Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 83 (2019) 490–494

www.elsevier.com/locate/procedia

11th CIRP Conference on Industrial Product-Service Systems

College Library Personalized Recommendation System Based on Hybrid Recommendation Algorithm

Yonghong Tian[a,*], Bing Zheng[a], Yanfang Wang[a], Yue Zhang[a], Qi Wu[a]

[a] College of Data Science and Application, Inner Mongolia University of Technology, Huhhot 010080, China

* Corresponding author. *E-mail address:* yonghong_tian@126.com

Abstract

When the number of books provided by library is relatively large, it becomes difficult for user to select appropriate book from a lot of candidate books. In this case, this paper designs a personalized recommendation system for college libraries based on hybrid recommendation algorithm. First of all, paper studies the application of collaborative filtering and content-based recommendation algorithm in the recommendation of university books, which involves reader classification, the establishment of user-item scoring matrix, the construction of vector space model and the calculation of similarity among users. And considering the characteristics of books and readers in universities, the user - item scoring matrix is improved, and clustering is used to alleviate the data sparsity problem. Do comparative experiments using the hybrid algorithm in data sets of Library of Inner Mongolia University of Technology. The results demonstrate that the hybrid methods can provide more accurate recommendations than pure approaches. Finally, the Spark big data platform combined with the hybrid recommendation algorithm is used to achieve the personalized book recommendation system design.

Peer-review under responsibility of the scientific committee of the 11th CIRP Conference on Industrial Product-Service Systems

Keywords: Collaborative Filtering, Content-based Recommendation, Cluster, Spark ;

1. Introduction

Every year, the university library will introduce a large number of books. The amount of collection is increased year by year. Users need to spend a lot of time choosing a book. At the same time, many books are not effectively used, resulting in a waste of library resources. These phenomena are caused by "information overload"[1]. To solve this problem, library needs to rely on information filtering mechanism. The information filtering mechanism is divided into two kinds: search mechanism and recommendation mechanism. The former uses keyword help users quickly find relevant books; the latter automatically recommend books to the users.

Personalized recommendation systems are seek to predict the preference based on the user's interest, behavior and other information. Personalized recommendation is not only can provide the user needs, but also to help users explore and discover new hobbies. The application of recommendation systems in the university library solves the problem of difficulty in choosing books, and improves the utilization rate of library resource[2].

This paper is organized as follow. In section2, the related works of recommender systems is detailed. In section 3, the hybrid recommender based on university students is provided, and collaborative filtering is improved by combining cluster. Hybrid recommender design methods are offered. Meanwhile, the user-item scoring matrix and clustering is used to alleviate the data sparsity problem. Recommended algorithm for detailed comparison of experimental results and discussion are shown in section 4. In section 5, the personalized book recommendation system is designed by using the Spark big data platform, and the recommendation system is displayed through actual operation. In the last section, the conclusions and future works is presented.

2212-8271
Peer-review under responsibility of the scientific committee of the 11th CIRP Conference on Industrial Product-Service Systems.
10.1016/j.procir.2019.04.126

2. Related work

Recommender systems have become extremely common, and are utilized in a variety of areas: some popular applications include movies, books, research articles, and social tags[3]. There are three basic categories of recommendation algorithms: collaborative filtering, content-based filtering, and hybrid recommendation.

Collaborative filtering methods are based on collecting and analyzing a large amount of information on users' behaviors, activities or preferences and predicting what users will like based on their similarity to other users[3]. The proposed algorithm does not need professional knowledge, and the recommendation effect will become better and better with the interest of the user, but there are data sparsity and other problems [3].

Content-based filtering methods are based on a description of the item and a profile of the user's preference. These algorithms try to recommend items that are similar to those that a user liked in the past[4].

Hybrid recommendation is combining collaborative filtering and content-based filtering. Hybrid approaches could be more effective in some cases. These methods can also be used to overcome some of the common problems in recommender systems such as cold start and the sparsity problem.

3. Hybrid recommendation systems

Hybrid approaches has three strategies. The first is by making content-based and collaborative filtering separately and then combining them. The second is by adding content-based capabilities to a collaborative filtering approach (and vice versa).The third is by unifying the approaches into one model [5]. In order to compare the performance of the two algorithms and the hybrid algorithm, the first strategies is adopted in this paper.

3.1. Collaborative filtering algorithm based on university users

User based collaborative filtering algorithm is divided into three steps, the establishment of user model, find the nearest neighbor set and generate recommendations. [6,7]

Suppose we have n users set U = {User1, User2, ... Usern} and book categories set I = {Item1, Item2, ..., Itemm} ; R is expressed as an $N \times M$ matrix, N is the number of users, M is the number of categories. The number of books borrowed by user i in the category j is represented as R_{ij}.

The user-based algorithm calculates the similarity between two users. Similarity computation between users is an important part of this approach. Multiple measures, such as Pearson correlation and vector cosine based similarity are used for this. The user based top-N recommendation algorithm uses a similarity-based vector model to identify the k most similar neighbors to an active user. Vector cosine between user u and v is shown in formulae (1):

$$S_{u,v} = \frac{\sum_{i \in I_u \cap I_v} r_{u,i} \bullet r_{v,i}}{\sqrt{\sum_{i \in I_u \cap I_v} r_{u,i}^2 \sum_{i \in I_u \cap I_v} r_{v,i}^2}} \quad (1)$$

According to the nearest neighbor set U_k, the recommendation list B is generated to make recommendation for the target use. First, the collection of books I_n = {book1, book2, ...} which is evaluated by all the users in the nearest neighbors is obtained, and then the candidate list B_u = {book1,book2, ...} which the target user has borrowed is removed.

3.2. Clustering and Collaborative Filtering

Collaborative filtering approaches often suffer from sparsity problems. [8] The number of university library books is extremely large. As a result, the user-item matrix could be extremely large and sparse, which brings about the challenges in the performances of the recommendation. The most active users will only have read a small subset of the overall database. The sparseness of the user-book matrix is as high as 99.99%.

In order to reduce the sparsity of data, this paper adopts the K-means clustering before calculating the similarity .Clusterin g is the task of grouping a set of objects in such a way that obj ects in the same cluster are more similar to each other than to t hose in other clusters.

The main idea of the K-means algorithm is described as follows:

Step 1. Given an initial set of k means $m_1,\ldots,m_k$.

Step 2. Assign each observation to the cluster whose mean yields the least within-cluster sum of squares.

$$E = \sum_{i=1}^{k} \sum_{x \in m_i} \left\| x - \mu_i \right\|_2^2 \quad (2)$$

Step 3. Calculate the new means to be the centroids of the observations in the new clusters.

$$\mu_i = \frac{1}{\left| m_i \right|} \sum_{x \in m_i} x \quad (3)$$

Step 4. The new centroids compared with the previously calculated centroids, if there is a change, transfer step 2, otherwise transfer step 5.

Step 5. Stop and output the clusters results.

The collaborative filtering algorithm based on university users is described as follows:

Input: target user u; book set B; clustering user characteristics matrix; borrowing records;

Output: a set of books recommended to the target user u;

Improved collaborative filtering flow chart see Fig. 1.

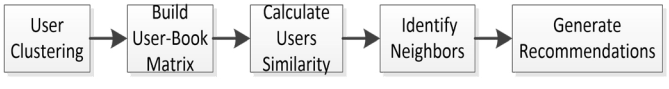

Fig. 1. Improved Collaborative Filtering Flow Chart.

3.3. Content based recommendation algorithm

Content-based filtering methods are based on a description of the book and a profile of the user's preference.[9] This algorithm try to recommend books that are similar to those that a user liked in the past.

First, abstract the features of books. In the library system, the information about the book includes title, classification number, index number, author, publisher, price, place of collection and so on. In this paper, the keywords of title, authors as features of the book.

Second, to create a user profile, the system mostly focuses on two types of information: history of the user's borrowed books and specialize. The profile of the user's preference can be expressed as a set of n-tuples *W:*

$$W_i = \{(w_1, v_1), (w_2, v_2) \ldots\ldots (w_n, v_n)\} \quad (4)$$

W_i denotes the preference of the user i, W_n denotes the feature of the user n, and the weights v_n denote the importance of feature to the user and can be computed from individually rated content vectors using a variety of techniques.

Finally, various candidate items are compared with the books previously borrowed by the user and the best-matching books are recommended.

3.4. Hybrid approach

Hybrid approaches, making content-based and collaborative-based predictions separately and combining them could be more effective in books recommender systems. The architecture of recommender systems shown in Fig. 2.

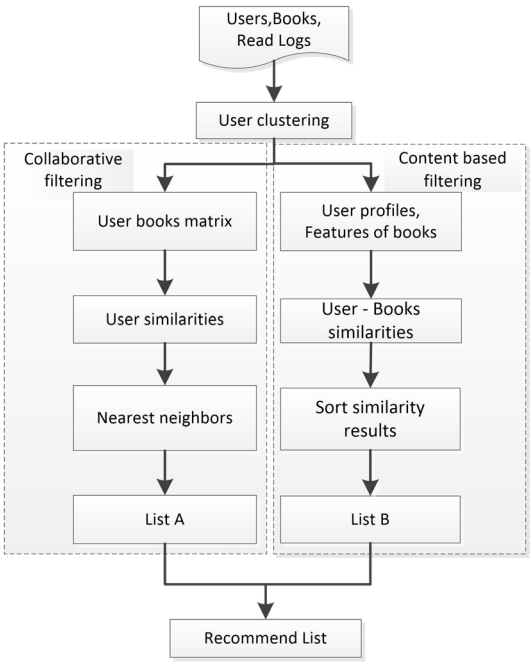

Fig. 2. Figure of Architecture of Hybrid Recommend System.

4. Experimental results and analysis

4.1. Experimental data

The experimental data were collected from Library of Inner Mongolia University of Technology from 2014 to 2016, including 43,153 users' information, 70,000 books and 200,593 users' books borrowing records.

4.2. Experimental Result

This paper used three methods to solve the sparsity problem. Method1 is to replace the book with the classification of the book, using the user-category matrix instead of user-books matrix. There are 255 categories of libraries. Method2 is to select 30 common categories as the user's eigenvector. Method3 is clustering, the cluster number is set to k = 15 and the maximum number of iterations is 50. The clustering results are shown in Table 1, and the number of users in each cluster is shown.

Table 1. Cluster Result.

Cluster	Number of Users	Cluster	Number of Users	Cluster	Number of Users
1	1347	6	895	11	1206
2	4205	7	1077	12	1288
3	340	8	1147	13	1100
4	703	9	1328	14	366
5	1181	10	1264	15	1168

Sparsity of the matrix can be computed by the formulae (5). The element 'a' denotes the number of nonzero elements. The element 'b' denotes the total number of elements.

$$s = 1 - \frac{a}{b} \quad (5)$$

The sparsity of the three improved methods is shown in Fig. 3.The sparsity of user-book matrix is 99.99%.The sparsity of method 1 is 98.7%, which is still high. Method 2 uses 30 major book categories instead of books, and the sparsity of user-categories matrix is 88.54%. Method 3 uses user clustering, and then generates 15 user-categories matrix. The sparsity is effectively alleviated to 76.42%.

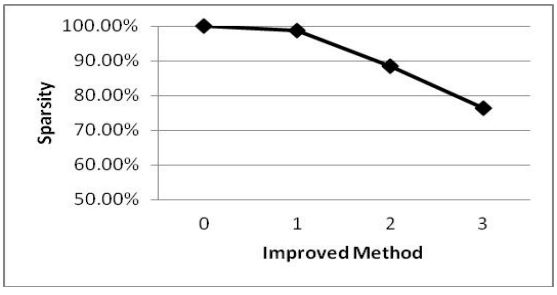

Fig. 3. Figure of Sparsity of Three Improved Methods

In order to verify the effectiveness of the hybrid recommender system, the experiments of single algorithm and hybrid algorithm are respectively performed to compare the precision of the three algorithms. Precision is a very important

indicator for evaluating the performance of recommender systems. Precision is defined as follows:

$$\Pr ecision = \frac{sum(R(u) \bigcap T(u))}{sum(R(u))} \quad (6)$$

R(u) denotes the list of recommended for made according to the behavior on the training set, and *T(u)* is the list of books on the test set. We run the collaborative filtering algorithm, content-based and hybrid algorithm with respect to different training set size. Fig. 4. shows changes of precision as the size of training set increases from 10 to 50. It can be clearly observed that the precision of each algorithm gradually increases.

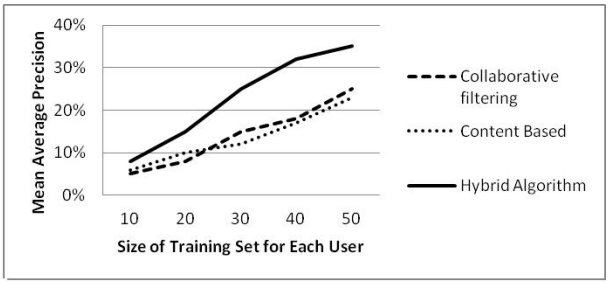

Fig. 4. The Precision of Different Recommend Algorithm

4.3. Experimental analysis

Similar to the traditional collaborative filtering algorithm, the number of books read by user affects the performance of recommend algorithm. For a small training set, users' preferences were not fully expressed, while, a large collection incorporates many users' preferences. Improved collaborative filtering algorithm solved the problem of data sparsity by combining clustering algorithms. Obviously, hybrid algorithm based on the collaborative filtering algorithm and content based algorithm improved the efficiency and quality of the recommendation algorithm. Meanwhile, it can also solve item cold start issues effectively.

5. Personalized book recommendation system design

5.1. Spark-based book recommendation system design

As the number of recommended system users and books increases, the amount of data calculated increases rapidly. When recommended for multiple users at the same time, the system computing pressure will increase exponentially, causing problems such as slow running speed and poor real-time performance. In order to solve the above problems, the article uses the Spark big data platform. The Spark big data computing platform is based on memory computing, with less reads and writes to the hard disk and fast execution speed of parallelization algorithm.

This paper is based on the Spark platform's personalized recommendation system for university libraries, and uses Spark cluster framework as a data mining platform for hybrid recommendation algorithms. Firstly, the Spark Core module is used to achieve the parallelization of the hybrid recommendation algorithm, and the scheduling management of related computing tasks is completed. Next, the Hadoop Yarn module is used to manage the cluster, and HDFS (Hadoop Distributed File System, HDFS) is used for distributed management and storage of book big data. Next, the Hadoop Yarn module is used to manage the cluster, and HDFS (Hadoop Distributed File System, HDFS) is used for distributed management and storage of book big data. Then the data transfer between HDFS and Oracle relational database is achieved by using Sqoop technology. he system structure diagram is shown in Fig. 5.

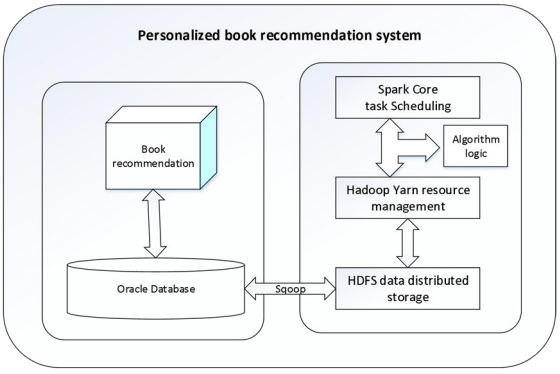

Fig. 5. Book Recommendation System Architecture Diagram

5.2. Book recommendation system display

The main task of the college library personalized recommendation system based on the hybrid recommendation algorithm is to recommend books to efficient library users. Combine the user's historical borrowing information and related book content information to achieve intelligent recommendation for new and old users. Based on the information records of 43,153 users, 200,593 borrowings and 76,638 books from the Inner Mongolia University of Technology Library from 2014 to 2016, this paper completes the training of personalized book recommendation models training and combines the Spark big data platform to improve the real-time and convenience of the model. Finally, the paper achieve the personalized recommendation system of university libraries. The following takes the user 201320801077 as an example to show the system recommendation results, as shown in Fig. 6.

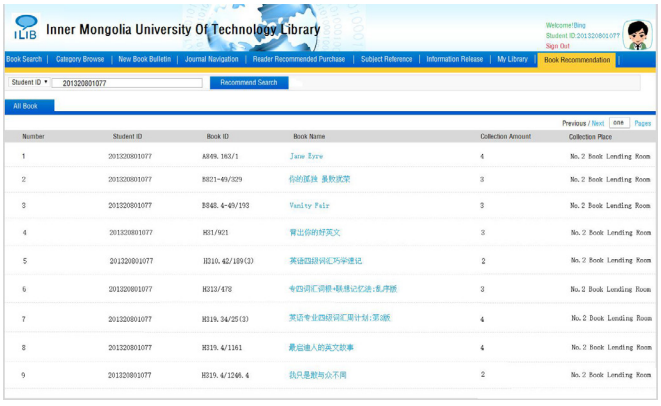

Fig. 6. Book recommendation system display

The intelligent book recommendation system based on the big data platform provides personalized book recommendation services for users. To some extent, it improves the efficiency of book recommendation and the number of books borrowed, and reduces the waste of university book resources.

6. Conclusion

When the number of books provided by library is relatively large, it becomes difficult for user to select appropriate book from a lot of candidate books. Aiming to help user overcome the problem, this article proposed a hybrid algorithm, which is based on the collaborative filtering algorithm and content based algorithm. On this basis, Spark big data platform is combined to achieve the personalized book recommendation system and improve the utilization rate of book resources. The main contributions of this paper include: (1) describe the architecture of the hybrid recommender system and the details of all procedures of recommender system; (2) several studies empirically compare the performance of the hybrid with the pure collaborative and content-based methods; (3) demonstrate that the hybrid methods can provide more accurate recommendations than pure approaches. In the future, we will study the following two aspects: (1) Investigate different technologies for improving the performance of system. (2) Solve the problem of user cold start.

Acknowledgements

The work is Supported by the natural science foundation of Inner Mongolia （ No.2013MS0920 ） , Inner Mongolia science and technology plan project（No.201502015）.

References

[1] Sun Yanchao, Han Fengxia. Research on Personalized Book Recommendation System Based on Collaborative Filtering Algorithm [J] .Journal of Library Theory and Practice,, 2015(4):99-102.
[2] Chih-Ming chen .An intelligent mobile location aware book recommendation system that enhances problem based learning in libraries [J].Interactive Learning Environments , 2011 , (1):45 -51.
[3] Waila, P.; Singh, V.; Singh, M.. A Scientometric Analysis of Research in Recommender Systems. Journal of Scientometric Research, 2016 (4): 71–84.
[4] Lin Xiaoe,Zhou Bin,Li Kechao. Research on Cold Start Problem of Digital Library Personalized Recommendation for New Readers and New Books [J] .Journal of Intelligence Theory and Practice,2014,08:100-104+99.
[5] Zilei Sun, Nianlong Luo. A New User-Based Collaborative Filtering Algorithm Combining Data-Distribution[J]. International Conference of Information Science and Management Engineering, 2010, 2(8): 19-23.
[6] Adomavicius, G., Tuzhilin, A. (June 2005). Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. IEEE Transactions on Knowledge and Data Engineering. 17 (6): 734–749.
[7] Sheng Jiagen, Liu Sifeng. A Knowledge Recommend System Based on User Model [J].International Journal of Digital Content Technology and its Applications, 2010,4(9): 168-172.
[8] Sanghack Lee , Jihoon Yang , Sung-Yong Park, Discovery of Hidden Similarity on Collaborative Filtering to Overcome Sparsity Problem, Discovery Science, 2007.
[9] Balabanovicm, Shohamy. Fab: content-based collaborativere commendation[J] . Communications of the ACM,1997, 40 (3): 66-72.