

A Top-N Algorithm-based Personalized Learning Recommendation System for Digital Library

<https://doi.org/10.3991/ijet.v11i11.6256>

Xin Gao, Huang Wen-xue, Ning Wang, Yan-chao Yang, Ying Yan
North China University of Science and Technology

Abstract—The digital library brings convenience, but meanwhile, it also brings the problems of overloaded information and over-diversified forms, thus search becomes difficult. Personalized Learning Recommendation System is the key to solve the problems, and suitable for the situation with user diversification and demand diversification. With the System, users spend the least time and energy in accurately finding the information they need, where efficiency is improved to the greatest extent. The research conclusion of personalized learning recommendation system based on Top-N algorithm is based on the calculation of the experimental results from the analysis of the related theory and technology based on Top-N algorithm.

Index Terms—Top-N algorithm, library, Personalized learning recommendation system

I. INTRODUCTION

Personalized learning recommendation service is a kind of creative and active service, and it is an effective solution to solve the problem of information overloading. Its biggest characteristic is that it has clear target[1-3]. This kind of new service model breaks the traditional passive service model. Under the support of various information technologies, it makes full use of the advantages of information resources, takes the initiative to meet the personalized information needs for individual users, and proceeds continuous service. At present, there are a large number of library management systems, most of which also provide readers with Personalized learning recommendation function, but the total effect is not satisfactory[4-5]. The main reason is that in the process of the recommendation, it is only dependent on the characteristics of the book explicit, while not taking into account the applicable environment of the book. Considering that the knowledge in books has the application value only in a specific environment, this paper introduces the application environment of books into the book recommendation system to improve the accuracy and efficiency of the book recommendation. In China, the research on personalized service system began in 1999, taking two dissertations Personalized Active Information Service Based on the Agent Technology and Achieve Personalized Active Information Service Based on Multiagent Evolving Ecosystem Algorithm published by scholars Lu Haiming and Xu Jin Hui as the mark. In recent years, many institutions have increased the research investment on personalized information service technology of library[7-8]. Overall, at present, China has not yet developed a set of perfect library personalized information service system. Whereas, with the development of information technology, more and more institutions have shown the trend of researching and developing personalized information service system, such as Beijing University library and Shanghai Communication University li-

brary, China Agricultural University library and so on. Thus, personalized information service system has a good development prospect in China.

II. ANALYSIS OF PERSONALIZED LEARNING RECOMMENDATION SYSTEM BASED ON TOP-N ALGORITHM

A. General Framework of Personalized learning recommendation System

In terms of traditional collaborative filtering recommendation algorithms, its biggest feature is not considering the specific content of the users and items, and compared to other algorithm, it has a great advantage. But it also has a very obvious defect that recommendation accuracy is relatively low. This is because, in the recommendation process, it does not consider the specific content of the users and items, while recommends the content that looks very similar on the surface to the users. This paper, Top-N algorithm is integrated into collaborative filtering algorithm, so as to solve this problem. After the calculation of similarity for the first time, the nearest neighbor set is obtained, and then according to the target user or item with the nearest neighbor set to calculate the similarity of knowledge scenario, finally superposes the similarity of these two parts, what is gained is the ultimate similarity of target user or item. Then, according to the similarity values, adopts Top-N algorithm to re-obtain the nearest neighbor set. The improved collaborative filtering algorithm absorbs the advantages of the two, which can effectively eliminate recommending error caused by similarity on the surface and it to large extent improves the accuracy of recommendation.

Personalized learning recommendation system framework of library personalization based on knowledge situation is shown in Figure 1. In this system, the key is the selection of knowledge situation. To select representative hidden situational characteristics, which is one of the factors that decides the whole recommendation performance.

B. Feature Model Based on Ontology

The reason why introduce ontology into digital library is that ontology can accurately describe the books' feature information, describe the relationship between knowledge, organize digital library knowledge, and reduce ambiguity in concepts and terminology. Book features, refer to the inherent characteristics of a book, which are born with, and they are the most obvious and direct features of a book. Figure 2 shows a book feature model based on the Ontology model and Figure 3 shows the Tree diagram of readers feature model. According to the principle of the tree diagram of the book feature model, it can also set up the tree diagram of reader feature model.

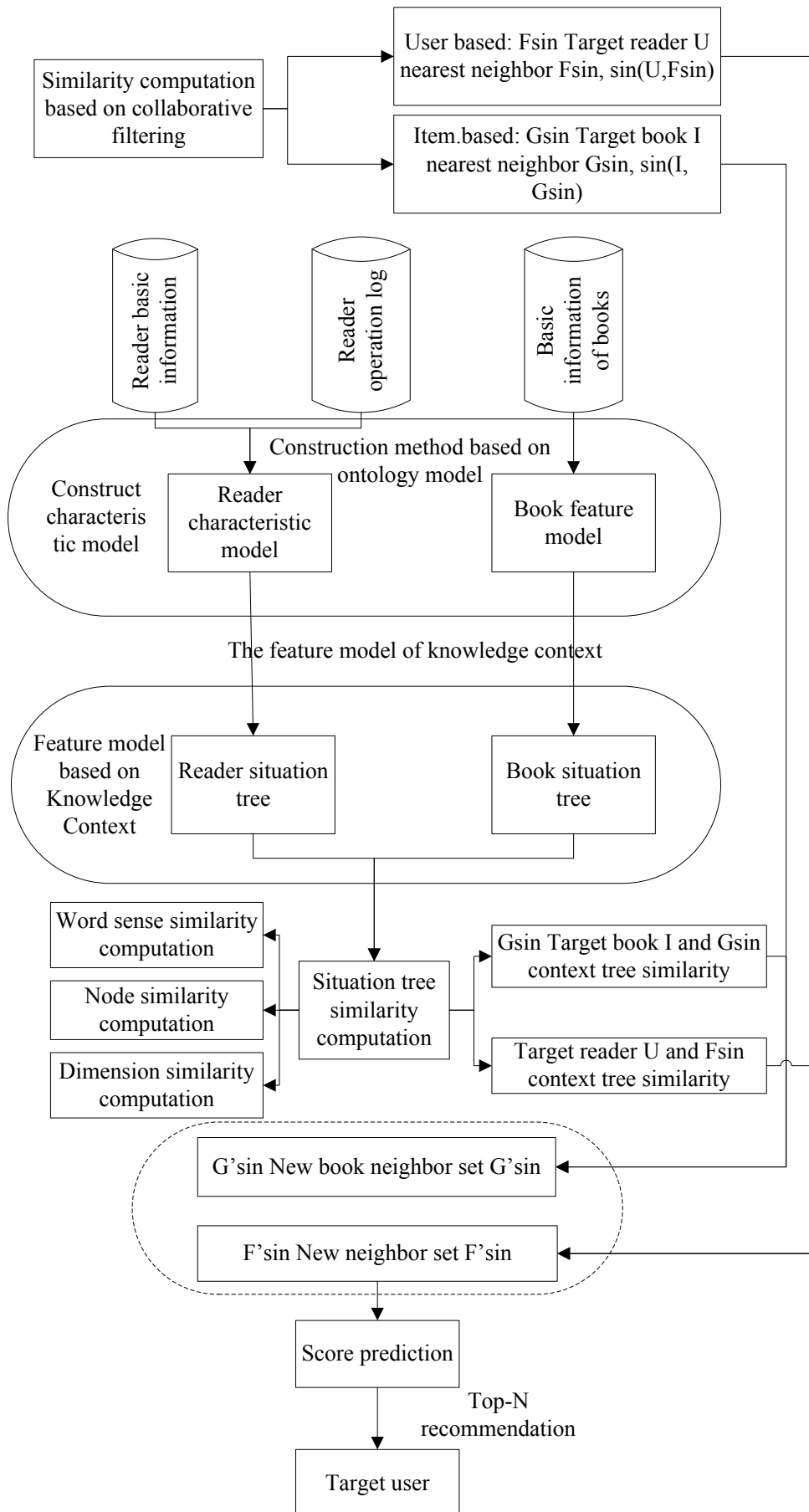


Figure 1. Personalized Learning recommendation system based on knowledge situation

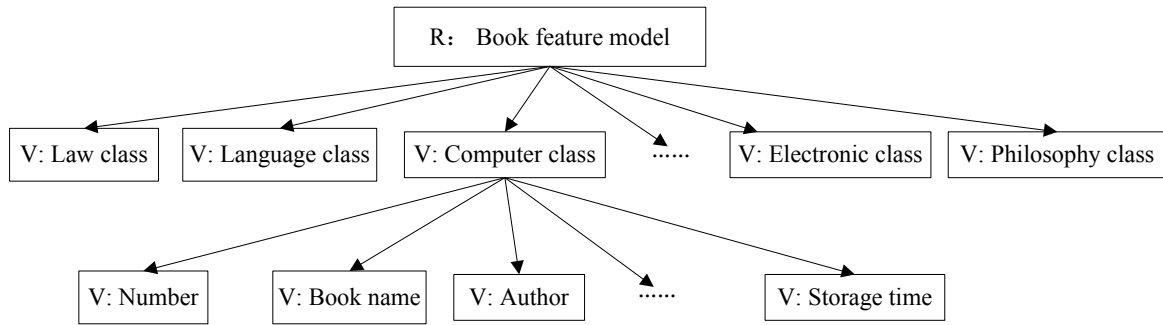


Figure 2. Tree diagram of book feature model

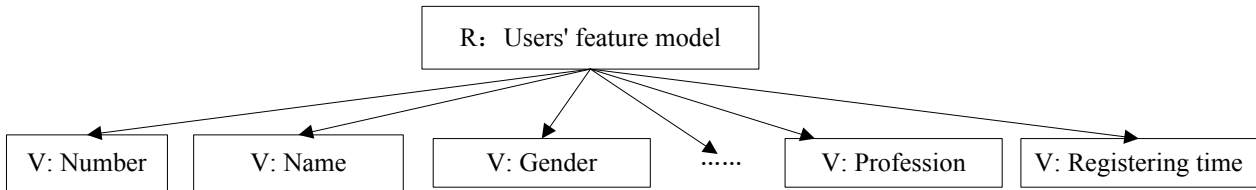


Figure 3. Tree diagram of readers feature model

III. EXPERIMENTAL RESULT

A. Experimental Data Sets and Experimental Environment

The experimental data used in this paper is the data set provided by the data center of Chang'an University library, which contains the book score of all the students of Chang'an University library. These scores constituted the user-item score vectors in traditional collaborative filtering algorithm, and using these score vectors, it is possible to calculate the similarity value without adding situation elements and get the nearest neighbor set; on the basis of these data, the reader situation - reader's hobby and interest, habits and skilled fields is added into the reader information; in addition, in book information, it is added book situation - the professional field of book research, problems books discussed, books innovation point, and readers field books face. Data format re-defined are shown in Table I and Table II. By comparison of the similarity of situational feature model, it is concluded of the similarity of the nearest neighbor set calculated by the collaborative filtering algorithm and situational feature model of target or user. The similarity generated by these two methods is summed up to obtain an ultimate similarity. Obtain a new neighbor set by the ultimate similarity value, generate recommendation by the new neighbor set, and recommend the result to users.

B. Evaluation Criteria of Recommendation System

In order to be able to accurately measure the effect of recommendation system, especially measure whether there is an improvement in recommendation after adding the situational knowledge compared to the original performance, we must have a reliable evaluation standard. The method of measuring the performance of recommendation system are of variety, and the criteria is different, while the comparatively commonly used methods are three kinds: MAE Mean Absolute Error, MSE, Mean Squared Error, and RMSE, Root Mean Squared Error, among which, MAE is the simplest and the most widely used performance evaluation criteria in recommendation system.

TABLE I.
 READERS' INFORMATION TABLE

| | |
|--|------------------------------|
| Readers' inherent characteristics | Reader's library card number |
| | Reader's name |
| | Reader's gender |
| | ID |
| | Major |
| Readers' situation | Reader's age |
| | Interest and hobby |
| | Habits |
| | Skilled field |

TABLE II.
 BOOKS' INFORMATION TABLE

| | |
|--|--|
| Books' inherent characteristics | Books' number |
| | Books' name |
| | Author |
| | Press |
| | Books' language |
| | Books' category |
| Books' situation | Professional field books researched on |
| | Problems books discussed |
| | Innovation of books idea |
| | Readers' field books face |

The specific method of Mean Absolute Error MAE is shown as follows: the absolute value of the difference between the reader's actual score values and the model prediction score calculated by model, accumulating the absolute values of the score difference of all the items, and finally divided by the total number of items.

Target user U , the number of items in training concentrated browse is m , the item score that the corresponding users has browsed but hasn't evaluated is obtained by the application of the formula, the actual score set of corresponding U users is $\{I_1, I_2, I_3, \dots, I_m\}$, and the prediction score set calculated by the system model is $\{\hat{I}_1, \hat{I}_2, \hat{I}_3, \dots, \hat{I}_m\}$, and then the Mean Absolute Error of the target user is:

$$MAE_u = \frac{\sum_{j=1}^m |I_j - I_j|}{m} \quad (1)$$

The Mean Absolute Error for all users (the number of users is n) in the entire system model is:

$$MAE = \frac{\sum_{u=1}^n MAE_u}{n} \quad (2)$$

In the formula 2, the smaller the value of Mean Absolute Error, it suggests that the value of system model predicting score and actual score is smaller, which shows that the higher the accuracy of recommendation system model, the higher the recommendation quality, and the better the recommending effect. This is a widely accepted and good method of measuring recommendation system performance.

C. Experimental Analysis

The experimental data is the data set provided by the data center of Nanchang University library. Select the evaluation data of 500 readers in the data set of 500 books as the experimental data. Among them, 70% is taken as the training set and 30% for testing set. Books information and readers' information is gained according to book storage table and readers information table of Nanchang University library. Take part of the information in these information as situation, and situational feature data are defined as shown in Table I and Table II.

During the experiment, three experiments were carried out, respectively test the original data, improve the filling rule and add the situational elements, and then do the experiments.

This experiment uses the modified cosine similarity calculation method. For the selection of the number of the nearest neighbor set, the general methods mentioned in the previous part are threshold method and Top-N method, and this paper uses Top-N algorithm.

For testing the recommending effects on different conditions, this paper selects MAE value of different groups' tests under different data sets, the experimental data sets and result are shown in Table III.

In Table III, U100I100 represents the evaluation data by selected 100 readers of 100 books, U200I200 indicates the evaluation data by selected 200 readers of 200 books, and by analogy. The purpose of doing this, in allusion to different amount of data, is to get the recommendation effect under the different amount of data.

TABLE III.
EXPERIMENTAL DATA AND RESULT

| Data set | The first time | The second time | The third time |
|----------|----------------|-----------------|----------------|
| U100I100 | 0.925 | 0.917 | 0.928 |
| U200I200 | 0.912 | 0.904 | 0.909 |
| U300I300 | 0.898 | 0.897 | 0.889 |
| U400I400 | 0.880 | 0.878 | 0.857 |
| U500I500 | 0.861 | 0.853 | 0.836 |

1) The first and second time experimental results analysis

In these two experiments (see Figure 4), it can be seen that in both of the experiments, the Mean Absolute Error MAE values are on the decline, indicating that with the increase in the amount of data, original data and experiment results based on filling value improvement method show that the efficiency and performance of recommendation system is increasing. It suggests that the recommendation effect of the collaborative filtering recommendation algorithm, with the increase of the amount of data, will be better. This is because the more the amount of data, the more the effective data can be used for the calculation, and the more accurate the selection of the nearest neighbor set.

2) The first and the third time experimental result analysis

From Figure 5, it is seen that with the increase of the amount of data, the Mae values are declining, which indicates that efficiency and performance of recommendation no matter what kinds of ways they are generated are improving.

From Figures 4 and 5, it is known that the improvement method of adding knowledge situation is more conducive to improve recommendation performance than the improvement method of filling value. Filling value improvement will help alleviate the data sparsity problem of the whole system, but it cannot fundamentally improve the effect of recommendation system. Recommendation system based on knowledge situation can alleviate the data sparsity problem in the system, but also enhances the effect of recommendation system. The combination of these two methods can alleviate the data sparseness problem but also enhance effect of recommendation systems.

IV. CONCLUSION

This paper, starting from the problem of information overloading existing in current digital library, discusses the method of solving the problem, analyses the origin of the problem of information overloading, and then proposes digital library Personalized learning recommendation system based on Top-N algorithm. First of all, it makes a brief introduction to the current most widely used recommendation system, studies various technologies used in recommendation systems, and analyzes the advantages and disadvantages of related technologies in recommendation systems. This paper puts forward introducing the concept of Top-N algorithm into the Personalized learning recommendation system. Based on ontology, construct feature model of users and items, and on this model, add items and users' situation, and finally obtain a feature model of users and items based on Top-N algorithm. The feature model similarity between users and the users, items and items calculated by this is more credible. This paper uses the improved Top-N algorithm to recommend N items that has the highest similarity for users. Eventually, the experimental results show that this method to a certain extent improves the accuracy of the recommendation.

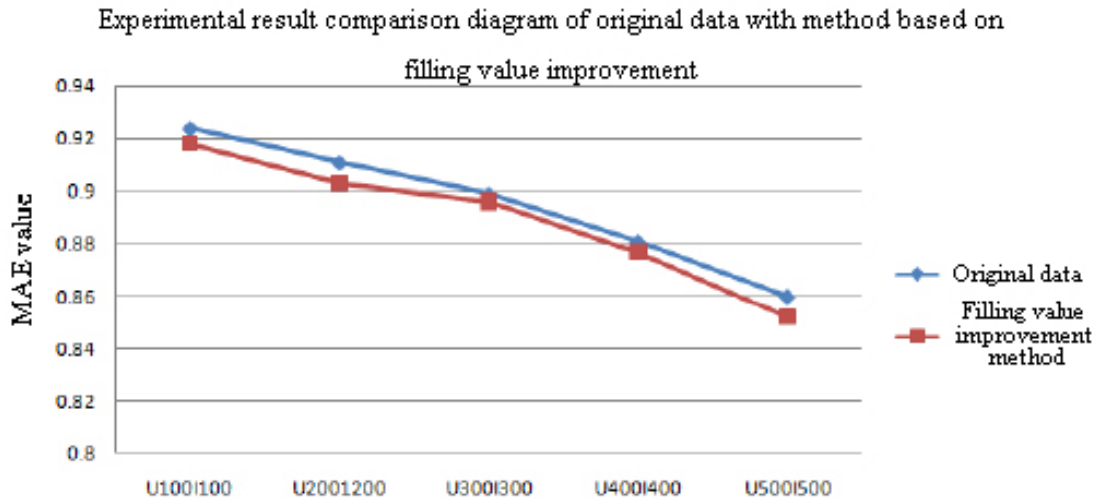


Figure 4. The first and second experimental results diagram

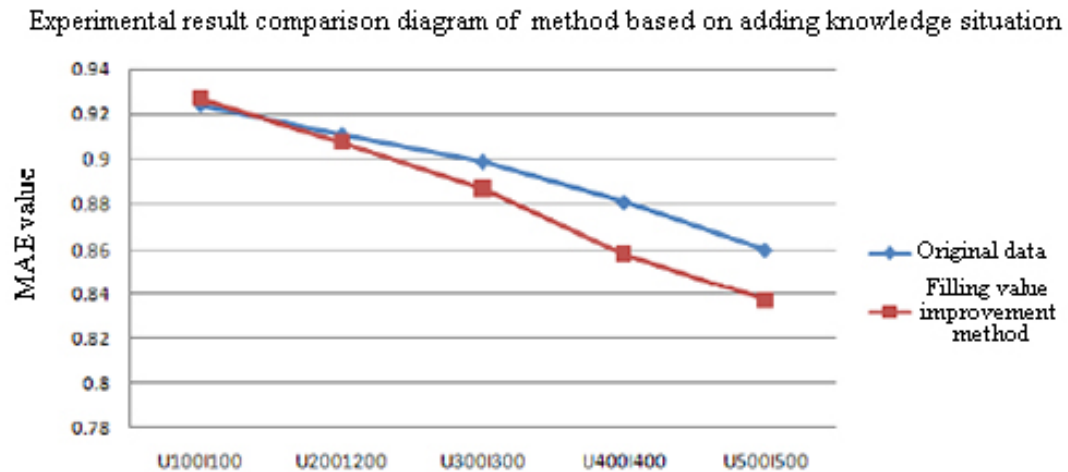


Figure 5. The first and the third time experimental result diagram

REFERENCES

- [1] H. Ping, "The Research on Personalized learning recommendation Algorithm of Library Based on Big Data and Association Rules," *Open Cybernetics & Systemics Journal*, vol. 9, no.1, pp. 2554-255, 2015. <https://doi.org/10.2174/1874110X01509012554>
- [2] H. Jing, "Routing optimization algorithm based on nodes density and energy consumption of wireless sensor network," *Journal of Computational Information Systems*, vol. 11, no.14, pp. 5047-5054, July 2015.
- [3] H. Jing, "The Study on the Impact of Data Storage from Accounting Information Processing Procedure," *International Journal of Database Theory and Application*, vol. 8, no.3, pp. 323-332, June 2015. <https://doi.org/10.14257/ijta.2015.8.3.28>
- [4] Y. Liu, J. Cheng, C. Yan, et al., "Research on the Matthews Correlation Coefficients Metrics of Personalized learning recommendation Algorithm Evaluation," *Journal of Higher Education*, vol. 37, no.6, pp. 39-41, 2015.
- [5] Q. Hou, Y. Sun, and S.O. Management, "Collaborative filtering recommendation system based on core node of overlap community detection," *Journal of Shenyang University of Technology*, 2013.
- [6] M. S. Cai, X. Li, and Y. T. Yin, "Hybrid top-N recommendation method based on social user tag," *Application Research of Computers*, vol. 30, no.5, pp. 1309-1308, 2013.
- [7] T. Fischer, K. B. Brothers, P. Erdmann, et al., "Clinical decision-making and secondary findings in systems medicine," *Bmc Medical Ethics*, vol. 2016, no.17, pp. 1-12, 2016. <https://doi.org/10.1186/s12910-016-0113-5>
- [8] H. Zhang, S. Zheng, and H. Xu, "Study on the Web Courses Resources Recommendation System based on WP Algorithm," *International Journal of Emerging Technologies in Learning*, vol. 11, no.6, pp. 62-67, June 2016. <https://doi.org/10.3991/ijet.v11i06.5666>

AUTHORS

Xin Gao, Huang Wen-xue, Ning Wang, Yan-chao Yang, and Ying Yan are with North China University of Science and Technology (g200478@163.com)

Submitted 09 September 2016. Published as resubmitted by the authors 27 October 2016.