ISSN: 2664-9640

Research on the construction of user dynamic portrait model based on small data

Yi Liu

School of Yunnan Normal University, Kunming 650500, China.

Abstract

User profile is one of the tools for realizing accurate service under the background of digital informationization. In order to solve the coarse granularity of user portrait, this paper proposes a user dynamic portrait model based on small data. Introduce individual-centered all-round and multi-level small data into the precise knowledge service of university libraries, and construct library user portrait models. Analyzed the concept, characteristics and composition of library small data, put forward a new idea of using big data technology to process small data, and researched the process of user portrait.This model can realize the dynamic portrait of the user according to the changes of the user's small data, which provides the possibility to analyze the user's needs in multi-dimensional vision, and can deeply dig out the needs and preferences of the users in different aspects, and provide the users with accurate knowledge services.

Keywords

Small data, user profile, precise knowledge services, university libraries.

1. Introduction

As an effective tool for delineating target users and improving decision-making efficiency, user portraits have been widely used in many fields [1]. Such as medicine, precision marketing. Ding Xingshuo et al. [2] proposed a method based on label hierarchical modeling to construct user portraits for enterprises in view of the disadvantages of traditional label modeling . Hana Maťová et al. [3] developed a corporate character scale from seven dimensions in order to measure company image and identity to construct user portraits. The granularity of the portrait is rough, and the "one-thousand-person" portrait model is still a difficult problem facing us. This article builds a more comprehensive and accurate portrait model from the level of user small data. Small data and user portraits have the same legal person input and output [4]. Therefore, the two should first combine the strengths and avoid weaknesses, and explore better user data collection and analysis methods, which can achieve the effect of 1+1>2, which is of great significance to the development of precision services. Chen Chen [5] and others explored the application value of small data in user portraits, and constructed reader user portraits from the perspective of small data. In order to meet the precise needs of users, Sun Danxia [6] etc. subdivided users based on small data of subject users, understood users from multiple angles, constructed user portraits, and tapped the potential needs of users. The introduction of small data that can fully characterize the individual characteristics of users can effectively refine the image granularity and improve the accuracy of services.

Therefore, this article intends to take university libraries as the research object, and build a dynamic user profile model that breaks through the differences in user surface behavior based on small data that is strongly related to university library users, high-density, omni-directional, and multi-level. On the basis of describing the characteristics of users' behaviors, look for motives that affect their behaviors, such as the analysis of personality and situational data, in

order to depict more scientific user portraits, and improve more accurate services for teachers and students throughout the school with predictability.

2. The connotation and characteristics of small data

2.1. The connotation of small data

In 2014, Cornell University professor Depohar Esting proposed the concept of small data for the first time. He got enlightenment from the abnormal changes in his father's daily behavior that were not medically examined. This small data that can fully describe individual characteristics can be used as a reliable medical evidence to be used in the diagnosis and treatment of patients [7]. After the concept of "small data" was put forward, it has attracted the attention of many scholars. Allen Bonde [8] (Allen Bonde) analyzes the definition of small data from two perspectives. One is to understand from the literal definition that small data is data smaller than big data, which is a public manageable and The amount used. Second, from the perspective of characteristics, it is defined as data that is usually based on a visual perspective, data is conveniently packaged, can be turned into personalized insight, and can help people use data in their daily work and life. Ma Xiaoting[9], Chen Chen[10] and others believe that small data is all-round data centered on the individual. Sun Haijing [11] believes that small library data is generated by users personally and collected through a comprehensive and personalized collection. In summary, library user small data is a collection of all data related to library users and library activities.

2.2. Small data characteristics

Small data is a collection of data perceived by individual users. It has a high data value density and obvious personal characteristics. Big data focuses on a large number of users, and discovers patterns in the same type of behavior and characteristics of different users. Small data emphasizes precision. Because the distinctive individual uniqueness of small data creates conditions for obtaining the personalized needs of users, there is usually a more rigorous identification of data sources. Small data emphasizes causality, uses results as a weather vane, pays more attention to the internal principles behind the phenomenon, and is more entangled in why. Big data often pays more attention to what the problem is, and uses correlation to arrive at a solution to the problem.

3. Literature References

As an important service target of library and information institutions, colleges and universities cannot explore the demand mode without collecting and analyzing users' all-round data. According to the nature of university libraries, the small data of library users in this article is centered on user learning activities, collected through management information systems, sensors, reading terminal equipment, wearable devices, etc. In addition, reading in third-party services Learning-related data is also a big source of small user data [12]. This article divides users into reader users (undergraduates, graduate students), research users (graduates, doctors, teachers), teaching classrooms, and subjects based on this article. The details are shown in Table 1.

Table 1: The composition of user small data in university libraries

	Data category	Data type composition	Data Sources
Reader user	Basic characteristic data	Reader's age, gender, major, grade, education level, etc.	University information database,

ISSN: 2664-9640

	Reading behavioral data	Readers 'browsing, borrowing, recommendation, downloading of electronic resources, etc.	wearable equipment, sensor equipment, information office network monitoring, library management system, scientific research management system
	Learning behavioral data	Classroom learning, online classroom learning, training courses, lectures, participating projects, etc.	
	Situational data	The time, place, method of the reading activity, etc.	
	Social activity data	Social networking sites, academic blogs, extracurricular activities, study notes, project cooperation, etc.	
	Reading psychological data	Changes in the psychological and physiological characteristics of readers in reading activities	
research user	Basic characteristic data	Age, gender, major, title, working years, research direction, etc.	
	Teaching activity data	Information on teaching subjects, training of graduate students or doctoral students, library lending, browsing and other information	
	Research activity data	The number of papers published and cited status, related information about hosting the project, access to the knowledge base, online learning and reading, etc.	
	Social activity data	Participate in academic conferences, team personnel changes, social networking sites, academic forums, exchanges of experience within the research team, etc.	
classroom	Basic characteristic data	Course outline, teaching purpose, teaching method, teaching design, etc.	Teacher monitor, video recorder, educational administration system, class evaluation system
	Generate data	The demeanor and expression of students in class,the frequency of teacher-student interaction, etc.	
	Evaluation feedback data	Student evaluation system scoring, quantity and quality of homework completed after class, peer evaluation, teacher self-evaluation, etc.	
Subject	Basic characteristic data	Subject category, major, subject background, teachers, source of students, etc.	The requirements of the state and the Ministry of Education for the development of disciplines, the school discipline development plan
	Scientific research data	Academic papers, monographs, patents, published textbooks, etc. related to the subject	
	Development planning data	Subject development requirements, school talent training plan, etc.	

ISSN: 2664-9640

4. Small data processing framework

Small data also has 5V characteristics similar to big data in some cases [13]. Through long-term tracking of individual users in multiple dimensions, a large amount (Volume) of small data can also be generated. These data sources are widely diverse (Variety). Because it is necessary to use the data of individual users to make effective perceptions and predictions of behavior, there is a high requirement for velocity (Velocity). The data of individual users is often affected by many subjective factors and there are many uncertainties. Various processing methods (such as multi-source data fusion) must be used to ensure the accuracy of the user's small data sources and processing results (Veracity). User small data also contains a lot of in-depth value (Value). It is precisely because small data and big data have some similar characteristics that the big data thinking can be used to store and process user small data, and the big data algorithm can also be used to mine and analyze user small data.

"Small data" in the library field is centered on readers and users, and dynamically analyzes the various needs of readers at different times, and conducts data collection, processing, calculation, analysis and application [14]. Provide scientific support for the library's decision-making, personalized customization and push. This paper draws on the general process of service mining and data mining, and proposes a system architecture for library user small data processing, which is divided into three levels, namely user data collection, data processing and storage, and data mining analysis. The basic research framework is shown in Figure 1. Shown.

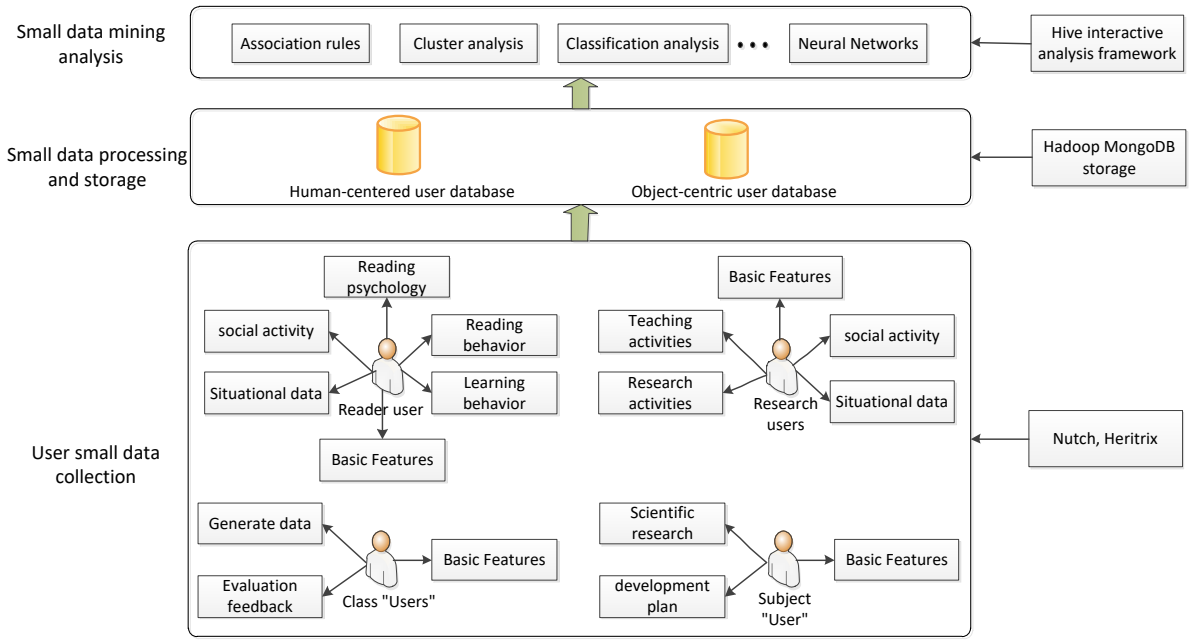

Figure 1: Small data processing framework

Small data collection is the foundation. It is divided into human-centered user data, that is, reader user data, and scientific research user data, according to the service objects of colleges and universities; object-centered user data is classroom "user" data and subject "user" data. The four types of small object data are divided into static data and dynamic data. Static small data with strong stability is composed of readers, subjects, basic characteristics data of scientific research, and subject development planning data. Dynamic small data consists of reading behavior data, learning behavior data, situational data, social relations, etc. Dynamic data is easily affected by the external environment and other related factors. The social relationship data in the dynamic data can be obtained using Nutch or Heritrix crawler technology.

ISSN: 2664-9640

Small data processing means cleaning and standardizing small data to filter out errors, low-value and invalid data and storage of small data. Use SQOOP incremental migration to extract the collected data into the Hive table, and load it into the HBase table by converting it into an HFile file. Store the cleaned data in MongoDB to form a human-centric user database and a material-centric user database. MongoDB is a document-based storage type. Data is stored in the form of documents. The small user data is stored in the system log and user operation records in the form of documents. Therefore, it is more suitable for storage with MongoDB, which can standardize the standardization of small data collection and improve the collection of data The relevance to ensure the accuracy of the reader's portrait model, so as to improve the efficiency and economy of data calculation, and reduce the loss of system resources in the portrait process.

Small data mining analysis is the core stage, and the most important thing in this process is the selection of data mining algorithms. The data mining analysis stage analyzes user behavior characteristics through data mining algorithms, including the construction of user behavior index system, user behavior analysis, and algorithm implementation. In this paper, static data is used to form short texts by splicing, and Word2vec is used to generate word vectors. For dynamic data, especially review data, it indirectly reflects the user's preference for a certain book and the needs of his research field. Therefore, this article first encodes and processes the review data, and then uses the long short-term memory network (LSTM) Learn the relevant language representations, and finally input the generated results into CNN to get more fine-grained feature vector extraction. as shown in picture 2. The obtained feature information is extracted through the TF-IDF model and finally stored in the mysql database.

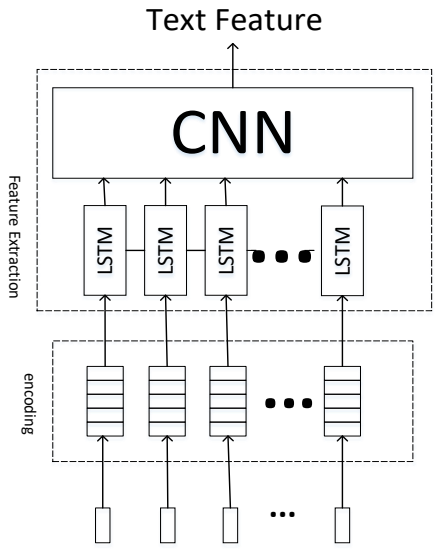

Figure 2: Feature extraction

5. Construction of user dynamic portrait model based on small data

User portraits were first proposed by Alan Copper [15] and occupied the mainstream status in early research. Alan Copper believes that user portraits are a target user model based on real data. Holden [16] regards user portraits as user-centered, restores typical user characteristics and builds user prototypes by giving specific scene descriptions, using names, photos, interests and preferences to describe users. In the user service, the library can use the small data closely related to the user to abstract the strongly relevant user tags according to the purpose of the portrait, and use computer technology to simulate the virtual portrait of the user. This article builds a user dynamic portrait model based on small data based on the processing of small data in the previous section, as shown in Figure 3. They are user-related small data collection, small

data processing, user portrait model, portrait quality evaluation, and application of user portrait.

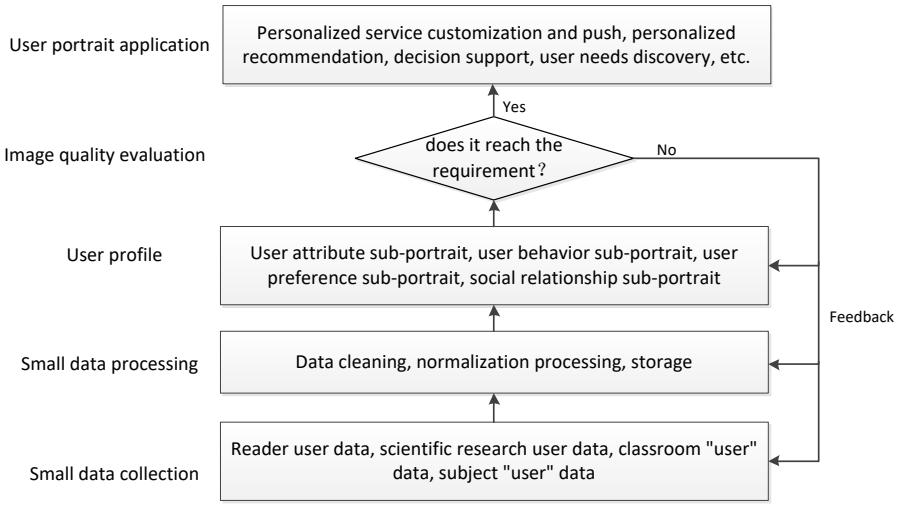

Figure 3: User portrait model

The user profile model is based on real and accurate small user data, abstracting reader tags that are strongly related to user learning activities. User portrait includes user attribute sub-portrait, behavior sub-portrait, user preference sub-portrait, and social relationship sub-portrait, which can comprehensively and accurately describe the user portrait of university user learning activities. The composition of the user portrait is shown in Figure 4.

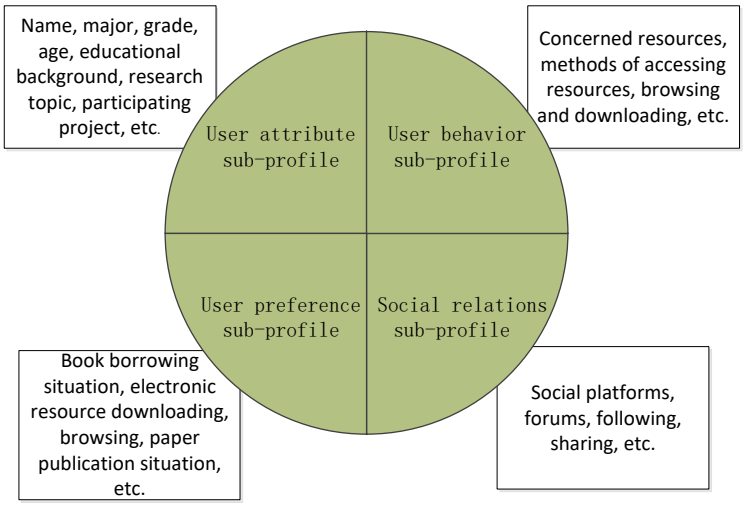

Figure 4: User portrait composition

Image quality evaluation refers to the evaluation and evaluation of the accuracy, usability, and timeliness of the image. If the user's needs are met, the image result will be sent to the library application system to provide a basis for the library's decision-making [17]. If the evaluation result is not satisfactory, the small data collection, processing, and mining analysis work should be improved to build a user portrait model that meets the real needs of users. Making good use of user feedback information can be targeted and customized services on demand.

According to the functional requirements of the library service system and the learning needs of users, the application system functions of accurate portraits can be divided into personalized recommendations for readers, personalized service customization and push, refined management of library operations and services, decision support, and readers' reading benefits The six parts, including satisfaction guarantee and reader value evaluation, involve library

construction, operation management, user service, reader service quality assurance, etc., which can effectively improve the service quality of the library.

Acknowledgements

At present, with the in-depth application of technologies such as cloud computing, big data, artificial intelligence, and the Internet of Things, in this environment, the user needs of university libraries are diversified and personalized, and require high service timeliness, professionalism, and accuracy. User portrait technology, as a user data analysis tool, can assist smart libraries in universities to tap the potential needs of readers and actively push personalized and professional precise subject services. Subject librarians can predict and analyze the potential needs of the group based on the user portrait platform, and can also analyze the different potential needs of individual readers, so as to provide users with more comprehensive and accurate subject services. Personalized services need to be adjusted at any time according to the dynamic needs of users, which requires the use of information channels to collect, classify, integrate, and mine data resources. Therefore, how to collect readers' small data comprehensively, accurately and with strong relevance to achieve accurate portraits and analysis of readers is a major issue that libraries must pay attention to to improve precision services and service quality assurance. The portrait process should also protect the user's privacy and safety to prevent leakage and abuse.

References

[1] Zhang Liman, Zhang Xiangxian, Wu Yawei, et al. Research on the construction of social academic app user dynamic portrait model based on small data_Zhang Liman[J]. Library and Information Service, 2020,64(5): 50-59 .

[2] Ding Xingshuo, Li Xiang, Xie Qian. A method for building corporate portraits based on label hierarc hical and deep modeling[J/OL]. Computer Application: 1-10[2021-09-22].http://kns.cnki.net /kc ms/detail/51.1307.TP.20210917.1124.002.html.

[3] MAŤOVÁ H, DZIAN M, TRIZNOVÁ, et al. Corporate image profile[J]. Procedia Economics and Finance, 2015, 34(03): 225-230

[4] Zhang Ruhao, Yang Zhiping. A Summary of the Frontier Research on Precision Services of Academic Libraries at Home and Abroad[J]. Journal of Library Science Research, 2019, 467(24): 2-12.

[5] Chen Chen, Ma Xiaoting. Research on Accurate Portraits of Library Users Based on Small Data_Chen Chen[J]. Information and Documentation Services, 2018, 224(5): 57-61.

[6] Sun Danxia, Wang Weijun, Jiang Yi. Research on Embedded Subject Service Based on User Small Data［J］. Library Work and Research, 2019(4): 84-90.

[7] Estrin D.. Small data, where n=me [J]. Communications of the ACM, 2014, 57(4):32-34

[8] Focus Vision.Understanding the Customer through Big Data and Small Data[EB/OL].[2019-05-07].https//www.focusvision.com/resources/understanding-thecustomer-through-big-and-small-data/.

[9] Ma Xiaoting, Chen Chen. Research on Library Personalized Service Based on Reliable Small Data[J]. Library and Information Service, 2015, 59(04): 70-75.

[10] Chen Chen. Library Personalized Service Based on Small Data Decision Support[J]. Books and Information, 2015(01):82-86.

[11] Sun Haijing. Research on the Library Personalized Reading Promotion Service Model Based on User Small Data[J]. Henan Library Journal, 2021, 41(01): 4-6.

[12] Niu Yong.Research on the Construction of Library Knowledge Community Based on Small Data[J].Books and Information,2016(2):58-61.

[13] Yang Xiaogang,Jiang Yi,Zhang Fan,Wang Weijun.User small data management based on big data technology[J].Information Theory and Practice,2018,41(03):29-33+15.

ISSN: 2664-9640

[14] Zhou Xin.Research on Subject Knowledge Service Based on "Small Data" Mining in Big Data Environment[J].University Library Work,2020,40(06):66-69+84.

[15] Alan C. The Road of Interaction Design[M]. Beijing: Publishing House of Electronics Industry, 2006:115-135.

[16] Holden R J,Kulanthaivel A,Purkayastha S,et al.Know thy eHealth user:Development of biopsychosocial personas from a study of older adults with heart failure [J].International Journal of Medical Informatics,2017,108(10):158-167.

[17] Li Yong, Meng Xiaofeng, Liu Ji. Research on the long-term evolution of online user interest based on small data[J].Computer Research and Development,2015(4):779-788.