

# Research on the Construction of Personalized Resources Recommender Platform in University Based on Big Data

Guang Liu<sup>1, a</sup>, Xu Zhao<sup>2, b</sup>

<sup>1</sup> Key Laboratory of Modern Teaching Technology, Ministry of Education, P. R. China, Network Information Center, Shaanxi Normal University, Xi'an 710119, China;

<sup>2</sup> Library, Shaanxi Normal University, Xi'an 710119, China.

<sup>a</sup>liuguang@snnu.edu.cn, <sup>b</sup>zhaoxu@snnu.edu.cn

**Keywords:** Big Data, Recommender System.

**Abstract.** Big data is one of the most popular concepts and technologies in recent years. It has brought profound influence to our work, life and study. This paper studies the construction of personalized resources recommender platform in Universities under big data environment. The first is to solve the difficulties faced by teachers and students in searching and selecting of resources in research, learning and work, and through the platform to provide a precise personalized resources recommender services for teachers and students in the school. At the same time, it is also aimed at further improving the utilization efficiency of various resources and promoting the informatization construction in universities. This paper introduces the related domain knowledge of big data and recommender system, expounds the process of collection, storage and calculation of big data platform in colleges and universities, and proposes how to build a personalized resources recommender platform based on big data in universities.

## 1. Introduction

In recent years, the development of mobile applications and the popularity of electronic products make the volume of information data showing explosive growth. The concept of big data has been widely concerned by industry, academia and even governments around the world [1]. In 2012, the United Nations issued the "Big Data for Development: Challenges & Opportunities" [2], which elaborated the profound impact of big data on social and economic development, and expected countries to seize the opportunity to promote social and economic changes and development through big data. The famous Periodicals "Nature" and "Science" published the special issue "Big Data" [3] and "Dealing with Data" [4] respectively for big data and explored the opportunities and challenges of big data.

Big data has triggered a new technological revolution, and it has also profoundly affected the work, life and learning of ordinary people. In universities, with the continuous deepening of information construction, teachers and students are getting more and more big and rich data resources, but also begin to face the problem of resource search and selection: how to quickly and effectively obtain the resources that they need. As an effective method to solve the problem of information overload, the recommender system [5] is playing a significant role in many fields, such as e-commerce, tourist recommendation, news recommendation and so on. To solve the problem of how to obtain the resources quickly, this paper proposes to build a resources recommender platform based on big data of universities through the combination of big data and recommender system to provide a precise personalized recommender service for teachers and students. At the same time, it can also make full use of all kinds of resources in schools to further promote the information construction in universities.

The following chapters are organized as follows: Section 2 gives an overview of big data and recommender system. Section 3 introduced the key technologies of big data and recommender system. Section 4 elaborated on the construction of personalized resources recommender platform based on big data. The last chapter is the summary of this article.

## 2. Big Data and Recommender System

### 2.1 Concepts of Big Data

There is still no uniform definition for the big data. The world-renowned consulting firm McKinsey firstly puts forward the big data era to come in the report “Big data: the next frontier for innovation, competition and productivity” and give a definition: Big data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze [6]. The industry summed big data up as 5V model [7]: volume, velocity, variety, veracity, value. In fact, that is the 5 characteristics of big data, including 5 levels meaning.

Volume means that the amount of data collected and analyzed is very large, from TB level to PB level. Velocity means that processing speed is fast. We need to analyze the data in near real-time. Variety means that big data comes from a variety of data sources, data types and formats are increasingly rich, including structured, semi-structured and unstructured data formats, such as web logs, videos, pictures, geographically located information and so on. Veracity means that the content of big data is closely related to the reality of the real world. The study of big data is the process of extracting and predicting real events from large network data. Value means that value density is low and business value is high. By analyzing the data, we can find out how to seize opportunities and harvest value.

### 2.2 Traditional Recommender System

Recommender system, as an effective tool to solve information overload, is a focus of attention in academia and industry and has been widely applied in many fields [8]. The recommender system, according to the user's needs, interests, and so on, uses the recommender algorithm to mine the items of users interesting in from the massive data and give the result to the user in the form of personalization. Now, the recommender system has been successfully applied in many fields, such as electronic commerce (Amazon, Alibaba etc.), information retrieval (Google, Baidu etc.), social networks (Facebook, Twitter, WeChat etc.), location service (Yelp, Meituan etc.), news push service (Toutiao, Google News etc.) etc.

There are two important stages in the process of generating recommender by traditional recommender systems [9]: data processing phase and recommendation generation phase. In data processing stage, recommender system needs to get preferences from data. In the generation stage, recommender system generates recommender items from data sets according to user preference information and related recommender algorithm.

The traditional methods of recommendation include collaborative filtering, content-based recommendation and hybrid recommendation. With the development of mobile devices, a context aware recommendation method is also presented [10].

### 2.3 Recommender System in Big Data

Recommender systems in big data are the extension of traditional recommender systems. Their core ideas are basically similar. But the big data environment is facing more complicated information providing environment and data characteristics than traditional environment. In big data environment, the speed of data generation is faster, the big data is high dimensional and sparse, and the channel of content sampling is much more. In the fusion of multi-source data, higher noise and redundancy will be introduced due to the difference of structure and acquisition mode. Data structure has also changed. Nonstructural data and semi-structured data become the main data. Streaming data is also a common data type. At the same time, the data content becomes richer, and the recommender system can collect a rich user implicit feedback data. A large amount of data requires more data processing capability, and the rich data enable users to have higher real-time and accuracy requirements for recommender system. As a result, the traditional recommender system cannot be directly applied to the big data environment, so it is necessary to improve and expand the algorithm [11].

### **3. Key Technology of Recommender System in Big Data**

Document [9] gives a basic framework of recommender system in big data environment. Under the big data environment, the recommender system is divided into 4 levels, namely data collection layer, data preprocessing layer, recommender generation layer and utility evaluation layer.

The data processing layer is responsible for preprocessing and calculating the collected data, and the data processing result is used as the input of the recommender system. The main tasks include user preference acquisition, social network construction and context user preference acquisition.

The recommender generation layer is the core of the recommender system. The main task is to introduce and process big data, and to generate the recommended results of high real-time, high accuracy and user satisfaction. Currently, there are the following recommender techniques in big data: recommendation based on matrix factorization, recommendation based on implicit feedback, recommendation based on socialization, and recommendation based on group.

The utility evaluation layer is mainly to present the recommended results to the user, and then combines the user feedback data to evaluate the recommender system by accuracy, real-time, novelty and diversity, and expand and improve it according to the requirements.

#### **3.1 Preference Acquisition Technology in Big Data**

User preference acquisition is the key technology of recommender systems. The accuracy of preference acquisition directly affects the accuracy and real-time performance of recommender systems. Because of the lack of data, the preference acquisition in traditional recommender system is difficult to accurately get users' short-term preferences and cannot accurately capture the dynamic changes of user preferences. Preference acquisition layer can collect richer user data in big data. How to get users' short-term preferences from data is also a research hotspot and difficulty in recommender system under big data environment [12].

#### **3.2 Recommender System Based on Matrix Factorization under Big Data**

In big data, the input data of the recommender system have high dimensional sparsity, high redundancy and noise, and the streaming data is also an important data type. The traditional recommender systems degrade performance very quickly when dealing with such data. Therefore, in recent years, matrix factorization has gradually replaced the traditional recommender algorithm, and has become the research hotspot and mainstream model, mainly because of its accuracy and scalability. The core idea of the algorithm is to transform the recommendation problem into a matrix complete factorization problem. The sparse user rating matrix is mapped to a given set of users and the set of items. The default rating is predicted by the matrix operation, which reflects the user's potential preference to the item. The advantage of matrix factorization algorithm is that it can reduce the sparsity of high dimensional data, and it is insensitive to noise and redundancy and has strong extensibility, but the disadvantage is that the algorithm is poor in interpretative, lack of clear meaning and high computational complexity [13].

In the big data environment, researchers introduce multiple contexts into modeling so that they can handle higher dimensional tensor data. Koren et al. [14] introduces the time context into modeling and proposes SVD++ algorithm to improve the accuracy of predicting user movie ratings. The personalized resources recommender platform based on big data is constructed by SVD++ algorithm.

### **4. Construction of Personalized Resources Recommender Platform Based on Big Data**

The big data and recommender systems are summarized and the key technologies are analyzed earlier. Next, we will elaborate the construction of the platform. Our proposed personalized resources recommender platform based on big data uses the mainstream big data framework Hadoop platform, using this platform to collect, process and analyze the data produced in the campus, obtain the data of the teachers and students' preference data on the research and study of the resources, and combine the mature recommender algorithm to process the big data, form a personalized recommender list.

The framework of personalized resources recommender platform based on big data is designed as shown in Fig.1.

The personalized resources recommender platform based on big data mainly includes data source, data collection and management, data storage layer, computing framework layer, data access layer, service and interface, large data analysis, recommender system and so on.

The data source is the source of the data required for the platform analysis, including various information management systems in the University, such as the library management system, the personnel management system, the scientific research management system and so on. It also contains many logs produced by various network devices, as well as the information related to teachers and students that are collected on the Internet based on relevant topics.

Data acquisition mainly uses various collection tools or interfaces to collect data from different sources to big data platforms, and then clean, organize and process data, and store it to the data storage layer for the upper analysis service.

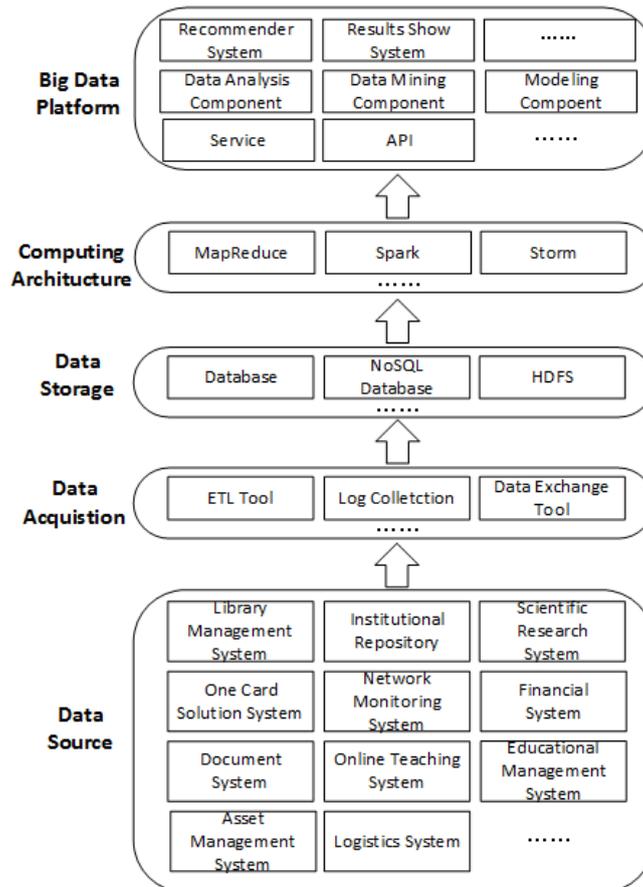


Fig 1. The framework of personalized resources recommender platform based on big data

The data storage layer mainly stores data after cleaning and processing, facilitate the upper layer analysis and use.

Computing framework layer: on the one hand, it accepts the analysis task from the upper level, on the other hand, it connects the data storage layer below, processes all kinds of data according to the analysis task, and gets the result feedback to the upper layer.

Big data platform layer: this layer contains data analysis and mining, recommender system, etc. First, data analysis and mining are based on the user preference data required by the recommender system to analyze and process the underlying data sources, obtain user preference information, and then recommender system use the obtained user preference information data and relevant recommender algorithm to process the data in order to generate personalized recommendation. At last a personalized recommender result list is generated and sent to the related users.

## 5. Summary

The personalized resources recommender platform based on big data is constructed by the popular big data technology. Collect data from different sources, such as various information management

systems, various network equipment logs, and various network data, to the big data platform. Then, the relevant models are used to analyze the data and get the preference of teachers and students in research and learning. Finally, a personalized recommender list is generated for users based on recommender system.

### **Acknowledgments**

Supported by Interdisciplinary Incubation Project of Learning Science of Shaanxi Normal University (SYSX201504).

### **References**

- [1]. Li Guojie, Cheng Xueqi. Research status and scientific thinking of big data [J]. Bulletin of Chinese Academy of Sciences, Vol. 27 (2012) No.6, p.647-657.
- [2]. Big Data for Development: Opportunities & Challenges (2012) [DB/OL].
- [3]. <https://www.unglobalpulse.org/projects/BigDataforDevelopment,2012-05>.
- [4]. Goldston D. Big data: data wrangling [J/OL]. Nature,2008,455:15. <https://www.nature.com/news/2008/080903/full/455015a.html>.
- [5]. O. J. Reichman, Matthew B. Jones, Mark P. Schildhauer. Challenges and Opportunities of Open Data in Ecology [J/OL]. Science,2011,311(6018):703-705. <http://science.sciencemag.org/content/331/6018/703>.
- [6]. Resnick P, Varian H R. Recommender systems[J]. Communication of the ACM, Vol. 40 (1997) No.3, p.56-58.
- [7]. Manyika J, Chui M, Brown B, et al. Big data: The next frontier for innovations, competition, and productivity. [EB/OL]. <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation>.
- [8]. Cheng Xueqi, Jin Xiaolong, Wang Yuanzhuo, et al. Survey on big data system and analytic technology [J]. Journal of Software, Vol. 25 (2014) No.9, p.1889-1908.
- [9]. Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering. Vol.17 (2005) No.6, p.734-749.
- [10]. Meng Xiangwu, Hu Xu, Wang Licai, et al. Mobile recommender systems and their applications[J]. Journal of Software, Vol. 24 (2013) No.1, p.91-108.
- [11]. Wang Licai, Meng Xiangwu, Zhang Yujie. Contextaware recommender systems [J]. Journal of Software, Vol.23 (2012) No.1, p.1-20.
- [12]. Meng Xiang-wu, Ji Wei-yu, Zhang Yu-jie. A Survey of Recommendation Systems in Big Data. Journal of Beijing University of Posts and Telecommunications [J]. Vol.38 (2015) No.2, p.1-15.
- [13]. Rafailidis D, Nanopoulos A. Modeling the dynamics of user preferences in coupled tensor factorization. the 8th ACM Conference on Recommender Systems. Silicon Valley, 2014, p.321-324.
- [14]. Dror G, Koenigstein N, Koren Y, et al. The Yahoo! music dataset and KDD-Cup'11. 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining,2011 ACM SIGKDD. San Diego, 2012, p.8-18.

- [15]. Koren Y. Collaborative filtering with temporal dynamics [J]. *Communications of the ACM*, Vol.53 (2010) No.4, p.89-97.