

A Heuristic Collaborative Filtering Recommendation Algorithm based on Book Personalized Recommendation

Chaoyang Ji*

College of Information Engineering, Xuchang University, Xuchang, 461000, China

Abstract

Through data mining technology, the design of intelligent and personalized book recommendation system is an important development direction of scientific library management in the future. This paper proposes a heuristic collaborative filtering recommendation algorithm based on book personalized recommendation and data mining technology. The proposed algorithm calculates the similarity between users by inputting the two-dimensional matrix of user items and using the similarity formula to get the set of user preferences, and finally generates a recommendation list for each user. The simulation results fully show that the proposed collaborative filtering recommendation algorithm has strong personalized recommendation function, can mine the relevance between readers and books, and recommend suitable book information according to readers' personal preferences.

Keywords: book personalized recommendation; collaborative filtering recommendation algorithm; heuristic collaborative filtering recommendation algorithm; personalized recommendation system; data mining

(Submitted on May 22, 2019; Revised on October 10, 2019; Accepted on November 24, 2019)

© 2019 Totem Publisher, Inc. All rights reserved.

1. Introduction

As the core of library information construction, library management system has become the focus of all kinds of information gathering and radiation in colleges and universities with its massive data and huge throughput. Book recommendation system is a subsystem of book management [1-2]. Designing an intelligent and personalized book recommendation system through data mining technology is an important development direction of scientific library management in the future. Personalized recommendation services have gradually moved from academic research to practical application, and gradually become the focus of industry attention [3-4].

Early collaborative filtering algorithms require users to submit their own ratings and preference data to the recommendation system to recommend items [5-6]. However, ratings often fail to truly express users' interest, which limits the development of collaborative filtering technology. In this paper, the main task is to scientifically manage the reader information, book information and borrowing information of the library, and to intelligently analyze the user's interests and hobbies based on the user's preferences for book categories, and to recommend books that meet users' needs comprehensively and accurately by using collaborative filtering technology.

The specific contributions of this paper include:

- A literature survey about various personalized recommendation algorithms.
- A heuristic collaborative filtering recommendation algorithm is proposed, and the recommendation principle is analyzed in detail.
- The performance of the proposed algorithm is analyzed and compared with other existing algorithms.

The rest of this paper is organized as follows. Section 2 discusses related work about data mining technology and

* Corresponding author.
E-mail address: jcypaper@163.com

personalized recommendation algorithms. Section 3 introduces the collaborative filtering recommendation algorithms and recommendation principle. The heuristic collaborative filtering recommendation algorithm is proposed in Section 4. Section 5 shows the simulation experimental results. Finally, section 6 summarizes the work of this paper and looks forward to the future work.

2. Related Works

Since the end of the 20th century, how to apply data mining technology to libraries has become a hot research topic in the International Library circles. Subsequently, many library-related technical theories and applications have been proposed by experts and scholars in the library circles. My library is based on query analysis and was developed by the library of Cornell University, a famous university in the United States. The system is a network personalized book recommendation system [7-8]. The personalized information service system consists of two parts: Personalized Connection and Personalized Update. Personalized connections are digital resources that allow users to collect and organize themselves from anywhere on the Internet. Personalized updating refers to informing readers of new resources in library, such as periodically sending new books, periodicals or announcements, which may be of interest to readers, so as to save readers' valuable time and increase their interest in using them. The Library of the University of Chicago in the United States has developed a system called Arrow smith, which has the function of distinguishing between words in biomedical literature.

China's recommendation system includes Dangdang's online book recommendation system, Jingdong's and Taobao's commodity recommendation system, Renren's friend recommendation system, and Interactive Publishing's online bookstore book recommendation system [9-10]. The continuous development of data mining technology promotes the application of recommendation system technology in library personalized information service, such as Huiwen Document Information Service System, which has simple data mining and statistical analysis functions. The National Science Digital Library of the Chinese Academy of Sciences, in providing personalized information service for readers, has adopted two ways of user customization and system recommendation, which have achieved good recommendation effect.

Collaborative filtering algorithm starts with memory-based collaborative filtering. First, the algorithm constructs a two-dimensional matrix by using the interaction between users and objects, calculates the similarity between users or objects, then searches for neighboring users, and finally makes recommendations. This algorithm can't really reflect users' preferences under the condition of fewer scores. Therefore, scholars at home and abroad propose model-based recommendation algorithms based on data mining, artificial intelligence and other technologies. For example, by combining clustering algorithm with collaborative filtering, items with similar interests or attributes can be grouped together and then recommended by collaborative filtering algorithm. This method can solve the problem caused by data sparsity.

Model-based collaborative filtering solves the problem of sparse scoring to a certain extent, and the recommendation effect of the algorithm has also been greatly improved. However, these algorithms are generally complex and take a lot of time to calculate. The usual method is to model offline, analyze users' preferences online, and then give recommendations. However, with the rapid growth of data in various industries, the limitations of model-based recommendation are becoming increasingly prominent. Therefore, many scholars have also given a lot of improvement measures, such as mixing several algorithms to achieve better recommendation effect.

3. Collaborative Filtering Recommendation Algorithms and Recommendation Principles

Collaborative filtering recommendation algorithms can be divided into user-based collaborative filtering and item-based collaborative filtering. User-based collaborative filtering algorithm obtains users' hobbies in this area through the historical records of books borrowed by users, and then classifies and evaluates these hobbies. In addition, we can also get the possible relationship between these users by their liking for the same kind of books or content, so that we can recommend similar books among users who have the same hobby.

The recommended estimates value sequence $\hat{X}^{(0)}(k)$ can obtained from the residual sequence of the system $E^{(0)}(k)$, which is

$$\begin{aligned} E^{(0)}(k+1) &= X^{(0)}(k) - \hat{X}^{(0)}(k) \\ &= [e^{(0)}(1), e^{(0)}(2), e^{(0)}(3), \dots, e^{(0)}(n)], \quad k = 1, 2, \dots, n \end{aligned} \quad (1)$$

Figure 1 is a schematic diagram of user-based collaborative filtering algorithm.

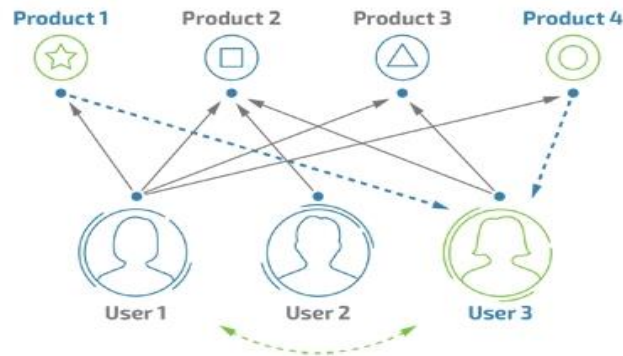


Figure 1. The schematic diagram of user-based collaborative filtering algorithm.

Item-based coordinated filtering algorithm is to query the borrowing history of each book for scoring. The scoring here represents the evaluation and hobbies of each user. If user A borrows books 1 and 2 at the same time, the contents of books 1 and 2 are highly correlated. When user 2 also borrows book 1, it can be seen that he/she should also have the need to borrow book 2. Among the recommendation status series of many books and the state equation of the recommendation chain can be given as follows:

$$Q \frac{dW}{dt} = Q \cos \alpha - \lambda g \sin \beta \quad (2)$$

$$Q \frac{dV}{dt} = Q \sin \alpha + \gamma g \cos \beta \quad (3)$$

In which, Q is the mass of the force unit in recommendation series.

Figure 2 is a schematic diagram of an item-based collaborative filtering algorithm.

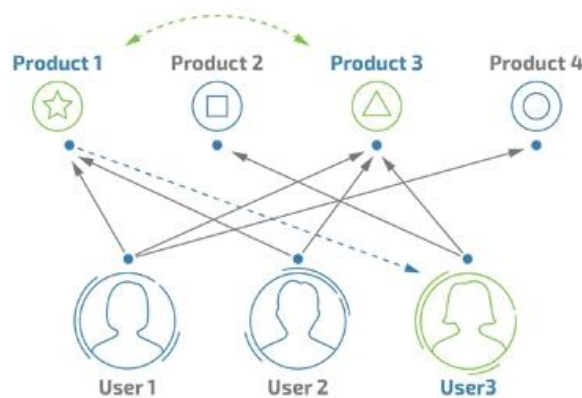


Figure 2. The schematic diagram of an item-based collaborative filtering algorithm

Books recommendation algorithm first obtains users' hobbies by consulting users' historical information. Based on different hobbies, users are individualized and recommended books suitable for their hobbies. A large number of borrowers complete the process of borrowing and returning through circulating libraries, forming a borrowing history. At first, the recommendation system will remove some unreasonable information, and then mine the data. When the borrower inquires about the personal center of the library, the system will give a reasonable recommended catalogue by comparing the reading history of the borrower. The book recommendation principle design diagram is shown in Figure 3.

4. Heuristic Collaborative Filtering Recommendation Algorithm

The first step of data mining is data preparation. The quality and source of data affect the recommendation effect of recommendation system to some extent. Log collection of digital library involves user's basic information, such as reader's name, gender, class, library card number, birth date and specialty, book's basic information such as book name, book

number, book classification number and related information, user's online inquiry information such as card number, date, time, book name and bar code, etc.

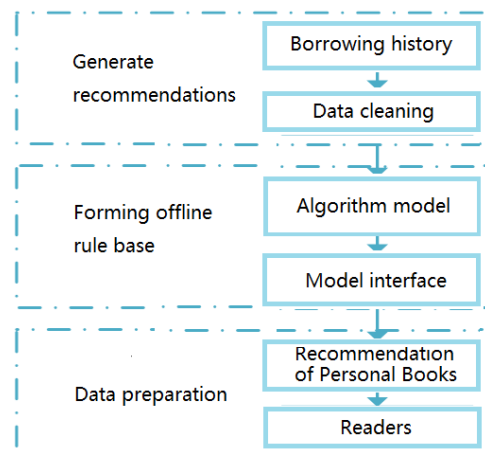


Figure 3. Book recommendation schematic diagram

The second step is data cleaning. When preprocessing data, we should consider all kinds of factors that may affect the mining results, and minimize the interference factors as far as possible. Data cleaning removes the data that are not related to each other, reduces the interference items and reduces the difficulty of data mining. Because there will be a double name between the reader and the reader, the borrowing card number must be taken as the object of mining, rather than the reader's name as the object of mining alone. After unifying and standardizing the data and eliminating irrelevant items, it is necessary to transform the data. Because the format of Microsoft is exported, the database of Microsoft is used in this paper, so the data should be imported into the database to effectively improve efficiency.

The purpose of data filtering is to remove irrelevant information from the original data and avoid unnecessary troubles in the mining process. Data filtering includes three aspects. First, clear the erroneous data. Second, data generation process will inevitably have some errors; these erroneous data will affect the accuracy of the results, so it must be cleared. Third, data conversion; some of the original data formats are not suitable for direct data mining. Simplified data can delete some irrelevant data or merge some of the same data, thus reducing the difficulty of data mining.

In this paper, the principle of collaborative filtering algorithm based on users and projects is applied to practice, and the real borrowing history of a library reader is taken as data reference. Based on the traditional collaborative filtering recommendation algorithm, a new heuristic collaborative filtering recommendation algorithm (HCFRA) is proposed in this paper. The basic idea of the algorithm is to input the two-dimensional matrix of user items, calculate the similarity between users by using similarity formula, get the set of user preferences, and finally generate a recommendation list for each user. The implementation steps of the proposed algorithm are as follows:

Step 1 Establish the two-dimensional matrix model of readers and books. Establish the database Recommend Books, which is used to store all kinds of data tables used in the process of recommendation system.

Step 2 The reader's borrowing records, which are derived from the library circulation system and cleaned up, are imported into the Borrow table already established.

Step 3 Readers borrow the history table Borrow to read the card number and user name data to the Users table. Define the dynamic array `ArrayList users = new ArrayList ()` to store the card number. Define the dynamic array `ArrayList usernames = new ArrayList ()` for storing user names. The card number field of the User table is read into the users dynamic array, and the user name field is read into the users dynamic array.

Step 4 Read the title of the book from the Borrow table to the Books table. Define the dynamic array `ArrayList books = new ArrayList ()` for storing book information, and read the title of the book from the Books table to the dynamic array books.

Step 5 Through the above data preparation, a two-dimensional matrix model of readers and books is established. A two-dimensional matrix `borrow [users. Count, books. Count]` is defined to store borrowing history. The number of rows

equals the number of users and the number of columns equals the number of books. Initialize the borrow [users. Count, books. Count] two-dimensional array and initialize the values to 0. The borrowing records of Borrow table are filled into the two-dimensional matrix borrow [users. Count, books. Count], and the location of the borrowing records is marked as 1.

Step 6 Find the nearest neighbor user set.

- In user-based collaborative filtering algorithm, user similarity is defined by the number of books they borrow together, and the ultimate goal is to calculate the user set similar to the target user.
- Computing user similarity. Firstly, we define a users. Count row, 20 columns of two-dimensional moments top20 booknums [users. Count, 20] to save the 20 users with the greatest similarity. The similarity is reduced in order from high to low. A two-dimensional matrix top20cardnums [users. Count, 20] is defined to store the user information corresponding to the two-dimensional matrix top20booknums.
- By calculating the similarity above, we get up to 20 similar users, and insert up to 20 user information and similarity values into the database table UserSimilarity.

Step 7 Generate book recommendation. The recommendation result is inserted into the data table RecommendBooks ByUCF, which is borrowed by neighbors but not by target users.

5. Experimental Results and Data Analysis

The key point of system pressure test is to test the pressure bearing capacity of the system. In the process of operation, the system will definitely be impacted by the peak flow period. Therefore, the pressure test also has the function of checking whether the system runs stably. This stress test takes login function as an example to conduct concurrent stress test. Apache JMeter is used to simulate 450 concurrent users to log in and log out. A concurrent user will be added every second and the system will run circularly. The JMeter listener provides access to the information collected by JMeter during the test. It can view the result tree to monitor the test operation, display the details of the sampler request and response, and display the basic part of the system response in HTML and XML format.

In order to test the performance of the proposed algorithm, three experimental scenarios are set up, and the performance comparison between HCFRA and CFRA (Collaborative Filtering Recommendation Algorithm) is also made. The first group of experiments compares the response time of the two recommendation algorithms. The experimental results are shown in Figures 4-5. The second group of experiments compares the execution time of the two recommendation algorithms. The experimental results are shown in Figures 6-7. The third group of experiments compares the accuracy of the two recommendation algorithms. The experimental results are shown in Figures 8-9.

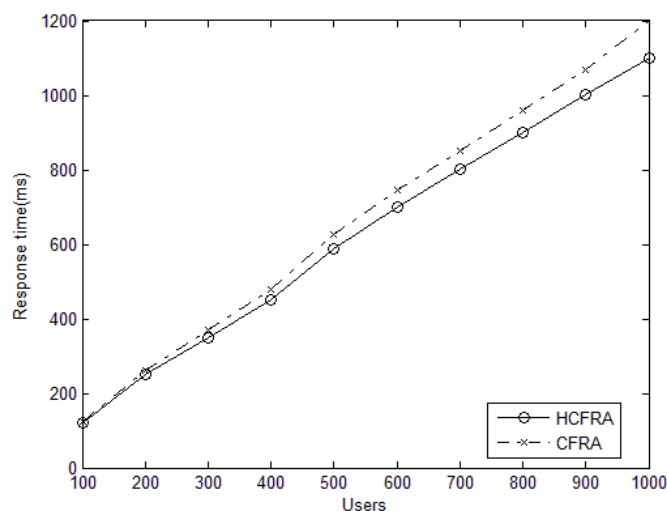


Figure 4. Comparison of the response time for the two recommendation algorithms in different user numbers

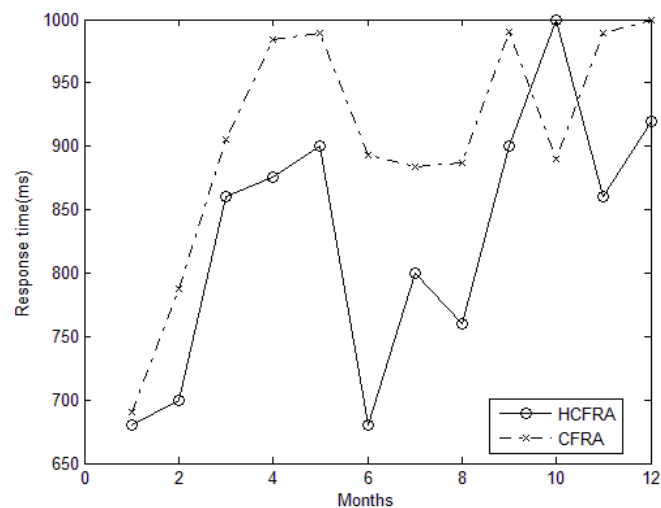


Figure 5. Comparison of the response time for the two recommendation algorithms in different months

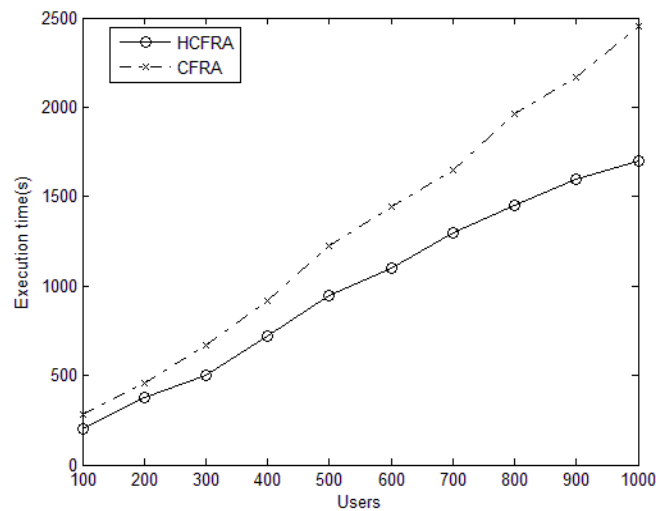


Figure 6. Comparison of the execution time for the two recommendation algorithms in different user numbers

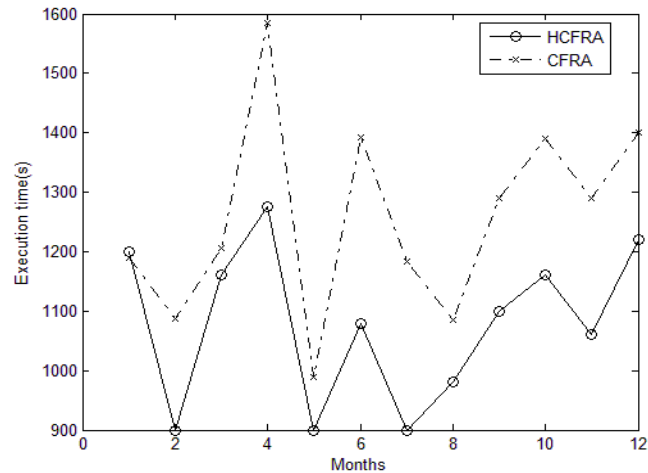


Figure 7. Comparison of the execution time for the two recommendation algorithms in different months

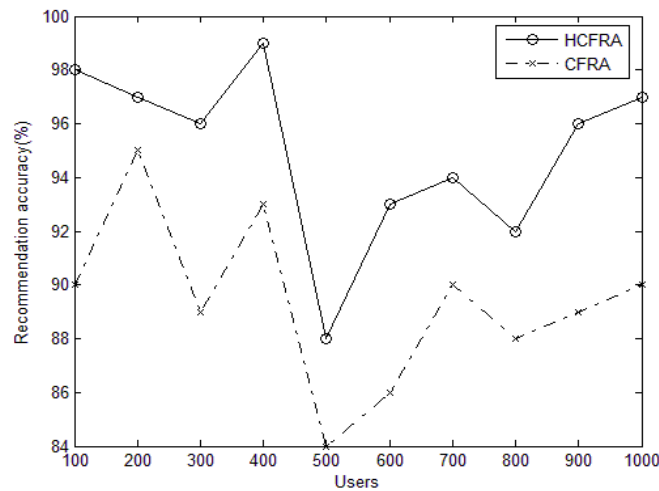


Figure 8. Comparison of the accuracy for the two recommendation algorithms in different user numbers

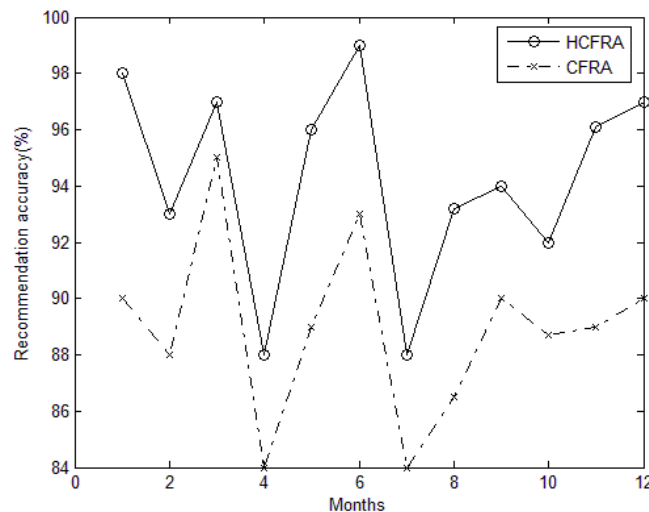


Figure 9. Comparison of the accuracy for the two recommendation algorithms in different months

From the experimental results of Figures 4-9, it can be seen that the corresponding results of 1 350 samplers are obtained during the test run. The response time of the results has reached 1ms through the latest sample of the sampler. The standard deviation of response time is 950ms from all sampler results, the number of sampler results per minute response is 17016.807/minute, and the median response time is from all sampler results.

6. Conclusions

In the era of big data, information will show explosive growth. By constantly tracking users, analyzing and recording keywords of their interests, and through data mining technology, the user's personalized needs are optimized to achieve the goal of customization with characteristics. This paper proposes a heuristic collaborative filtering recommendation algorithm based on book personalized recommendation and data mining technology. The simulation results fully show that the proposed collaborative filtering recommendation algorithm has strong personalized recommendation function, can mine the relevance between readers and books, and recommend suitable book information according to readers' personal preferences. In the future research work, we will recommend the book recommendation subsystem based on collaborative filtering algorithm as a key module to users, and constantly improve the algorithm through customer feedback.

Acknowledgements

This research is supported by Science and Technology Research Project of Education Department of Henan Province.

References

1. P. Mrutyunjaya and R. P. Manas, "Mining Association Rules for Constructing Network Intrusion Detection Model," *International Journal of Applied Engineering Research*, Vol. 4, No. 3, pp. 381-398, March 2009
2. Y. L. Chen and C. H. Weng, "Mining Fuzzy Association Rules from Questionnaire Data," *Knowledge-based Systems*, Vol. 22, No. 1, pp. 45-56, January 2009
3. M. Lamine, N. Aouadi, and T. M. Kechadi, "Performance Study of Distributed Apriori-Like Frequent Itemsets Mining," *Knowledge and Information Systems*, Vol. 23, No. 1, pp. 55-72, January 2010
4. M. Avouk, "Cloud Computing—Issues, Research and Implementations," *Journal of Computing and Information Technology*, Vol. 16, No. 4, pp. 235-246, August 2008
5. S. Marston, Z. Li, S. Bandyopadhyay, J. Zhang, and A. Ghalsasi, "Cloud Computing—The Business Perspective," *Decision Support Systems*, Vol. 51, No. 1, pp. 176-189, January 2011
6. Z. Q. Huang, J. L. Zhang, and H. Z. Zhou, "Preliminary Discussion on the Applications of Cloud Computing in the Bank System," *Applied Mechanics and Materials*, Vol. 5, No. 8, pp. 273-277, May 2011
7. F. Ozdamli and H. Bicen, "Effects of Training on Cloud Computing Services on M-Learning Perceptions and Adequacies," *Procedia - Social and Behavioral Sciences*, Vol. 13, No. 4, pp. 5115-5119, August 2014
8. Q. Alfonso, C. Andrea, G. Antonella, and D. Daniele, "Hybrid Clouds Brokering: Business Opportunities, QoS and Energy-Saving Issues," *Simulation Modelling Practice and Theory*, Vol. 39, No. 2, pp. 121-134, February 2013
9. K. Yacine, M. Nouredine, and E. G. Talbi, "A Multi-Start Local Search Heuristic for an Energy Efficient VMs Assignment on Top of the OpenNebula Cloud Manager," *Future Generation Computer System*, Vol. 29, No. 1, pp. 1-20, January 2013
10. M. Q. Dang, M. Federico, S. Domenico, and G. Raffaele, "T-Alloc: A Practical Energy Efficient Resource Allocation Algorithm for Traditional Data Centers," *Future Generation Computer Systems*, Vol. 28, No. 2, pp. 791-800, February 2012

Chaoyang Ji is received the MS degree in Computer Science from Central China Normal University. He is an associate professor of Computer Science in the College of Information Engineering, Xuchang University. His current research interests include Artificial Intelligence and Information Retrieval.