

Analysis and research on the Algorithm of university book recommendation

-- Taking the Library of Hebei University of Architecture as an example

Likun Jia, Yaohui Wang, Shiyu Yan and Dan Xiao*

Department of Information Management, Hebei University of Architecture, Zhangjiakou, China

*Corresponding author e-mail:519090265@qq.com

Abstract. This paper analyzes the performance of user-based collaborative filtering algorithm and item-based collaborative filtering algorithm in university library lending data and the applicable environment of different algorithms through accuracy and recall rate. The borrowing data of university library are processed by different methods and the processed data are run by different algorithms. The running results of different cutting ratios of training sets and test sets, different recommended quantities and different data processing are recorded. And then look for variables that are relevant. Finally, the best running environment of different recommendation algorithms is found by comparing the data.

Keywords: Accuracy, recall, data analysis.

1. Introduction

Book recommendation algorithm plays an important role in promoting book circulation in university libraries. Personalized recommendation algorithm can improve the utilization rate of books and the service level of libraries^[1] can be enhanced. This paper aims to find the best recommendation environment of the two algorithms by comparing and analyzing the performance of the two traditional collaborative filtering recommendation algorithms in different environments, and then apply them to university libraries.

This paper first analyzes the readers' lending data, then selects the effective dimension^[2] of lending data through multiple algorithm experiments with different division methods. Afterwards, by calculating the similarity between users or items, we can get the recommended results that have higher accuracy rate and records. Finally, analyzing data of the recommendation results under different treatments could draw conclusions that advantages of algorithm implementation. Finally, it compares the two recommendation system models through the corresponding evaluation indexes^[3]. Data preprocessing and algorithm implementation are all run in PyCharm software^[4].

2. The two collaborative filtering algorithms' introductions

2.1 User-based collaborative filtering algorithm

The user-based collaborative filtering algorithm judges the distance between users according to the preference of different users for the same item.

Compute User Similarity Using User 's Historical Behavior. Given user u and user v , $N_{(u)}$ denotes the set of items that user u feedbacks (borrowing records can be understood as borrowing a book), $N_{(v)}$ represents the set of items that user v feedbacks. The interest similarity between user u and user v is calculated by the following cosine similarity:

$$W_{uv} = \frac{|N_{(u)} \cap N_{(v)}|}{\sqrt{|N_{(u)}| \cdot |N_{(v)}|}} \quad (1)$$

According to the above formula, the interest similarity matrix between users is obtained, and the k users with high similarity with the target user are found. The first N bits are selected as the recommendation results to complete the algorithm. The preference degree of user u to item i is calculated by the following formula:

$$p^{(u,i)} = \sum_{v \in S(u,k) \cap N(i)} w_{uv} r_{vi} \quad (2)$$

Among them, represents the K user sets that are closest to the user's interest, $N(i)$ represents the user sets that have behavior on item i (such as borrowed books), r_{vi} represents the similarity between user u and user v, and W_{uv} represents user v's interest in item i [5].

2.2 Item-based collaborative filtering algorithm

The item-based collaborative filtering recommendation algorithm calculates the similarity of items by analyzing the user's behavior records [6]. The algorithm steps are:

Step one: Establish item similarity matrix. The key is to define the method for calculating the similarity of items. In order to compare with the user-based collaborative filtering recommendation algorithm, the similarity calculation formula of chord is still used here:

$$w_{ij} = \frac{|N(i) \cap N(j)|}{\sqrt{|N(i)| \cdot |N(j)|}} \quad (3)$$

$N(i)$ represents the number of users who like item i and $N(j)$ represents the number of users who like item j.

Step two: Generate a recommendation list for the user based on the similarity of the items and the user's historical behavior. User u's interest in an item j is calculated by the following formula:

$$p^{(u,j)} = \sum_{i \in N(u) \cap S(j,k)} w_{ij} r_{ui} \quad (4)$$

Among them, $N(u)$ represents the collection of items liked by user u, $S(j,k)$ represents the collection of the k items that are most similar to item j, w_{ij} represents the similarity between item i and item j, and r_{ui} represents the interest of user u to item i [7].

3. Experimental process and results

3.1 Data description and processing

The experimental data come from the borrowing data of undergraduate students in the library of Hebei University of Architecture for 35 months in 2018, 2019, September-December 2020 and January - July 2021 (The following data for nearly one year are collectively called one year data, and all data are collectively called three years data). Due to the low borrowing amount of a single reader and the large number of books, the accuracy and recall rate of the algorithm are extremely low, the credibility of the results is not high, and the actual research value is not large. Therefore, the reader's license number and the call number are finally retained. Class, professional information are acquired by the reader card number. The call number (book category) is used to replace the specific book as an item [8].

3.2 Algorithm evaluation index

Character definition: $R(u)$: recommendation list (according to the user's behavior on the training set); $T(u)$: List of user behavior on the test set (actual selection) [9].

3.2.1 Precision.

Accuracy rate refers to the ratio of the target user's favorite products to all recommended products in the recommendation list. In this experiment, the ratio of the number of book types that users like in the recommendation list to the number of all book types in the recommendation list is calculated as follows:

$$Precision = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |R(u)|} \quad (5)$$

3.2.2 Recall ratio.

Recall rate refers to the ratio of the product that the target user likes in the recommendation list to all the products that the user likes in the system. In this experiment, recall rate is the ratio of the number of books that the user likes in the recommendation list to the number of books that the user conducts in the list. The calculation formula is as follows:

$$Recall = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |T(u)|} \tag{6}$$

3.2.3 Algorithm operating speed. In this paper

the running time of the algorithm is calculated by inserting a timestamp at the beginning and end of the algorithm.

3.3 Running results of two algorithms in different data processing

3.3.1 Data processing is professional-call number.

Tables 1, 2 and 3 respectively list the changes of user-based collaborative filtering recommendation algorithm and item-based collaborative filtering recommendation algorithm in accuracy, recall rate and running time in the processing mode of 80 %, 75 %, 70 %, 60 % training set and ' year data ', ' 1-year data ', user object is ' professional ', item object is call number. (The data used in the subsequent algorithm operation is the same as this time, and will not be repeated. The result retains three decimal places.)

Table 1. Accuracy Operation Results.

Precision Proportion of test set	one-year data		three-year data	
	User _ based	Item _ based	User _ based	Item _ based
80%	0.765	0.725	0.800	0.600
75%	0.765	0.667	0.800	0.640
70%	0.818	0.727	0.759	0.586
60%	0.759	0.759	0.767	0.700

Table 2. The Operation Results of Recall Ratio.

Recall Proportion of test set	one-year data		three-year data	
	User _ based	Item _ based	User _ based	Item _ based
80%	0.279	0.264	0.251	0.189
75%	0.255	0.222	0.214	0.171
70%	0.213	0.190	0.164	0.127
60%	0.150	0.150	0.125	0.114

Table 3. Running Time.

Time Proportion of test set	one-year data		three-year data	
	User _ based	Item _ based	User _ based	Item _ based
80%	1.59	1.49	8.96	9.18
75%	1.52	1.63	8.94	8.75
70%	1.52	1.71	9.42	8.78
60%	1.44	1.54	8.73	8.58

3.3.2 Data processing is class-call number.

Table 4, Table 5, Table 6 lists the changes of user-based collaborative filtering recommendation algorithm (User _ based) and item-based collaborative filtering recommendation algorithm (Item _ based) in accuracy, recall rate and running time when the user object is ' class ' and the item object is the call number.

Table 4. Accuracy Operation Results.

Precision Proportion of test set	one-year data		three-year data	
	User _ based	Item _ based	User _ based	Item _ based
80%	0.619	0.577	0.669	0.644
75%	0.631	0.641	0.598	0.623
70%	0.595	0.613	0.659	0.643
60%	0.590	0.607	0.662	0.603

Table 5. The Operation Results of Recall Ratio.

Recall Proportion of test set	one-year data		three-year data	
	User _ based	Item _ based	User _ based	Item _ based
80%	0.260	0.242	0.243	0.234
75%	0.267	0.272	0.214	0.223
70%	0.193	0.200	0.178	0.174
60%	0.144	0.149	0.137	0.124

Table 6. Running Time.

Time Proportion of test set	one-year data		three-year data	
	User _ based	Item _ based	User _ based	Item _ based
80%	2.17	1.78	10.86	10.00
75%	2.12	1.76	11.39	10.13
70%	1.78	1.81	10.58	10.03
60%	1.76	1.91	10.05	9.84

3.3.3 Data processing is reader-call number.

Table 7, table 8 and table 9 respectively list the changes of user-based collaborative filtering recommendation algorithm (User_based) and item-based collaborative filtering recommendation algorithm (Item_based) in accuracy, recall rate and running time when the user object is 'reader' and the item object is the call number.

Table 7. Accuracy Operation Results.

Precision Proportion of test set	one-year data		three-year data	
	User _ based	Item _ based	User _ based	Item _ based
80%	0.235	0.277	0.340	0.356
75%	0.229	0.248	0.334	0.353
70%	0.237	0.300	0.336	0.370
60%	0.231	0.281	0.342	0.389

Table 8. The Operation Results of Recall Ratio.

Recall Proportion of test set	one-year data		three-year data	
	User _ based	Item _ based	User _ based	Item _ based
80%	0.189	0.222	0.247	0.259
75%	0.185	0.201	0.236	0.250
70%	0.172	0.219	0.214	0.236
60%	0.155	0.188	0.189	0.215

Table 9. Running Time.

Time Proportion of test set	one-year data		three-year data	
	User _ based	Item _ based	User _ based	Item _ based
80%	7.06	5.15	42.03	33.18
75%	8.00	5.07	44.69	34.33

70%	6.79	5.04	40.68	33.39
60%	6.36	5.34	40.41	32.23

4. Results Analysis and Explanation

4.1 User defined as class and user defined as profession

In terms of accuracy, the results after average processing show that when the user is expanded from class to major, the accuracy is increased by about 10 %. According to the calculation formula of accuracy, the ratio of the target user's favorite products to all the recommended products in the recommended list when the number of recommended results remains unchanged, the larger the range of $T_{(u)}$ in the molecule is, the greater the hit rate of the recommended items is, that is, the accuracy is improved.

Whether three-year data or one-year data, when users change from class to specialty, the accuracy improvement of user-based algorithm is more obvious than that of item-based algorithm, and the former is better than the latter in accuracy.

In terms of running time, when the user is defined as class or specialty, the amount of data changes from small to large, and the increase of item-based collaborative filtering algorithm is smaller than that of user-based collaborative filtering algorithm. Therefore, the speed advantage of the item-based recommendation algorithm is more obvious when dealing with large amount of data or personalized recommendation.

4.2 User defined as class and user defined as individual

When the user definition is changed from class to individual, the accuracy of the algorithm decreases by about 30 %. Due to the small amount of personal borrowing and the books borrowed are too scattered, the range of $T_{(u)}$ is greatly reduced, the intersection of $T_{(u)}$ and $R_{(u)}$ is smaller, the hit rate is lower, and performance on the results is reduced accuracy in the formula.

When the user definition is reduced to individual, compared with the user-based collaborative filtering algorithm, the item-based collaborative filtering algorithm performs is better in accuracy and running time, especially for three-year data. It can be seen that if the amount of data is large, the differentiation advantages will be particularly prominent.

4.3 Global analysis

According to the analysis of the results, the user-based collaborative filtering recommendation algorithm is used to recommend books with the specialty as the user and the call number as the item, and the effect is the best for university libraries. The reasons are as follows:

(1) Compared with the processing method of taking readers as users and taking call number as items, the processing method of ' professional-call number ' does not need to consider the problem of user borrowing. Since the specialty is a collection of "similar readers" in a large range, the user's borrowing is sufficient and the credibility is relatively high when the specialty represents the user. For ' reader-call number ' processing, users borrow less is an unavoidable problem. When calculating the accuracy, it has been around 20 %, which cannot be improved.

(2) Compared with the processing method of taking the class as the user and taking the call number as the item, the specialty is a collection of classes and students in different classes of the same specialty have the same courses and professional knowledge. The ultimate goal of the recommendation algorithm is to recommend books of interest to readers. Whether for class or professional in the recommendation algorithm, the recommended books are interested to readers. When the processing method defined by class as user is replaced by professional as user, the advantage of the algorithm is not only the improvement of accuracy, but also the improvement of algorithm operation speed.

After the above analysis, the most reasonable way of data processing is professional as users, call number as items and 75 % test set.

5. Conclusion

Item-based collaborative filtering algorithm is more suitable for the situation of large amount of data and personalized recommendation than user-based collaborative filtering algorithm. The results show that the item-based collaborative filtering algorithm is superior to the user-based collaborative filtering algorithm in terms of accuracy, recall rate and running time when using three-year data and recommending objects to readers

User-based collaborative filtering algorithm is suitable for class or professional recommendation. Especially when the amount of data is small, user-based collaborative filtering recommendation algorithm can still maintain high accuracy. Compared with item-based collaborative filtering algorithm, the running time is shorter.

For the cold - start problem of the algorithm, the recommended object can be changed into a group, that is, class or specialty. Because college students use more professional books, professional recommendation books can not only solve the problem of cold - start, but also improve the accuracy of recommendation algorithm to a certain extent. The use of class-based processing in data processing can cultivate students' interest in reading and solve the problems of which library books is only collected but not read, and students' difficulty in selecting books.

Acknowledgments

This work was financially supported by Cloud Recommendation-Book Recommendation Applet Based on Collaborative Filtering Algorithm fund. (NO.202151001010544)

References

- [1] Xu Zhengzheng. Research and optimization of user-based collaborative filtering algorithm and system architecture in personalized recommendation system [D]. Shandong: Shandong University, 2017.
- [2] Xiong Huixiang, Li Xiaomin, Li Yueyan. Research on group recommendation based on attribute mining of book reviews [J]. Data analysis and knowledge discovery, 2020,4 (2) : 214-222.
- [3] Zhang Geyi, Zhu Yueqin, Lv Pengfei, etc. Research on book recommendation method coupling collaborative filtering recommendation and correlation analysis [J]. China Mining Industry 2017,26 (S1): 425-430.
- [4] Song Qian. Research on Personalized Recommendation Algorithm of Library Collections [D]. Sichuan: University of Electronic Science and Technology, 2016.
- [5] Sun Xiaojuan; Qian Bo; Wang Shengtao. Design and Implementation of IT Book Recommendation System Based on Collaborative Filtering Algorithm [J]. Software Guide, 2017,16 (4): 70-73.
- [6] Chen Jiawei,Huang Liting,Lin Geng. TV program recommendation strategy of collaborative filtering algorithm [J]. Journal of Henan Institute of Engineering (Natural Science Edition).2020,32 (01): 61-65.
- [7] Liu Yu,Zhu Wenhao. Hybrid recommendation algorithm based on content and label weight [J]. Computer and digital engineering. 2020,48 (04): 77777.
- [8] Ji Wei. Research on the algorithm and model of university library recommendation system [D]. Inner Mongolia : Inner Mongolia University, 2017.
- [9] Geng Lixiao, Jin Gaojie, Li Yahan, etc. Research on personalized recommendation of university library literature resources based on improved content filtering algorithm [J]. Library and information work, 2018,62 (21) : 112-117.