

## Simulation research of university library recommended system based on big data and data mining

Rui Li

HongHe University, Yunnan, PR China

Lirui009@163.com

**Keywords:** University library; recommendation system; information visualization; data mining

**Abstract.** This paper based on the specific needs of the university library, analyzes construct necessity, feasibility and technical solutions to select automatic recommendation system for books, and a brief introduction of collaborative filtering algorithms, in particular the correlation calculation and different evaluation criteria. Then nearly 5 million loan offline test history data validate our ideas through the use of the university library, and focused on the specific needs due to the university library, the recommended herein may be based on local search for similar user, saves computing resources and at the same time without sacrificing significant accuracy. Finally, we have also introduced some of the content related to the implementation, including the consideration of a cold start and user interaction design, and lack of presence through the implementation of current research findings. This article provides some practical guidance and reference value for the field of application of the recommendation system of university library.

### Introduction

Along with the rapid development of information technology, we have entered an era of information technology [1]. In this context, automation and information technology work university library is moving forward steadily [2]. China's digital library, in particular the construction of university digital library is tight with the development of digital libraries abroad and development. Recommended system library constructed meaning can be summarized as follows: (1) a supplement to the traditional library book lending service, pushed the books need readers, making library services more humane; (2) enhance the collection of books utilization, the system automatically recommend books to readers, readers can explore the potential reading needs, reduce idle books, and enhanced frequency borrow books; (3) to promote interaction between the reader, through the analysis of records calculate the degree of similarity between the reader, the reader Available similar reading preferences of each other to each other, to help users identify potential book club [3, 4].

At present, China has been some discussion papers library recommender systems. Based on the credibility and actual reference value considerations, we focus on selected articles contain real experimental data, they are mainly divided based on clustering, collaborative filtering and correlation analysis techniques [5]. Currently the domestic university library personalized service system in general is relatively simple, the content and form of service is relatively simple, mostly only to provide users with customized information services, and unable to borrow on the user's log analysis, and as a basis initiative to provide intelligent book recommendation service [6]. Compared with the current development of foreign university library personalized services, domestic university library there are obvious gaps.

This paper discusses the necessity of recommendation systems applied in the university library, the university library on the basis of full investigation recommended a systematic study on the status quo at home and abroad, introduces the principle recommendation system, illustrating the recommendation algorithm design and information visualization technology in the recommended system, and elaborated on the design and implementation process recommended by the university library system.

## System Key Techniques

**Django Web Framework.** Django is a using simple, pragmatic design style, suitable for advanced Python Web framework for rapid development [7, 8]. Django built-in user management, group management, messaging systems and other site components can be multiplexed. Developers no longer need to write cumbersome using repetitive code Django development, and the ability to apply part of the page shows part of the code and logic implementation separated, thereby improving the efficiency of the code written [9, 10]. Figure 1 shows the Django framework for processing HTTP requests and generate responses. HTTP request from the user's browser will first be processed through ModPython module, followed by the form of HTTP request object is passed to the requesting middleware. Improper HTTP request object request might trigger an exception and generate status code 404 or 500 response object, the response object created in this step will return directly to ModPython modules via the response middleware. Next, the correct HTTP request object is passed according to URL View middleware configuration to be processed by the appropriate View and generate a response object, the response object is generated by an exception after exception middleware in this process may be triggered to capture. Response object returned by the response middleware ModPython module will eventually HTTP response back to the user's browser.

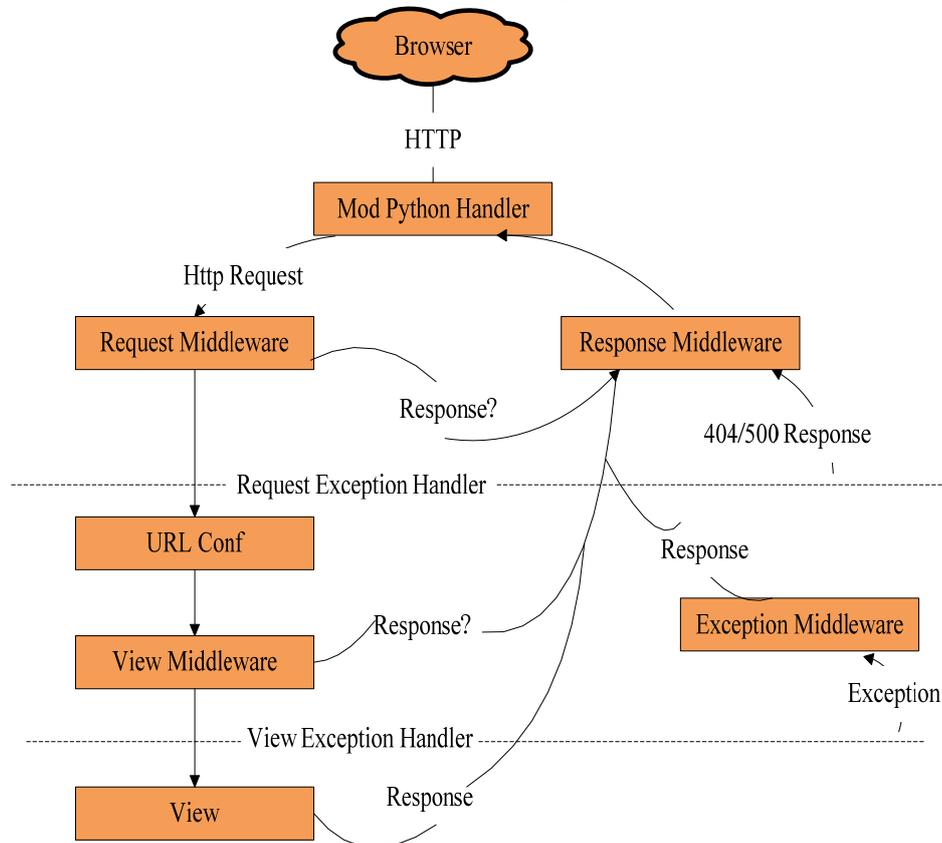


Figure 1. Django framework for processing HTTP requests and generate responses

**System Requirements Overview.** Description from the user needs recommended by the University Library System, through the analysis and recommendation system refined obtain library use case model shown in Figure 2. Participants recommended system library contains the class 2 Roles: Readers system administrator.

## University Library recommendation system

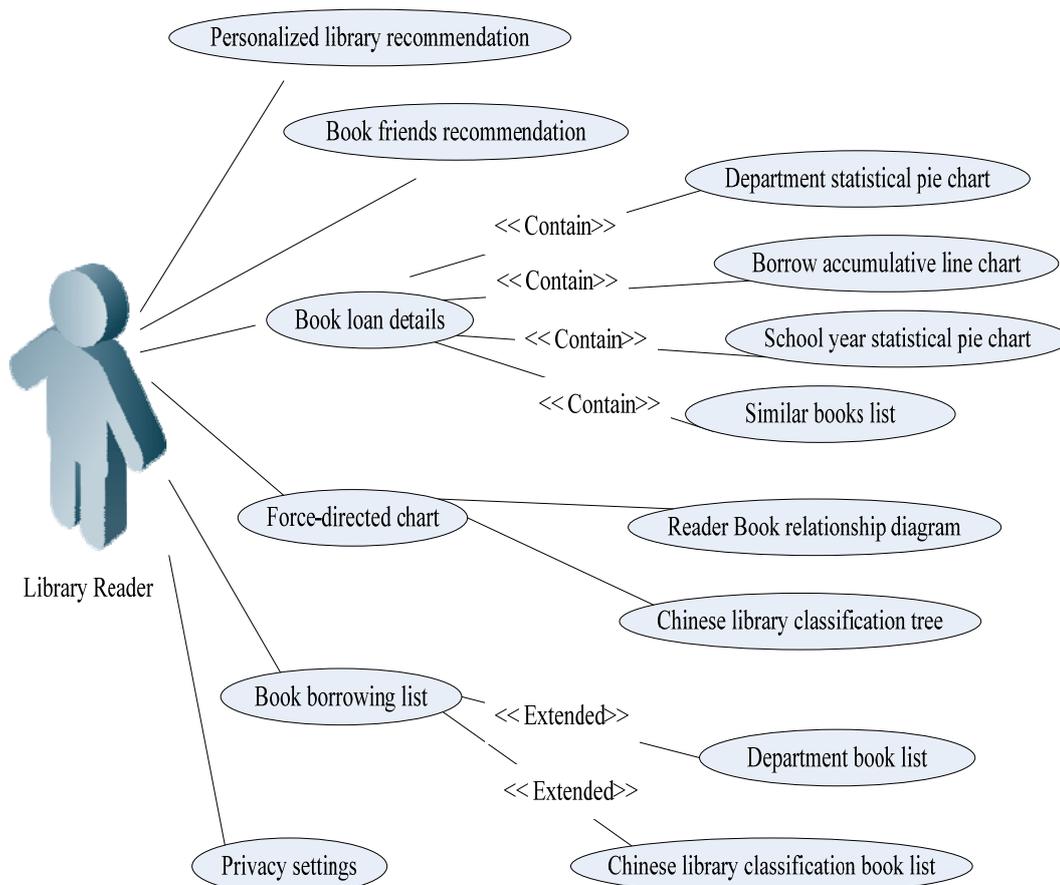


Figure 2. Use case diagram of recommendation system

Readers can obtain personalized book recommendation system recommended by the library information and book club recommendation information; readers can view the details borrow books, including faculty statistical pie charts, line charts cumulative lending, sub-year statistical pie, similar to a list of books; readers can view the reader of books and in the diagram in Figure classification tree; readers can view charts borrow books, including the chart and diagram in Figure faculties classify books list.

## Experimental results

**Data set.** In this paper, the experimental data from the University Library in 1998 - 2011 users borrow logs, user information and book information data. Summary statistics of the data in Table 1.

Table 1. Offline experimental data set summary statistics

All only books #	445722	Cumulative number of loan #	4932579
Have been borrowed books #	211776	Average borrowing amount of the user with the borrowing history #	≈ 75
Accumulated user #	76524	Proportion of books that have a history of borrowing #	≈ 47.51%
Users who have a history of borrowing #	65483		

Current data, due to historical reasons unknown (possibly early readers borrow logs and no storage) there are 11,041 users do not have to borrow logs. Even if these users exclude users - borrowing books matrix is still very scarce.

**Experimental methods and objectives.** In this paper, a 10-fold cross validation method, the data set is divided into ten parts, each decimation do a test set, and the remaining nine were used for the training set. Experimental take collaborative filtering method based on collaborative filtering method users. The purpose of this is to choose: (1) library users is relatively stable, generally one year have added a number of new users; (2) have to borrow record number of 65,483 users, the number of books to 445,722, the number of users is far less than the number of books, computing users inter similarity between the books than the calculated similarity save a lot of time; (3) due to the professional, the user demand for books but also because of the professional relationship is different. Accuracy different topN was shown in Figure 3.

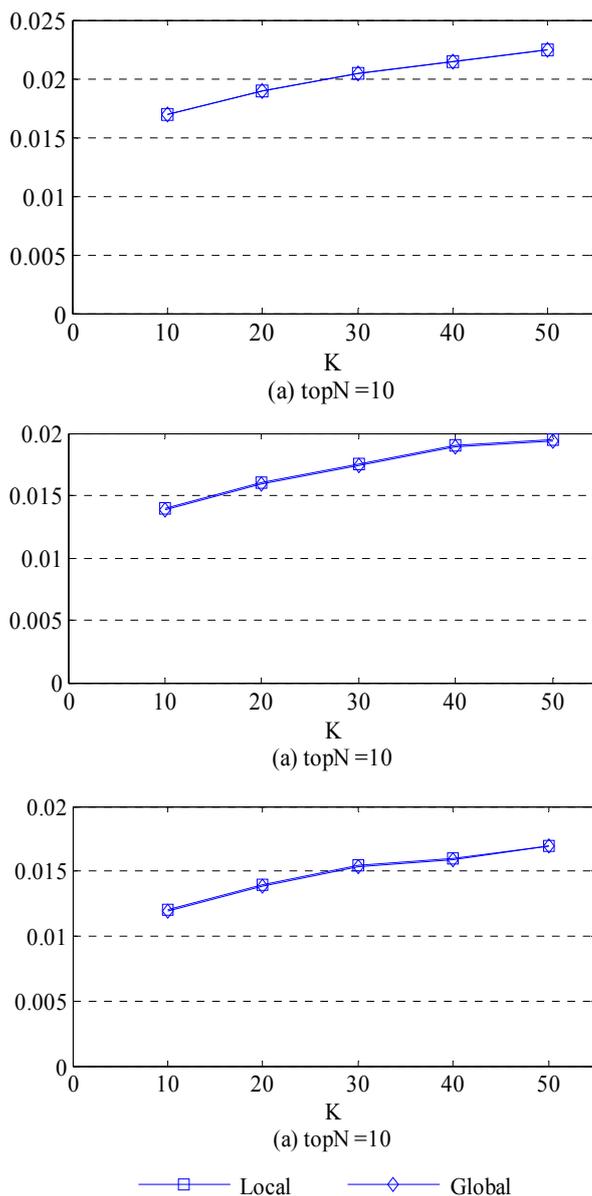


Figure 3. Accuracy different topN

School loan data for each user on average borrow 75 books; the classified data material by professional colleges in each user's average borrowing 75; the classified data according to specialty School of Public Administration in the average lending 80 per user. Whether global data or local data, the user can be seen as the average loan amount is the density matrix of user ratings, Materials Science, and all the data on the same density, School of Public Administration is even higher, but the result was different with varying densities. The impact of the results of the reason may be because the amount of data is not enough, can not contain sufficient "information."

## Summary

Based on actual user data discussed borrow the necessity and feasibility of the introduction of the University Library collaborative filtering recommendation system. University of borrowing by using historical data offline user experiments and found that due to the high professionalism of the reader background, may be based on historical data with college readers to recommend to other readers the same college, and almost without sacrificing accuracy. This discovery will our neighbor search space greatly reduced, thus saving enormous computing resources. In addition, the findings may also solve the common problem of cold start. In this paper, the results of those discussions or research fellow practitioner have a certain reference value.

## References

- [1] Qiang Li, Zuomi Shen, Ling wang Gao, Huiyong Yang, Wei Liu, Jian Qiao. An Internet-Based expert system platform for assistant identification of agricultural pests [C]//World Automation Congress (WAC),2010 International Conference on IEEE.2010:141-144.
- [2] Rahimi-Vahed A, Mirzaei A H, "A Hybrid Multi-Objective Shuffled Frog-Leaping Algorithm For A Mixed Model Assembly Line Sequencing Problem", Computers and Industrial Engineering, , vol.53,no.4, pp.642-666, 2007.
- [3] C.C.Aggarwal, P.S.Yu. "Data Mining Techniques for Associations." Clustering and Classification. 2002, pp.22-67.
- [4] Anu Vajdyanathan. "Malcolm Shore and Mark Billinghurst." Data in Social Network Analysis. 2008, pp. 23-25.
- [5] Monika Henzinger. "Link Analysis in Web Information Retrieval." IEEE Data Engineering Bulletin. Sep 2000, pp. 3-8.
- [6] M. Sadek, A. Tarighat, A. H. Sayed, "A Leakage-based Precoding Scheme for Downlink multi-user MIMO Channels" , IEEE Transactions on Wireless Communications, vol. 26, no.8, pp.1505-1515, 2008.
- [7] Zhu Yazhou, Zheng Guoxin, Rui Yun, , Li Mingqi, "A Novel Distributed Precoding Scheme Based on THP for Downlink Multi-Cell Multi-User OFDMA Wireless Systems", IJACT: International Journal of Advancements in Computing Technology, vol. 5, no. 9, pp. 213-220, 2011.
- [8] Raore Soungalo, Li Renfa and Zeng Fanzi, "Evaluating and Improving Wireless Local Area Networks Performance", IJACT: International Journal of Advancements in Computing Technology, vol. 3, no. 2, pp. 156-164, 2011.
- [9] T. Okamoto. A digital multisignature scheme using bijective public-key cryptosystems. ACM Trans. Computer Systems, 6(4), ACM Press, New York, 1988, pp.432–441.
- [10] A. Boldyreva. Threshold signature, multisignature and blind signature schemes based on the gap-Diffie-Hellmangroup signature scheme. In Proceedings of PKC 2003, LNCS 2567 , Springer, Berlin, 2003, pp.31–46.