Research and design of recommendation in book circle system

Hui Xu, Shanshan Ren, Jingjing Yan and Zheng Li

School of Information Science & Technology, Nantong University, Notong 226019, China

Corresponding author and e-mail: Hui Xu, xu.h@ntu.edu.cn

Abstract. Reading is the ladder of human progress. It is the real needs of readers to quickly find good books and needed books in this fast-paced times. It will be more enjoyable if we can find the book friends who are interlinked with each other through reading. Digital reading records the whole process of reading, which contains readers' reading preference. In this paper, we discussed the techniques based on the data related to reading: (1) book friends are personally recommended using books as the media; (2) good books are personally recommended taking readers as media; (3) reader recommendation model and book recommendation model are built based on the relationship between "book-reader" in reading events. A reading system with such a recommendation function can make it easy for readers to find the books they want and find book friends who have sympathetic thoughts.

1. Introduction

There are too many wise sayings about "books" and "reading" to show the importance of reading. Books are the chronicle of mankind (Kanye), the ladder of human progress (Gorky), and the spiritual legacy from one generation to another (Herzen). Books are our teachers, without flogging or beating, without words or reprimands, without tuition and form (DE berry), and especially the young people's inseparable life partner and mentor (Gorky); Books are the nourishment of the world, without books in life, it seems like there is no sunshine (Shakespeare); Books introduce us into the best of societies, and introduce us to the great minds of the ages (H. H. Smith).

Nowadays, reading is more and more convenient. Traditional paper books, in addition to the library, the book bar with diversified functions and flexible business hours, have blossomed everywhere, attracting a large number of readers; The development of information technology has opened up a new form of books -- digital books and a new mode of reading -- digital reading. Books can be read anytime, anywhere, and more and more readers and more and more books are read.

Reading is already necessary for people living in this rapidly developing society to understand the society, integrate into the society, and increase their knowledge and skills. So in the vast ocean of books, it is the dream of every reader to quickly find books that meet the needs, find good books that are worth reading, and find people who can communicate spiritually by using books as medium. This article studies the recommended techniques in the book circle to solve such problems.

2. Recommend relevant technologies

Recommendation system is a very potential solution to the problem of information overload [1]. It is based on the user's previous behavior (such as collecting, downloading and rating), and it will
specifically recommend the products that users can like or the information they may need, so as to make the way for users to obtain information from massive data more simple and more fast.

Recommendation algorithm is the core of recommendation system. According to different calculation methods, recommendation technology can be divided into recommendation based on association rules, recommendation based on content, recommendation based on knowledge, recommendation based on collaborative filtering, and recommendation based on hybrid.

2.1. Recommendations based on association rules
Recommendation based on association rules [2] is to find the internal connection between different projects from a large amount of data, that is, to dig out strong rules and make recommendations based on the rules.

The generation of association rules is divided into two steps: (1) find all frequent item sets in the item set, that is, the item set whose support is not less than the preset minimum value. (2) find strong rules in the frequent item set, that is, the associated item set whose support degree and confidence degree are not less than the preset value. The more classical association rule algorithms are Aprior algorithm and FP-Growth algorithm.

2.2. Content-based recommendations
Content-based recommendation is to recommend similar products according to the products users liked in the past.

In the process of content-based recommendation, the most important step is to extract the user's interest features and the features of the recommended items, as well as to establish the user's interest model. In terms of feature extraction, the most studied method is to use TF-IDF algorithm through word frequency statistics [3]. In the aspect of establishing user interest model, the algorithms that have been widely used at present mainly include decision tree, neural network and vector-based representation method.

2.3. Collaborative filtering recommendation
Collaborative filtering recommendation [4] is to find the interest and interest relationship between users and the recommended objects by collecting the historical behaviors of users, and then make recommendations based on this relationship. This recommendation algorithm can be divided into two forms: user-based recommendation and item-based recommendation.

The basic principle of user-based recommendations [5] is to discover their level of preference for a project through the historical data of all users. The calculation finds neighboring user groups with similar interest preferences to the target user, and finally makes recommendations to the target user based on the neighboring user group's historical preference information.

The basic principle of project-based recommendations [6] is very similar to user-based recommendations, except that it is necessary to change the calculation of similarity between users to calculation of similarity between items, then find the nearest neighbor of the unrated item, predict the score of the corresponding unrated item based on the user's rating of the nearest neighbor, and finally feedback the top-ranked predicted item to the user.

2.4. Mixed recommendations
In order to make up for the shortcomings of various recommendation technologies, scholars have proposed a mixed recommendation design idea. Hybrid recommendation is to flexibly combine different recommendation algorithms according to different application scenarios and specific problems to implement the recommendation function to improve the quality and performance of the recommendation by complementing the advantages of different recommendation algorithms.

3. Recommendation research in book circles
Reading online not only records what the user has read, but also all the behaviors associated with reading. Hidden in this mass of information are the user's reading habits, preferences, thoughts and opinions. Based on the analysis of these data, it is possible to discover the relationship between "reader - reader", "book - book", and "(reader, book) - (reader, book)" (as shown in Figure 1). These relationships can be used for personalized recommendations of books, personalized recommendations of book lovers, building reader models and book models.

![Figure 1. Recommendation in "Book Circle".](image)

3.1. Recommending books by medium of "reader"
For the most part, people read with a purpose. Time is like gold, what should people do to find the books they need in the sea of books? Based on online reading data, there are multiple ways to recommend books to readers. You can recommend books through reader attributes, readers' bookshelves, and readers' book reviews, as shown in Figure 2.

![Figure 2. Recommending books using "readers" as a medium.](image)

3.1.1. Recommend by reader attribution. It is often said that things are grouped together by class, and readers with the same attributes imply similarities in cultural level, level of awareness, level of needs, etc. The simplest way to recommend is to classify books according to the attributes of readers, i.e., readers with the same attributes read books as recommendations for each other. The recommendation technique used is a collaborative filtering recommendation based on the user. The methods are:

- **Step 1**: Calculate the similarity between other readers and target reader B through the reader property;
- **Step 2**: Find the nearest neighbor reader or reader set A to the target reader B;
- **Step 3**: Analyze A's book collection, and recommend it to B according to certain criteria (such as ranking)
3.1.2. **Recommending books by readers' book bookshelf.** Infer the user's preferences based on the user's bookshelf, infer the common book preferences based on the books in the bookshelves of other users with similar hobbies, and then recommend books to specific users accordingly to achieve personalized book recommendations.

3.1.3. **Recommending books by readers' book reviews.** Book review features are available on general reading platforms. In this way, the user-label and book-label labeling relationships can be used. Use the K-means algorithm to perform user clustering and book clustering to form user groups and book groups, and calculate the cosine similarity, compare different book types from the perspective of book review attributes, and score and sort the books and finally realize Implement personalized recommendations.

For example: according to the emotion value of the emotion word closest to the attribute word, score the comment attribute of each book in the book group, multiply each attribute score by a certain weight and add up to get the score of each book, cluster the books in each group again, and recommend the top n books after clustering to the user group.

3.2. **Recommend "book friends" based on "books"**

Everyone's today is shaped by the past learning and experience. Different people will like different books. People who like the same books are likely to have resonance in their thoughts and have similar views and knowledge. Using books as a medium to recommend book friends, you can find similar readers by analyzing the similarity of bookmarks, content, ratings, associations, etc., and make friend recommendations. as shown in picture 3.

![Recommendation diagram](image)

**Figure 3.** Recommending book friends using books as a medium.

3.2.1. **Finding similar readers.** Take the book view as an example.

*Step 1:* Analyze the user's evaluation and obtain the main rating matrix $A_{User, item}^{M \times N}$.

*Step 2:* Calculate the similarity between all users based on the user's evaluation. Each row of $A$ is the user's evaluation vector for all items. The similarity calculation uses the cosine similarity formula to calculate the vector similarity between the user and the user to obtain the user similarity matrix $B_{User, User}^{M \times M}$. Key codes are as following:

```java
// Calculate cosine similarity
double dis = 0; // Specified user vector * user i vector
double asum = 0; // Square of user i vector
double bsum = 0; // Specify the square of the user vector
for (int i = 0; i < targetFeature.length; i++) {
    dis += ele.getValue()[i] * targetbookFeature[i];
    asum += ele.getValue()[i] * ele.getValue()[i];
    bsum += targetbookFeature[i] * targetbookFeature[i];
}
if (asum == 0 || bsum == 0) {
    dis = 0;
} else {
    dis = dis / (Math.sqrt(asum) * Math.sqrt(bsum));
}
```
3.2.2. Recommending based on the content of the book. Books record facts, state stories, or give the author's feelings, really touch the reader is the content of the book. Therefore, two readers with similar reading content are likely to become book friends. The content-based recommendation method can be used.

   To analyzes the commonality of the books that the readers have previously read or interested in, and then compares these commonalities with other readers' book characteristics, and finally recommends the readers with the highest similarity ranking as book friends. The specific method is as follows:

   Step 1: Getting the collection of readers' reading books;
   Step 2: Mining readers' interest preference from the collection;
   Step 3: Calculating the similarity between the interest preference and interest preference of other readers;
   Step 4: The readers corresponding to the most similar book collections are recommended as book friends.

3.2.3. Recommending based on Book Association. Reading is a process of expanding knowledge and constantly improving. A book read by a reader in a certain period of time usually has some connection. For example, a computer major student has successively read "Fundamentals of program design", "data structure and algorithm", "algorithm design and analysis", "ACM competition training", etc. This indicates that the student is likely to be preparing for a programming competition. Those with the same interests can be like-minded people. Therefore, the association between books can be used as the recommendation of books or the recommendation of book friends.

   Apriori algorithm is the simplest recommendation technology based on association rules. The specific operations are as follows:

   Step 1: Scan the data set of the all read book to get frequent k-item sets;
   Step 2: Calculate the distance between the reader's book collection and frequent K sets;
   Step 3: The readers who are close to a certain k-item collection will be regarded as book friends.

3.3. Building a model based on the "reader-book" relationship

Whether you use "people" as a medium, recommend a book, or use "books" as a medium, and recommend "friends", you can't do without the relationship between people and books in the book circle. This relationship is the basis for establishing a user model and recommendation goals.

3.3.1. Reader model. In the book circle system, the reader model can be obtained through data analysis of all actions of readers’ book search, book reading, and book review. For example:

   (1) Reader model based on preference. To analyzes the books that readers consult to obtains readers' favorite authors, themes, characters, events, etc., so as to make the model object attribute associated with readers.

   (2) Reader model based on book review. Analyze and excavate the book review information to obtain the readers' positive or negative evaluation of a book, such as: like, dislike, hate, poor, beautiful words, positive energy, philosophy, etc.; give the readers' portraits through the book review.

   (3) A model based on reading time. To analyzes the characteristics of readers' reading time and classifies them. For example, the average daily reading time is more than 2 hours, the weekend reading time is more than 3 hours, the night readers, etc.

3.3.2. Book model. Through all actions such as querying, reading, commenting and recommending by readers, the model of the book is established. For example:

   (1) A collection of books read by the same reader.
(2) A model based on readers' attributes. From the attributes of readers and their book collections, the book collections with different levels of readers' preferences can be obtained.

(3) Frequency-based model. Books are classified according to the number of occurrences in the statistical analysis system recommendation or author recommendation, the number of queries by readers, the number of readings by readers, and the number of repeated readings by the same reader.

(4) A model based on characteristics of reading time. Classify books according to the length and time period of reading, etc.

(5) A model based on book reviews. Analyze and mine the words and semantics of the book, and get different types of books such as good reviews, bad reviews, philosophy, positive energy, and long knowledge.

4. Conclusions
Through the media of "book", we can get more book friends who love reading, are similar to their own thoughts, or are even different. Though reading, on the one hand, people can expand their own thinking dimension and invest in the sea of thoughts; on the other hand, people can expand their circle of friends, know more various friends, and escape from the isolated island of life. This paper studies the recommendation of books and the discovery methods of book friends in the book circle, so as to help readers find the books they need conveniently, find people who have ideological resonance to share books, and let book lovers share their warm feelings with each other.

References
[1] Goldberg D, Nichols D, Oki B M, etc. Using collaborative filtering to weave an information tapestry[J]. Communications of the ACM, 1992, 35(12): 61-70.
[2] Mooney R J, Roy L. Content-based book recommendation using learning for text categorization[C]//Proceedings of 5th ACM Conference on Digital Libraries. San Antonio, USA, 2002: 195-204.
[3] Wu H C, Luk R W P, Wong K F, etc. Interpreting TF-IDF term weights as making relevance decisions[J]. ACM Transactions on Information Systems, 2008, 26(3): 55-59.
[4] BOBADILLA J, O R TEGA F, HEMANDO A, et al. Improving collaborative filtering recommender system results and performance using genetic algorithms [J]. Knowledge-Based Systems, 2011, 24(8): 1310-1316.
[5] Zhang L, Tao Q, Teng P Q. An Improved Collaborative Filtering Algorithm Based on User Interest[J]. Information Science, 2014, 9(4): 999-1006.
[6] Karypis G. Evaluation of Item-Based Top-N Recommendation Algorithms[J]. Evaluation of Item-Based Top-N Recommendation Algorithms, 2001, 22(1): 247-254.