

The Research of Reader Portrait in University Library based on Big Data

Huazhi Lin^{1, a}, Denghui Zhang^{2, b, *}

¹Department of Book and Information Center Zhejiang Shuren University Hangzhou, Zhejiang, China

²Department of College of Information Science and Technology Zhejiang Shuren University Hangzhou, Zhejiang, China

^a67060675@qq.com, ^{b, *}zdh13957192365@qq.com

Abstract. This paper takes the reader of university library as the research object, analyzes the current research status of the existing libraries, points out the problems existing in the library research process, and proposes the use of big data and user portrait technology to solve the existing problems. Based on the user portrait system architecture under big data, using SPARK technology to integrate offline and real-time data and practiced in a university. We combined with the management experience of higher education industry to design a label system to portrait all readers. Finally, we pointed out the future research direction.

Keywords: big data, library, deep learning, user portrait.

1. Introduction

Big data was first proposed by American data scientist Victor Meyer Schonberg, and pointed out that the biggest change in the era of big data is to abandon the desire for causality and instead focus on related relationships. Gartner Research Institute defines big data (Big data): Big Data is an information asset that requires a new processing model to have greater decision making, insight and process optimization capabilities to accommodate massive, high growth rates and diversification. The McKinsey Global Institute defines big data as a collection of data that is large enough to capture, store, manage, and analyze much beyond the capabilities of traditional database software tools, with massive data scales, fast data flow, and diverse data. IBM proposes 5V features of big data: Volume, Velocity, Variety, Value, Veracity [1]. User portrait is a virtual representation of the real user, is built in the target user model above the series of data (Usability data). Through user research to understand users, according to their differences in goals, behaviors and opinions, they are divided into different types, and then each type extracts typical features, giving names, photos, some demographic elements, scenes, etc., a user portrait (personal prototype) is formed. The user portrait technology was first applied in product design, that is, in the product design process, according to the constructed user portrait (personal prototype) to carry out personalized product design and implementation.

In the context of the era of big data, the information behavior of college readers has undergone major changes. Libraries need to take advantage of big data technology in the new situation, and deeply explore the information needs of readers and provide personalized services in a timely manner.

2. Related Work

The existing researches mainly start from the reader's borrowing method, instead of constructing a complete reader model from the reader's omnidirectional perspective. Many of the research is mainly based on the data owned by the library itself, but does not expand the data source. At present, many scholars in China have studied the borrowing behavior of reader. Zhu. Ding Yong analyzed the use of books in the library for the library borrowing rankings and readers' borrowing rankings in Nanjing Library [2]; Yu Yu used the reporting function of Unicorn system to distribute the library books, book borrowing volume and different contents of Tianjin Normal University Library. The number of readers and the amount of borrowing are used to analyze the reader's borrowing behavior and psychological reasons [3]. Zhang Liping conducts research on the low utilization rate of

collection resources and mainly discusses the authors, publishers, and publication dates that affect users' borrowing behavior [4]. Xiong Lijun conducted a statistical analysis of the reasons for the decline in the number of books borrowed by the readers in the library after the merger of colleges and universities, and the number of books borrowing books [5]. Li Yan introduced social network analysis methods to established the social network relationship matrix of library readers borrowing behavior. On the basis of this analysis, Li Yan built the relationship between the library readers' borrowing behavior network, groups and central figures [6]. Sang Qing bing concerned personalized learning and prediction and point out some big data application points in university [7]. Zhao Liang mainly discusses the application prospects of big data in colleges and universities from educational data mining, learning analysis, personalized course analysis, learning behavior analysis, work-study analysis and school public opinion analysis [8]. Gong Dan is based on Data mining analyzes the role of college student cadres [9]. Li Xianglong and Yang Guifu use campus card data to construct user digital behavior trajectory [10]. Yan and Zhao Zeyu study the bringing of big data from the perspective of building smart campus services. They also point out that in the process of information in universities, the establishment of data collection and application mechanism with “people” as the core. The exploration of data standards based on the accumulation of basic data, the improvement of data production and consumption cycle mechanism, and the protection of data security for data protection [11]. At present, there is little research on libraries based on big data. This article is different from previous research in focusing on library readers. We take library readers as the core of research and construct reader models using user portrait., and use the recommendation system to carry out more effective knowledge services.

In the current library management, it is urgent to adopt new technologies and new methods to build a scientific and complete reader model to improve library management and knowledge service efficiency, library collections utilization and reader satisfaction thus improve the overall competitiveness of the university.

3. Proposed Model

This article makes user portraits of library readers under the conditions of big data and global data, better understands and serves readers, and carries out competitive and personalized services, thus solving the problem of low utilization of library collections.

This paper focuses on the user's portrait with the reader, using the reader's full information table for user portraits. The user model framework is shown in Figure 1:

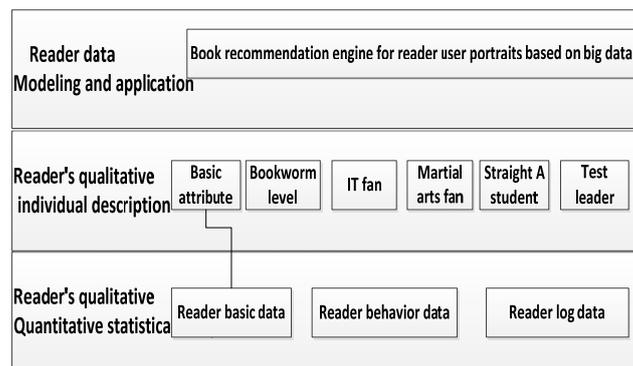


Fig 1.Reader portrait model

The user image is actually mapped from data abstraction to user granularity. The data language is transformed into a concise, visual and understandable phrase tag, which is easy for humans and computers to recognize and understand. Through the reader's user portrait, the basic attributes of the reader, the bookworm level, the interest tendency characteristics, the book preference characteristics, the academic level characteristics, and the consumption characteristics are extracted. According to the reader's differentiation in the borrowed books, the characteristics of each type of readers are mined,

and then the model is built to abstract the reader user characteristics into phrase tags. Under the massive data, the more readers' behaviors, the more features can be extracted. The richer user tags, we can know more about the description of the reader's basic attribute characteristics, interest tendency, book preference. Given enough information, we even can reveal the reader's mood fluctuations and personality to provide readers with accurate personalized book recommendation services.

In addition, when the reader group is divided, it is proposed to improve the clustering algorithm based on the user characteristics constructed by the deep confidence model. The preliminary idea of the method includes the following steps: 1) obtaining the initial network weight of the deep neural network; 2) randomly grouping and mapping the samples to the feature space; 3) adding the objective function of the original deep neural network to the intra-class constraint function of the feature layer 4) update the network weight of the deep neural network, calculate a new feature layer; 5) assign each sample to the cluster of the nearest cluster center, calculate a new cluster center; 6) use a new cluster center Instead of the clustering center of the intra-class constraint function [12], the returning network weight update step is iterated. The final clustering grouping result is obtained and output. And then it can improve the speed and effect of the user grouping.

The advantage of this model are followings:

User portrait based on big data use a quantitative and qualitative portrait technique to build a reader model. The reader's portrait is multi-dimensional in terms of the reader's frequency of accessing the library, such as using the search engine to retrieve the retrieval information of the book, the statistical information of borrowing books, using various electronic resource library information, living consumption information, educational affairs related information, and scientific research related information. The avatar of the target reader is constructed, so that the library can better identify and understand the reader. so that they can provide a good knowledge service and reading experience, and then realize the active push and active guidance of the library service.

Based on the big data background also means in view of global and full-scale, the continuous development and extension of big data technology, so that can promote the informationization of university. The big data processing model is used to realize the dynamic cross-analysis of reader-related data. Through breaking the previous data processing and network bandwidth capabilities, it is impossible to realize the linkage of multiple data and massive data in time. The existing big data platform can realize massive data association, which can more accurately determine the behavior pattern of the group. The correctness of the verification mode can be achieved, and the individualized service of the individual can be realized in consideration of the group.

They can actively extend the service connotation and service scope of the library, analyze the relevant information of the readers, draw the interest points of interest to the users, send the collection information and recommendation materials to specific groups. For the university library, they can provide suggestions for the university decisions such as the opening of classes and elective courses, selecting courses for students and tutors, point out the direction of graduation thesis, provide readers with valuable reference information, guide the research direction of readers to a certain extent, introduce social training institutions and interest alliances as an intermediary.

The technical architecture of this model is as follows: Under the SPARK system, the book recommendation engine system framework based on the reader user portrait is shown in Figure 2:

The user portrait table comes from the reader's statistical information full scale. The reader's statistical information full scale information comes from the offline reader behavior information table and the online behavior statistics information table [13]. The online reader behavior information comes from the log data stream with the offline user bolt collection [14]. The reader behavior information comes from the database hive of the spark. The system can use both historical data and real-time online data.

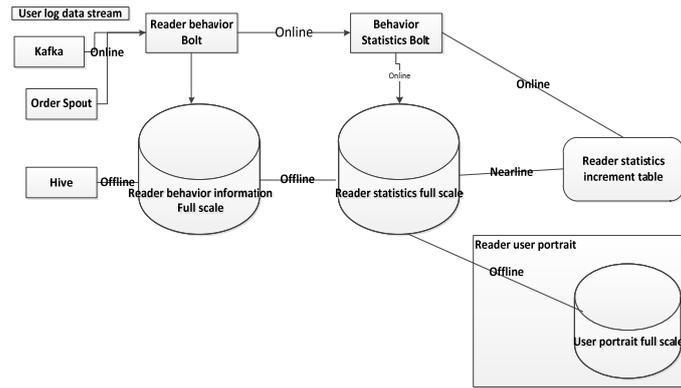


Fig 2 .The recommendation engine architecture based on user portrait

Table 1. If each clicked site has more than 10 times per month, they are tagged

IT related	IT fans	Foreign exchange	Money-grubber	search	Fashion	mailbox	Social		
IP site		fund		entertainment		online chat			
Software automatically collects information URLs		Securities		Entertainment news		Weibo			
operating system				Travel traffic		Blog			
IT industry				car		forum			
software download		Futures market		Sports		faith		Military and weapons	Military fans
Virus database upgrade		gold		religion				Science & Technology	Technology enthusiast
cloud storage	Financial information	Geomantic Omen							
Online Shopping	Consumerism	stock	Game fans	Government	Society	mailbox	knowledgeable		
online banking		bank		Transportation industry		Financial information			
search engine		lottery		Energy industry		News portal			
Business		game		manufacturing		ad			
ad		game information		Legal information		Life information			
Life information		Video and audio		food		Love life		Literary novel	
Merchant		cloud storage		Property					

4. Practice

This paper mainly completes the practice of user portraits. The user portrait is the basis for the recommendation. We divide the data of teachers and students in school into static data and behavior data according to the variability perspective. From the perspective of activities, it can be divided into teaching related data, consumption-related data, and life-related data. These data are scattered in various business systems. Nowadays, through the data integration platform, each data is integrated into the data center. Different from other people's research only individual business data is passed through the data center, we incorporate relational data and non-relational data generated by each business system into the big data platform for unified management and analysis, so that the analysis of users is more comprehensive and accurate. We set the user portrait conditions based on the business of the library. The data mainly includes static data (basic data) and dynamic data (behavior data). The core data includes user-based data, Internet access, card-related information, grades, book borrowing, consumption, number of visits. References to the regularity of university management, readers are tagged.

(1) Internet behavior

According to the type of website the students visit, they are classified and labeled. The following indicators are selected. If each clicked site has more than 10 times per month, they are tagged, as shown in Table 1. The sites to be visited will be classified according to navigation network such as 360 navigation, Baidu website navigation and so on.

(2) Library

According to the different types of books read by the readers, the labeling condition is that the books of the category are borrowed more than once per semester as shown in Table 2. We choose some of the labels.

Table 2. We choose some of the labels

Industrial Technology	Technology	Culture, science, education, sports	Literary fans
Aerospace	enthusiast	literature	
Environmental Science, Safety Science	Society	art	
Transportation			
History, geography	Geography fans	Languages	Benevolent
Astronomy, earth science		Medicine & Health	

(3) Card

The card information is relatively rich, including meals, shuttles, laundry, supermarkets, medical treatment, dormitory electricity, etc., as shown in Table 3. The following information is generally used for each semester as the timeline.

Table 3. Semester as the Timeline

Supermarket frequent guest	Consumption exceeds 200 yuan
Shuttle frequent guest	Consumption exceeds 30 yuan
Infirmery frequent guest	Consumption exceeds 200 yuan
Washing machine frequent guest	consumption exceeds 80 yuan
Power saving pioneer	Monthly consumption is less than 30 yuan
Canteen fans	More than 30 times in the month
Off-campus food fans	Less than 10 times a month.

(4) Grade

The following table is a conditional label. The average grade of the historical selection is the division condition, as showed in table 4.

Table 4. A Conditional Label

Straight A student	The average score is greater than 85.
Average person	Average between 60 and 85
slacker student	Average score below 60

(5) Basic information

The main basis information includes gender, age, region, constellation and birthday. Constellation is widely known and recognized by students.

The above are the main label system. From the perspectives of online behavior, book borrowing, card consumption, Grade, basic information, etc., more than 140 million students in a university have been portrayed. The school is mainly composed of media students. The results of the portraits are shown in Figure 3. It is found that most of the students in the school are off-campus food control. They usually consume less in the cafeteria, use less electricity, and have less time in the dormitory. For this university, he should improve the level of food and beverage to attract student.

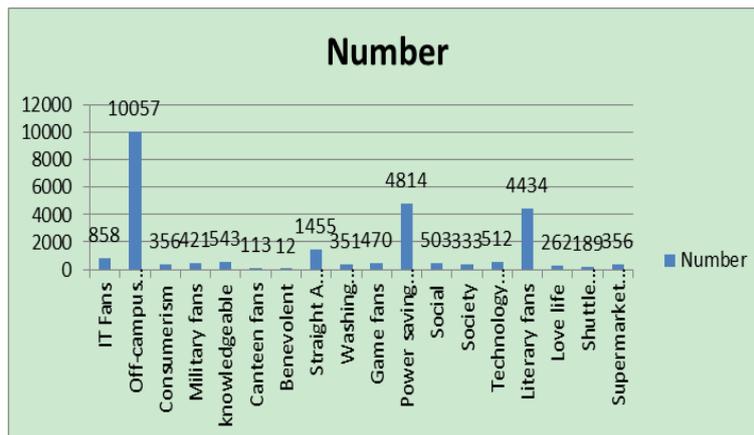


Fig 3. The results of the portraits

5. Conclusion

This paper takes the participants of higher education as the research object, uses the user portrait technology to describe the reader's behavior in the context of big data, fully integrates the data of all parties, and proposes the technical framework and scheme under the condition of big data. In the research, the reader's understanding is deeper, and the students are selected as the practical objects and the user portraits are carried out. However, the research on the relationship between different labels is relatively lacking. The future research directions are:

Collaborative recommendation: such as library recommend books between related tags; Tag threshold should be more accurate and scientific; Providing data support for student management, such as the consumption of the card and borrow book times, reveals its economic ability to some extent, it is helpful for the evaluation of the award; Early warning mechanism: If it is found that students have no behavior data for a period of time, they should promptly inform the counselor to pay attention to it. In short, the user's portrait makes the user's features clearer, which lays a solid foundation for further research.

Acknowledgments

This work was supported by Zhejiang Provincial Natural Science Foundation of China under Grant No.LY16F020021 and Zhejiang Provincial Department of Education under Grant No. Y 20163685.

References

- [1]. Li Xuelong, Gong Haigang. Review of Big Data Systems[J]. Science in China, 2015, 45 (1): 4-30.
- [2]. Ding Yong. Analysis of readers' borrowing behavior in public libraries: Take Nanjing Library as an example [J]. Library World, 2011 (4): 46-49.
- [3]. Yu Yu. A Study on the Reader's Borrowing Behavior Analysis and Guided Reading Strategies in Normal University Libraries[J].Journal of Library and Information Sciences,2011,29(1) : 71-75.
- [4]. Zhang Liping, Yang Lin, Wang Yuhe. Research on Collection Resource Usage Mining Based on User Borrowing Behavior[J].Library Theory and Practice,2014(9) : 9-12.
- [5]. Xiong Lijun. The borrowing statistics of the library in Nanchang University and the reasons for its decline[J]. Chinese Journal of Medical Library and Information Science, 2012, 21(1): 22-24, 34.
- [6]. Li Yan. Research on Reader Borrowing Behavior of University Library Based on Social Network Analysis[J].Innovative Science & Technology,2013(8) : 74-75.
- [7]. Sang Qingbing. Application and Thinking of Big Data in Colleges and Universities [J]. Journal of Nantong Textile Vocational and Technical College (Comprehensive Edition), 2013, 13(2): 84-87.
- [8]. Zhao Liang. Journal of Heilongjiang Institute of Education [J]. Journal of Nantong Textile Vocational and Technical College (Comprehensive Edition), 2014, 33(9): 14-15.
- [9]. Gong Dan. Analysis of the role of college student cadres based on data mining [J]. Heilongjiang Education (Higher Education Research and Evaluation), 2013, 12, 81-82.
- [10]. Li Xianglong, Yang Guifu. Research on Constructing User Digital Behavior Trajectory by Using Campus Card Data[J]. Information Systems Engineering, 2015,11,120-121.
- [11]. Yan, Zhao Zeyu. Big Data Innovation and Smart Campus Service [J]. China Education Informatization, 2013, 23, 3-7.
- [12]. Chen Xianchang. Deep learning algorithm and application based on convolutional neural network [D]. Zhejiang Gongshang University, 2013.
- [13]. SPARK.[EB/OL][http://baike.baidu.com/link?url=R14_c6_CLbRvLzjcx7f3JLe7VnS118EKmLSDRbgNb2w2QlBin2sa-v8OmPl2eyn7fp_rLEpefi7hktIxgyA8CK].
- [14]. Li Cuiping, Lan Mengwei, Zou Benyou, et al. Big Data and Recommendation System [J]. Big Data, 2015(3).