get a useful recommendation result after a close look at the recommendation list. When the recommended list is limited in length (the general recommendation list contains 5-20 items) and the number of items is large, whether the entire recommendation list is effective or not is much more practical and meaningful than simply considering the performance of the recommendation system.

## 3. ALGORITHM VERIFICATION AND DATA ANALYSIS

The experimental data used in this study are the historical flow data of Shanghai Jiao Tong University Library from 2013 to 2015, including more than 1.85 million records, among which the number of books is more than 220000 and the number of people who have borrowed books is more than 39,000. Historical data entries can be decomposed into user behavior tables (user's student numbers, book number, borrowing behavior number, borrowing behavior time), user table (user's student number, user type, the college) and book list (book number, ISBN number, the title of a book, author and press).

In the verification phase, the user behavior after the data coverage of the recommendation algorithm is used as the actual demand list of the user and the evaluation index of the recommendation list given by the algorithm. The presentation of a recommendation list has been always based on historical data for a certain period of time, therefore, if the items in the recommendation list are successfully recommended, the user's needs are found and met, and the user subsequently will start to borrow, if not successfully recommended, it means that the user's needs have changed or that the user's needs have not changed, while the books are not within the scope of the requirements. The following algorithm verification stages will first prove that user growth has an impact on the recommendation effect and then verify the performance improvement brought by the recommendation algorithm by considering user growth.

The results of the collaborative filtering algorithm can be summarized as books that the user does not select in a book that is similar to the one chosen by the user. According to the idea of considering the growth stage of users in this study, in terms of the university libraries, it is reasonable to speculate that students registered in the same year (that is, at the same growth stage) are more likely to become the similar users in collaborative filtering. However, for reasons such as college and course requirements, etc, the borrowing behavior of two similar users at the same growth stage does not give each other enough valid information. Only the borrowing behavior of users at a higher stage of growth can be of reference significance. In order to observe the actual situation of this assumption, the recommended results are formed based on the following three ways:

Method I: (Classic collaborative filtering) The number of books borrowed by the user over a period of time is used as a weight to form a certain user's borrowing behavior vector, which can be used to find a similar user and form a list of recommendations using books that are similar to those that a particular user has not borrowed.

Method II: (At the same growth stage): the users are divided into 6 groups according to the time of registration: such as 0-1 /1/2/2/2/3/3/3/4/4/6/100, and the recommended results are given according to the classical collaborative filtering algorithm in each group.

Method III:(Higher stage of growth): According to the different registration time, the users are divided into 6 groups: such as 0-1/1/2/2/2/3/3/4/4/4/6/100 years. Each user and the user who are not in the same group give the recommended results according to the classical collaborative filtering algorithm.

The difference between the above three recommendation methods is only the differentiation of user groups. The results of the recommendations are as follows:

In the method I, II, III, I, II and III to form a recommended list with a length of 10 for January-June 2013, borrowing pipeline data. Using the July 2013 pipeline data as the validation data, the results are as follows:

Method	Top 10 accuracy	F value	10 columns inch availability
1. Classic collaborative filtering	0.021%	0.0399%	0.223%
2.The same stage of growth	0.014%	0.0321%	0.188%
3. Higher growth stage	0.021%	0.0400%	0.223%

Table 1. comparison of recommended results in method I, II, and III

It can be seen that the recommendation result of collaborative filtering among users in the same growth stage is worse than that of classical collaborative filtering, while the recommended results are slightly better than that of classical collaborative filtering from the users in higher growth stage. From the point of view of effective information, the borrowing information of users in the same growth stage cannot give each other enough effective reference information, and the real reference information comes from the users who are at other growth stage.

At the same time, the F value of the recommended results of the three recommended methods is compared.



Figure 1. Similarities in Method I, II, and III

It can be considered that the similarity between method I and method II is 0.67, the similarity between method I and method III is 0.47, and the similarity between method II and method III three is 0.16. Therefore, the recommended result of classical collaborative filtering mainly comes from the information borrowed by users in the same growth stage. To take advantage of the information of the user group at a higher growth stage, it's merged the recommended list of method I with method III to get a recommended list of borrowing pipeline data from January to June 2013 with a length of 20 and to use the July 2013 pipeline data as the validation data, which is as follows:

Method	Top 20 accuracy	F value	20 columns inch availability
Classic collaborative filtering	0.021%	0.0380%	0.383%
The combination of Method I and III and the introduction of a higher stage of growth	0.020%	0.0424%	0.401%

## Table 2. Resultant merger of Method I, II, and III and the comparison of the recommended results

Complement of user information from different growth stages can effectively improve recommendation performance and the validity of the entire recommendation list. Therefore, it can be determined that user growth is a highly important consideration for book recommendation.

In order to get a more general recommendation method it doesn't need to consider the stages of growth and the inherited properties of the relationship between books in getting a recommendation every time, for example, modern algebra always appears after the linear algebra books. As a result, taking the period from the time of user's registration to the time of book borrowing as a weight instead of the number of times for the user to borrow the book, it can reasonably speculate and obtain more accurate recommended results.

In order to observe the actual situation of this assumption, the recommended results are currently formed in the following manner:

Method IV (collaborative filtering with growth as weight): taking the reciprocal of the length of time between the borrowing time of a book borrowed by the user over a period of time and the time of registration to form a lending behavior vector for a particular user as a weight and finding out a similar user with this vector. A recommendation list is formed by the use of books similar to those that a particular user has not borrowed from a book.

Collaborative filtering, weighted by growth, forms a recommended list of borrowing pipeline data from January to June 2013 and using the July 2013 pipeline data as the validation data, and the results are as follows:

Table	3.	Result	table	for	method	IV
rabic	<b>J</b> •	Result	table	101	memou	<b>T</b> A

Method	Top 10 accuracy	F value	10 columns inch availability
Classic collaborative filtering	0.021%	0.0399%	0.223%
4.Collaborative filtering with growth as weight	0.021%	0.0492%	0.278%

In the performance of each performance, the recommended results with the growth as weight are all superior to those of the classical collaborative filtering algorithm and the introduction of user growth information and introduction of sequential information about books greatly improves the accuracy of collaborative filtering algorithm.

Note: similar results can be obtained after validation of the above recommendations with data from 2014 and 2015.

Place Tables/Figures/Images in text as close to the reference as possible (see Figure 1). It may extend across both columns to a maximum width of 17.78 cm (7").

Captions should be Times New Roman 9-point bold. They should be numbered (e.g., "Table 1" or "Figure 2"), please note that the word for Table and Figure are spelled out. Figure's captions should be centered beneath the image or picture, and Table captions should be centered above the table body.

#### 4. CONCLUSION

In view of the characteristics of the growth of the user group in the university library, and by distinguishing the growth stage of the user, the introduction of book information from similar users at different growth stages in the formation of recommendation results can increase the effective information contained in the recommendation results to a certain extent and the collaborative filtering based on growth weight can better match the recommendation results with the users' actual requirements and can obtain better performance of the recommendation system at the same time.

From the point of view of access to information, it's a must to notice that the repeated information contained in the recommendation results for the user is not effectively removed, but only the addition of non-repeated information for the user as valid information to enhance the recommendation result. That is, if the two books are not removed, in the eyes of the user, it is a situation where the same content is included and the mutual exclusion effect is formed in the borrowing behavior. In addition, in subsequent studies, the information contained in the keywords in the title of the book can be used to eliminate the invalid repeated information in the part. The heading of a section should be in Times New Roman 12-point bold in all-capitals flush left with an additional 6-points of white space above the section head. Sections and subsequent sub- sections should be numbered and flush left. For a section head and a subsection head together (such as Section 3 and subsection 3.1), use no additional space above the subsection head.

#### 5. REFERENCES

- Li Min, Wang Yingchun and Liu Yanquan, Investigation on the Recommendation System of Collection Resources in University Library of "211 Project", Library and Information Service, 2016, (9): 55–60.
- [2] [2] Huang Yan, Research and Design of Bibliographic Recommendation System in the Library, Library Research, 2011, 41(2): 92–96
- [3] Xie Linhui, Research on the Application of Recommendation System in University Digital Library, Modern Intelligence, 2006, 26(11): 72–74, 2006.
- [4] Dong Kun, Research on Book Recommendation System of University Library based on Collaborative Fltering Algorithm, Modern Library and Information Technology, 2011, (11): 44–47.

- [5] Yao Wang, Book Recommendation System of University Library for Students, Journal of West Anhui University, 2014, (4): 149–152.
- [6] Liu Kai, Wang Weijun, Huang Yinghui and Fang Wei, "Theoretical Exploration of Personalized Recommendation System: From System to User-Centered Evolution", Intelligence Theory and Practice, 2016, 39(3): 52–56.
- [7] Xu Jiali and Chen Jia, A Fast Personalized Bibliographic Recommendation Method, Modern Library and Information Technology, 2010, 26(2): 79–84.
- [8] Dou Lingyuan, Wang Xinhua, and Sun Ke, Collaborative Filtering Recommendation Algorithm Integarting Tag Feature and Time Context, Small Microchip Systems, 2016, 37(1): 48–52.
- [9] Zhang Wenhua, Hu Chun, Hu Guanglin and Feng Kai,

"Research on Clustering Analysis Based on Library Circulation Data", Journal of Agricultural Library and Information Science, 2010, 22(1): 109–111.

- [10] Li Shuqing, Xu Xia and Xu Minjia, The Measurement Method of Book Referral Quality and the Personalized Book Recommendation Service based on the Binary Network of Readers' Lending and Reading, Journal of Library Science of China, 2013, 39(3): 83–95.
- [11] Li Weihua, Lu Yumin and Mei Hong, Discussion on Personalized Information Recommendation System of Digital Library, Science and Technology Plaza, 2007, (9): 109–110.
- [12] Yang Xu, Research on Collaborative Filtering Algorithm based on User Context Information, Jilin University, 2014.