

Design and Application of Personalized Recommendation Technology in Research Information Service of College and University

Qing-cong Zhao

School of Information Management, Beijing University of Science and Technology
Information,
Beijing 100192, China

Keywords: College and university, Personalized recommendation, Research information service

Abstract: With continuous progress and development of scientific research work in college and university, there are more and more scientific research results, including a variety of documents. In this paper, a personalized information service platform for scientific research in college and university was designed and applied to solve the related research literature information management confusion and other issues in college and university. The study results showed that the system can support and provide personalized search services for users, change passive service to active recommendation, so as to help users to filter out valuable resources in the massive literature resources to meet the user's demand.

Introduction

University and college is not only the cradle of cultivating students to become professional talents, but also is the field of having research and developing new technology. Teachers in colleges and universities can not only continue to ensure and improve the quality of teaching, but also can organize and carry out scientific research actively, combining the teaching and the scientific research together. In this way, it can not only improve the educational level and the academic level, but also can enhance the comprehensive strength of the school, so as to cultivate groups of qualified professional talents for schools. Then there will be more and more scientific research projects and scientific research achievements, which are mainly various types of literature, thus, how to help teachers or researchers to find out the needed information efficiently and accurately becomes an urgent issue to solve ^[1]. The proposed personalized information service platform in this paper can provide personalized information service for teachers and management staff of scientific research, it can promote the fusion between different majors, which also can reflect the disciplines development trend of schools, so as to guide whether the next step needs to increase the corresponding investment, introduce the relevant personnel or to declare the corresponding topics and so on.

Personalized Recommendation Technology

The Status of Research

Personalized recommendation can be understood as a special advanced information retrieval tools, which is mainly composed by three elements, namely, the target users, the candidate objects (projects or products), as well as the recommendation algorithm models, including the following steps: firstly, according to the characteristics and attributes of the users, recording the users' historical information and behavior; then combined with the recommended algorithm model, using model to analyze the potential preferences of users and preferences gradually; finally sorting objects in accordance with the appropriate level of recommendation, according to the actual needs of the project from the collection of selected higher ranking as a part of the recommended results to the target users ^[2]. The specific recommendation process can be shown in Fig. 1, one of the most important steps is Finding Similar Users, namely, calculating out user's similarity. The commonly used methods are as follows: Jaccard Index, Person Correlation Coefficient, Cosine-based Similarity, Euclidean Distance Score, Distance Calculation and so on.

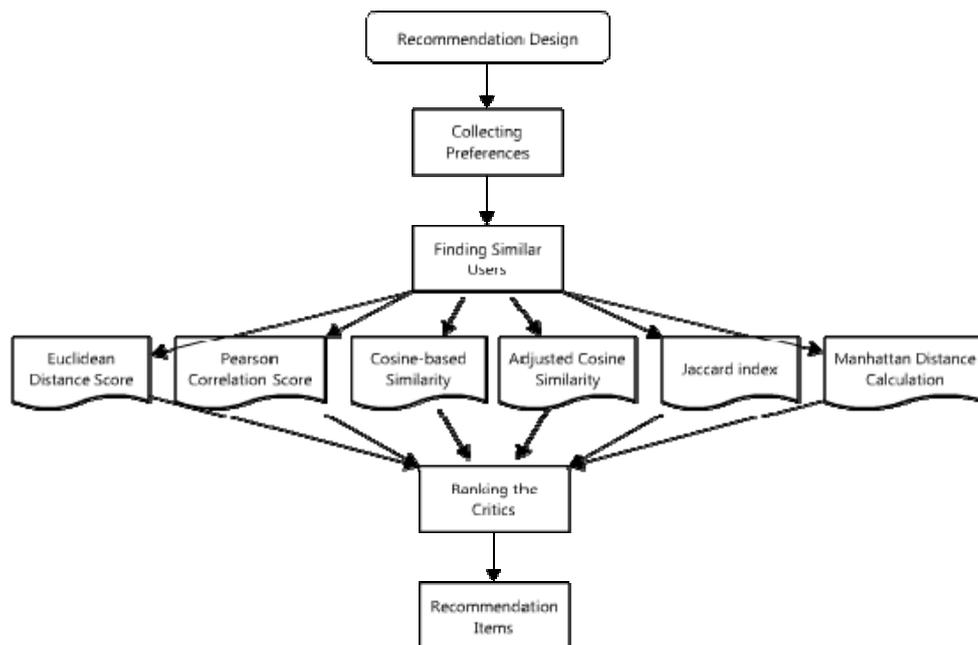


Fig.1 Recommendation Process Graph

Commonly Used Personalized Recommendation Algorithm

At present, personalized recommendation algorithms emerge in an endless stream, which are especially popular in e-commerce applications. Since the data characteristics are so different, there is no one algorithm can lead the others in all aspects, therefore each of them has its own advantages and disadvantages. The commonly used recommendation algorithm mainly includes: Association Rule-based Recommendation, Content-based Recommendation and Collaborative Filtering Recommendation.

Association Rule-based Recommendation

Association rule aims to find out the correlation between different items appearing in the same event such as famous beer and diapers event ("shopping basket analysis"), its core is to calculate the correlation in one purchasing transaction of a user to buy different commodities. Recommendation based on association rule is to use the association rules for data mining, so as to find out the potential association relationship from a large number of historical records, and automatically produce recommendation results in the next transaction. The advantage of this algorithm is that with the continuous accumulation of test data and the increasing number of copies, the personalization and accuracy of recommendation results will be improved accordingly. While the disadvantages are as follows: the algorithm requires more data and the association rules are difficult to get; some discrete values of the test data may lead to the loss of valid information.

Content-based Recommendation

Content-based Recommendation is take the content of the product information as basis to make recommendation, while the content of product information is defined through the relevant attributes as the basis, among them, the most important part is the user's data and the forecast calculation of the tested products, so as to choose the product with the maximum similarity and matching degree as the recommended results from all products. The commonly used content-based modeling methods include decision tree, neural network and so on.

The advantages of this recommended algorithm: this kind of algorithm is simple and effective; it can solve the problem of "Cold Start" of data that new registered users have in the most of systems, which only needs the attribute of the recommended products, keywords and other characteristics, the possibility of recommendation that one new product of new entry in the database can be the same as that of the existed product in the database; another advantage is the independence between users, namely, the target users and other users can't have mutual effect and influence on each other, thus the recommended results are not the same. While the disadvantages of the algorithm are as follows: the use of this kind of algorithm is usually limited to the recommendation of products that are easily analyzed or extracted.

Personalized Research Information Service Platform of College and University

Platform Overall Function Design

The personalized research information service platform can be as part of the management of scientific research, its function should ensure the integrity, which should fully meet the actual needs of users and recommend personalized information resources for users, which also needs to satisfy the administrator so as to meet the requirement of the chart data information analysis and prediction. The platform mainly includes document management, document statistics analysis and user management etc.. In document management module, it can support the maintenance and retrieval of document information. During the period of implementing the retrieval function, personalized recommendation technology is adopted, so as to realize personalized retrieval service for different users.

Personalized Literature Recommendation Design

The users of personalized research information service platform in college and university are teachers and researchers of the school, which can be divided into two categories, one is

new registered users, these users do not browse or download any records, as for this type of users, content-based recommendation algorithm can be used, namely, in accordance with the direction of the registered that users filled in the field and direction of science and research to recommend the most hot literature (the amount of the related areas most downloaded); the other is the user that had download records and literature of scores, as for this kind of users, it can choose user-based collaborative filtering algorithm, namely, by calculating Euclidean Distance Score among users to judge the similarity of users. According to the similarity of users to recommend literature resources.

According to the analysis of the system function, the home page of the platform can set personalized recommendation -- "recommended for you" module, the function of this module can be shown as follows:

1) Recommending for new registered users. Because these users did not have the records of retrieving documents or downloading documents and records, as well as having no historical records, so in accordance with the direction of the users belongs to fields and research, it can recommend the same discipline in the field of the "hottest literature", or recommended users with the same discipline, the same research literature resources that other people have downloaded or published.

2) Recommending for users who have retrieved documents and downloaded documents and scored before. This kind of users should give scores after downloading and reading documents, so as to illustrate whether documents are useful for them, and how useful the documents are for them, whether it's relevant to what they have studied. By ordering the calculating result, users with high similarity can form neighborhood sets. Using the score data of the pretreated neighbor set in the backstage database, then it can get the retrieval and scoring prediction of target users, finally it can take the prediction value that is sorted forward as the result of the recommendation to target users. As for the user's retrieval records (inputting keywords), downloading records, scoring records, etc., system can collect and save automatically, according to the updated data in the database, the recommendation results can have real-time changes.

The Implementation of Personalized Recommendation Service

The Implementation Steps of Content-based Recommendation

Document Model:

This platform is mainly literature materials, which can extract the structured data from the literature resources, namely, the attributes of the documents (subject, keywords, etc.).

User Model:

Extracting the structured data from basic user's information, such as subject, research direction, etc..

Generating Recommendations:

By comparing the literature model and user model, we can get the literature resources that users are interested, so as to generate Top10 related list of the most relevant information for users.

The Implementation Steps of Collaborative Filtering Recommendation based on Users

User Information Model:

By using Scoring Matrix of User -Literature to calculate the similarity of users, then user information model can be established. This matrix is established according to the rating information for each user on the documents, among them, line can represent the user number vector, while column can represent the interval score vector, 0-5 can be used to describe the degree that users are interested in literature. In general, one $m \times n$ scoring matrix can be represented in the form of Table 1, among them, m is the number of users, while n is the number of documents. User's set $U = \{U_1, U_2, \dots, U_m\}$, document set $T = \{I_1, I_2, \dots, I_n\}$. R_{ij} is the rating value of user U_i to document I_j . In general, the higher the score value R_{ij} , the greater the degree of interest that the user U_i is interested in the document I_j . If a user U_i fails to score the document I_j , then the value of R_{ij} is null.

Table 1 Scoring Matrix of User-Literature

User	Item						
	I_1	I_2	...	I_j	...	I_{n-1}	I_n
U_1	R_{11}	R_{12}	...	R_{1j}	...	$R_{1(n-1)}$	R_{1n}
U_2	R_{21}	R_{22}	...	R_{2j}	...	$R_{2(n-1)}$	R_{2n}
...
U_i	R_{i1}			R_{ij}		$R_{i(n-1)}$	R_{in}
...
U_{m-1}	$R_{(m-1)1}$	$R_{(m-1)2}$...	$R_{(m-1)j}$...		$R_{(m-1)n}$
U_m	R_{m1}	R_{m2}	...	R_{mj}	...	$R_{m(n-1)}$	R_{mn}

Calculation of User's Similarity:

To calculate the similarity between users is to calculate the similarity of the vectors in the evaluation matrix in nature. The similarity between the target users and other users in the system can be calculated according to the document grading value given by the users together. We can adopt Euclidean Distance Score (Formula 1) to calculate, using users scoring literature as the axis to set up coordinate system, taking the participation of users as the score points, then it can calculate out the distance between them ^[4].

$$d(x,y) = \sqrt{\sum (x_i - y_i)^2} \quad (1)$$

In order to show the similarity of users, so as to facilitate the comparison of the processed data, we use Z-score standardization method to normalize the calculated values and reduce the range to (0, 1). The calculation of user's similarity, namely, Formula 2 can be shown as follows. When the distance is 0, the normalized value is 1, the larger the distance is, the closer the normalized value is to 0.

$$sim(x,y) = 1/(1 + d(x,y)) \quad (2)$$

As is shown in Fig. 2, user A and user B give scores to literature X and literature Y respectively, the score that user A gave to literature X is 1 point, while to literature Y is 3 point, so the coordinate of A point in coordinate system is A (1, 3); user B gave to literature X

is 4 point, while to literature Y is 1 point, so the coordinate of A point in coordinate system is B(4, 1). By using the formula, the Euclidean Distance Score between two users of A and B can be calculated to be 13. Finally, the user's similarity is 0.07.

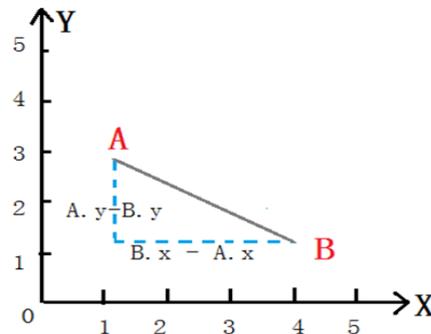


Fig.2 Euclidean Distance Score

Generating Recommendation Results:

Based on the calculated similarity with the target users , it can order all users from large to small, selecting Top2 users that had the highest similarity with the target users as the neighborhood set of the target users, denoted as $N(U_a) = \{U_1, U_2, \dots, U_k\}$. Among them, U_a is the target user, while U_k is one neighbor user of U_a . According to the nearest neighbor prediction score for each document that is given by the target user A, firstly it can calculate out the sum of similarity score between the target user A and its nearest neighbor user, as for the literature that user A gave no scores, while other users gave, using the score given by the nearest neighbor user times its similarity m , which van be divided by the sum of similarity so as to get the recommended value. Finally, listing the Top10 documents with the highest prediction scores but having not given scores yet by the target users in the recommendation list to the target user A.

Conclusion

Personalized recommendation technology is an important means to realize the personalized service, in this paper, the personalized information service of university scientific research platform is designed firstly to collect all the information after examination of the school, then according to the basic personal information that users filled in (the subject, research direction, etc.), combined with the user's downloading records, as well as the scoring record of documents, adopting the mixed model of Content-based Recommendation and Collaborative Filtering Recommendation, so as to provide personalized recommendation service for users.

Acknowledgement

This research was financially supported by the Importation and Development of High-Caliber Talents Project of Beijing Municipal Institutions (CIT&TCD201504059).

References

- [1] Li Wenwen, Cheng Ying. Information behavior analysis of scientific research staff and implications for library personalized scientific research services. *Information Science*[J], January 2017, Vol. 35 (No.1): 74-105.
- [2] Liu Rujian. From KNS to KDN: Review of knowledge discovery network platform of CNKI [N]. Beijing: *Library Science Research*. 2016.
- [3] Shan Liu, Yao Dong, Jianping Chai, Research of Personalized News Recommendation System Based on Hybrid Collaborative Filtering Algorithm[M]. Beijing: 2016 2nd IEEE International Conference on Computer and Communications, 2016.
- [4] Lee J, Sun M, Lebanon G. A comparative study of collaborative filtering algorithms[J]. *Computer Science*, 2012.