Study on the Construction of User Portrait Model for Accurate Dissemination of Public Welfare Books in China

Xu,Zhenyun[1] Hu,Zhihua[2] Guan,Yurong[1]

1Academy of Computer, Huanggang Normal University, Huanggang, Hubei, 438000, China
2Academic Affairs Office, Huanggang Normal University, Huanggang, Hubei, 438000, China

Abstract: There are six steps to construct user portraits for public welfare book users: user data collection, user data processing, user data label extraction, user data interaction association, user data application implementation, and user data feedback. Constructing the user portrait model of public welfare books in China is theoretically and technically feasible. Still, it faces some difficulties in the process of promotion and application: it is difficult to collect data, it is difficult to achieve the accuracy of real-time user portraits, there is a lack of feedback process, and the analysis of user growth process is neglected. It is suggested that effective targeted countermeasures should be adopted to solve the problem: diversification of data collection indicators, detailed classification of user tags, maintenance of the positive image of public welfare book reading service providers, and a friendly visual interface for feedback.

Keywords: Public welfare books; Precise communication; User portraits; Model building

DOI: 10.62639/sspjinss02.20240104

The concept of user portrait was first proposed by A. Cooper, the father of interaction design, who believed that user portrait is a virtual model based on real data related to users and is the virtualization of real users. "User portrait is to condense one or a type of user label according to the user's social attributes, personal behavior, preferences, and interests." The premise of the construction of user portraits of public welfare books in China is that the data records left by users of public welfare books in the process of use, according to the user's use data, the data information is collected and classified, the valuable data is screened to describe the user, and the whole picture of the user is depicted as much as possible, to understand the user. User portrait theory and technology have gradually matured in e-commerce and have been used by major business platforms.

Many scholars have found that user portraits have a positive significance for maintaining and managing library readers. Currently, the research on user portraits in the field of library information mainly focuses on the construction of reader portraits based on the data generated in the process of library user service and user behavior, intending to improve the accuracy and personalization of reader service and resource utilization. Most of these studies are practical and applied research. Most of the researchers are library staff, and "there are few and insufficient research on the construction methods and multi-dimensional ontologies of user portraits, although the subdivision topics are diverse, but the application subject is single." The construction and ontology research of user portraits involves subdivided topics such as data mining and intelligent

(Manuscript NO.: JINSS-24-4-F001)

Funding

2021 Philosophy and Social Science Research Project of Colleges and Universities of Hubei Provincial Department of Education, Project No is 21D114, Project Category is Key Project, Name of Person in Charge is Xu Zhenyun, Project Name is Research on the Precise Dissemination of Public Welfare Books in China Based on User Portraits.

In 2022, the general project of the Hubei Provincial Social Science Fund (later funded project) was approved; the project name is Research on On-demand Publishing and Sharing of Innovative Services for Public Welfare Books in Hubei Province, project number is HBSK2022YB403, project leader is Xu Zhenyun.

technology, which are closely combined with computers, data modeling, and other technologies and intersect with computer disciplines. Most current research is conducted on specific service architectures, which are difficult to generalize. Therefore, one of the important tasks is to build a unified framework for user portraits and realize the integration and sharing of user data. The current user portrait research and application are mainly based on offline data structure, which is a kind of static portrait, and the user's long-term behavior data changes, and the changed behavior data is challenging to track and correct in time, resulting in a significant deviation between static user portrait and dynamic user behavior. The next thing that needs to be studied and further expanded is the construction of real-time portraits of public welfare book users.

1. The Process of Building a User Portrait Model for the Accurate Dissemination of Public Welfare Books in China

There are six steps to construct user portraits for public welfare book users: user data collection, user data processing, user data label extraction, user data interaction association, user data application implementation, and user data feedback. Public interest book reading service providers collect users' basic attributes and behavioral data through information equipment, platforms, and management systems. Once the data is collected, it needs to be preprocessed. Analyze users based on user data, extract user data labels, and construct user portrait models. Then, the user portrait model is implemented into the application service. Finally, feedback detection was carried out using the user portrait model.

(1) User data collection

In the era of big data, the records left by users when they use the network can be traced back to their needs by collecting and analyzing user data. The information collected by user portraits can be roughly divided into four categories according to attributes: basic demographic attributes, social attributes, lifestyle, and psychological attributes. For users of public welfare books, according to the above classification, it can be summarized into four categories: user basic information data, user behavior data, situational data, and user social data. The basic information is user registration data, such as name, gender, age, education, address, work unit, income, family status, etc. User behavior data is mainly user historical behavior data collected through the network, and online behavior mainly includes login, retrieval, browsing, downloading, and other usage records. Offline behaviors mainly include records of borrowing and returning books, participation in book activities, etc. Contextual data includes time context, location context, resource contextual data, service contextual data, mobile contextual data, terminal contextual data, etc., obtained through intelligent chips and sensing devices. The context dimension reflects the user's current situation, which refers to the impact of the surrounding environment on the user, including not only the physical context (time, location, weather, etc.) but also the resource context (resource distribution, resource utilization, etc.), the social context (cooperative relationship, social relationship, etc.), the system context (network status, security, etc.), and the terminal context (device type, interaction mode, etc.). Contextual data is usually obtained through GPS, sensors, etc., and contextual awareness information is interconnected with basic user information, which can more accurately reflect users' specific preferences and needs at specific times and places. User social data includes discussions, exchanges, cooperation, and concerns between users, which are obtained from social platforms and social software, and social attributes include public welfare book reading service providers' WeChat official accounts, network access log data, user service information group data, offline activity feedback data, Weibo data, etc. The social dimension is a graph reflecting user preferences and online community relationships. The mobile library constructs many online communities with theme spaces containing relatively comprehensive, rich, real-time user data. In the online community, users are information users, creators, and disseminators of information content. Through interactive behaviors such as forwarding, sharing, and commenting, users

form a close community relationship and reflect users' interests and preferences. The user's social dimension data can be obtained through the user's homepage, and the social dimension information between users can be aggregated, which can effectively classify users, find users with similar interests, and build group user portraits, which can not only facilitate communication between users with similar interests but also make practical information recommendations. Based on the above user data, this paper constructs user portraits. However, the portrait will change according to the user's own and external conditions.

(2) User data preprocessing

The so-called user data preprocessing mainly refers to cleaning, transformation, specification, and integration, and the main task is to judge and review the accuracy and applicability of the data, screen the data that does not meet the requirements, and the data with apparent errors, and clear them, and finally form regular and structured data. The massive user data collected by public welfare book reading service providers from user information systems and smart management terminals is repetitive and disorganized, and this massive data with low-value density should be reprocessed to extract useful information. The big data mining algorithm can process user data to achieve twice the result with half the effort. The association rules, neural networks, support vector machines, K-means clustering algorithms, etc., can be used to analyze the rules and patterns of user behavior big data, which helps discover user behavior characteristics, read interest preferences, activity and other information, and lay the foundation for the label extraction of user portraits.

(3) User data label extraction

In constructing user portraits, tags are the concise introduction and differentiation of public welfare book users, and the most frequently used words are often the most concise and easy to understand. After the label is set, the user portrait is convenient for the subsequent assignment and calculation of the data model. Through collecting, collating, and refining user information data and behavioral data of public welfare books, user tags can be obtained, and corresponding labels can be attached to users so that public welfare book users' preferences, needs, and interests can be visualized through tags. The larger the weight value of each tag, the more important the tag is, which can more intuitively show users' preferences, interests, and needs of public welfare books.

The individual portraits of public interest book users can be determined according to five index systems: user basic information labels, user behavior labels, user activity labels, user preference feature labels, and user satisfaction labels. The user basic information tag presents the basic characteristics of the object by collecting the user's basic information. The user behavior tag describes the characteristics of the user's access to information resources at the present stage by obtaining the information left by the user in the behavior of using the public welfare book reading service system. User activity tags mainly refer to the time path and spatial path of user behavior, which describe the time span and length characteristics of user acquisition of information resources. User preference tags predictively describe the potential needs of users through mining user data. User preference data is divided into explicit preference data and implicit preference data. The former refers to the behavior data users affirm, such as targeted target retrieval, likes, favorites, subscriptions, etc.; The latter refers to the user's potential interest and preference data obtained by combining user operation behavior and recommendation algorithms, such as browsing time, frequency, attention, etc. The user preference data of public welfare books mainly comes from two aspects: one is the data of users' collections, likes, and frequent browsing, that is, explicit preference data; Second, based on the mining and analysis of the information of the user's contextual dimension and social dimension, the user's preference needs in a specific environment are clarified, which is also used as the data of the user's preference dimension. User satisfaction is mainly from the users' perspective, and the comprehensiveness and accuracy of the public welfare book reading service model are comprehensively evaluated. The user satisfaction label can provide a scientific

basis for the decision-making of the precise dissemination of public welfare books, allow users to actively participate in the activities of accurate dissemination of public welfare books, maximize the value of users, and form a closed loop of user portraits.

(4) User data interaction and association

In constructing user data interaction association, it is necessary to use data mining algorithms to dig deep into the potential connections between users, statistically analyze the similarity of multiple users, gather users with similar characteristics, and generate several user groups or clusters. The methods of constructing group user portraits generally include feedback algorithms, multi-dimensional analysis, clustering algorithms, filtering algorithms, weighting algorithms, etc. Collaborative filtering algorithms are divided into two types: collaborative filtering algorithms based on user similarity and collaborative filtering algorithms based on item similarity. On the one hand, it is based on the user's collaborative filtering algorithm USERCF; that is, the information is pushed by imitating the user's choice, which is pushed according to the user's similarity. Simply put, it is to recommend to the user the reading information selected by other users who have similar preferences, that is, to start with the similarity of the users. On the other hand, it is a collaborative filtering algorithm based on reading information, ITEMCF. Push the user through similar reading information, according to the user's recommendation of information similar to the information he chooses or follows, starting from the similarity of information. Tapping the connection between users is the key to analyzing user needs and realizing the accuracy of public welfare book reading services.

Through the collection of users' information browsing and query data, and after screening and removing dirty data, select the data that meets the requirements of public welfare book users, formulate user labels, and establish a user data model. The user data model includes seven dimensions 5W2H, and it involves analyzing the user's login information and their online data to determine who the user is (who), corresponding to user tags; at what time (when), corresponding to user activity tags; in what location (where), corresponding to user context tags; what actions they have taken (what), corresponding to user behavior tags; which one they are more accustomed to (which), corresponding to social platform tags; how to achieve precise communication (how to), how to utilize feedback mechanisms (how many), corresponding to user satisfaction tags. These are simple and operable, and it is feasible to construct a public welfare book user data model through the above seven dimensions.

(5) User data application implementation

The entire public interest book user service framework can include the following modules.

First, the user feature recognition module. Firstly, the service users' fine-grained feature recognition is carried out. The characteristics of users' identity, knowledge structure, and current tasks are identified from the natural, behavioral, situational, and social dimensions, and the feature analysis is carried out after identification. The feature variables that are conducive to optimizing the consulting results are summarized. Second, the user portrait matching module. The similarity between the served users and the existing user portrait database is matched based on user characteristics. The approximate weights are obtained by calculating the basic and exact matching weights between user attribute labels and user portraits, and the likelihood of the user belonging to the portrait is measured. The approximate weight calculation results are optimized through the user attribute aggregation algorithm, and then the user is matched with the portrait to which he belongs. Third is the smart service module. The application of technologies such as cloud computing, the Internet of Things, and big data enables public welfare book reading service providers to provide users with all-round and integrated resources and services without being limited by time and space, realizing a high degree of integration of resources and services, and making their service objects and service content more extensive. The smart service realizes the interactive integration of online and offline services to accurately

disseminate public welfare books, mainly including intelligent retrieval services, intelligent Q&A services, and intelligent recommendation services. The objects of personalized recommendation services can be divided into single users and group users. The similarity of users is calculated through user tags. Then, similar users are clustered to obtain the characteristics of the clustered groups. Then, according to the characteristics of different groups, the fine division of service content is realized to improve the efficiency of recommendation services. Through repeated training of many user data and optimization of feature expression, group features can be further mined, and relationship graphs can be established.

On the one hand, it provides accurate information dissemination and push and accurately provides the required information according to the user's behavioral preferences to achieve accurate services. On the other hand, based on grasping the needs of specific user groups, select the most appropriate promotion mode and design the promotion content that matches users' needs to improve the pertinence of accurate communication. Based on conforming to the behavior and habits of the target user group, various promotion forms are adopted to attract user attention and mobilize user enthusiasm.

(6) User data feedback

After users receive the push information, the degree of acceptance and satisfaction are essential to evaluate the effectiveness of accurate communication. Suppose undifferentiated information is pushed to different users of public welfare books; the user's attention is not high, and it will also produce a sense of boredom, which can easily lead to a confrontation between the user and the service organization. Therefore, it is imperative to disseminate reading service information accurately. User feedback is a key part of the user portrait process, from one-way to two-way transmission, and this link function is the most easily overlooked.

The construction of the user portrait model of public welfare books in China is theoretically and technically feasible, but it faces some difficulties in the process of promotion and application.

2. The Dilemma Faced by the Construction of User Portrait Model for Public Welfare Books in China

(1) Difficulty in data collection

Since the data of public interest book users is implicit or perceptual, it is challenging to collect them. A lot of data involves users' privacy, and users are unwilling to share it. To accurately grasp users' reading preferences, it is necessary to collect personal perceptual data about users of public interest books. The user may directly reject the collection of perceptual data, or the data obtained may be the data corrected by the user, which is distorted and has no application value. The accuracy of basic user information is the premise of building user portraits. If the data accuracy is not high, it will be more challenging to carry out other work effectively.

(2) It is difficult to achieve the accuracy of real-time portraits of users

The accuracy of a persona is highly dependent on the accuracy of the user data itself. If it is difficult to obtain the user's own data, it can easily cause the user's portrait to be distorted. In addition, user data is updated and iterated rapidly. Readers' preferences and needs are a rapidly changing process, and the analysis and labeling process of user portraits is often challenging to keep up with the change speed of data and user needs, resulting in the deviation of user portraits from accurate user personas. Finally, the classification standards for users are not uniform. The complexity and diversity of needs also lead to the diversification of users' preferences, and user portraits are often based on the preferences and needs of recent times, and the real-time, accurate, and primary reading needs of users are not necessarily consistent with user tags.

(3) Lack of feedback process

The public welfare book information push service based on user portrait technology is a one-way support system, and most public welfare book service providers provide a one-way delivery process from data acquisition to user information push. The evaluation and feedback function of user portrait technology by users and public interest book service providers has not been paid enough attention, and the feedback information collection channel has not been fully established. User feedback is often more difficult; the collection and sorting of user feedback depends on the consciousness of the service organization and is not a mandatory indicator of work performance appraisal. Therefore, the system feedback is the least valued, and the application is the lowest. Still, the feedback has a huge counter-effect on improving the accuracy of the user portrait.

(4) Ignore the analysis of the user growth process

After receiving the public welfare book reading service, users' growth process differs from before, and the user portrait is passive and weak in excavating user growth rules. First, the utilization rate of user historical data tags is low. The user's historical tag data may be recorded. Still, the probability of the historical data being used for association, classification, and clustering operations is small, resulting in the user tag being only in the current state, and the user's growth experience cannot be genuinely restored and visualized. The important tag of growth characteristics cannot provide assistance services for accurate communication. Second, the analysis function of user learning progress is weak, and there is a lack of analysis of user learning progress. As mentioned in the previous point, the user's historical portrait has not been analyzed and processed in detail. Its change laws and patterns cannot be thoroughly analyzed and utilized, so the progress law and key nodes in the user's learning progress cannot be accurately grasped, and the reasons for the differences between individuals and the core of improving the learning progress cannot be accurately traced. Third, there is a lack of analysis of user goals and attainment. In reality, using usage indicators to evaluate service satisfaction is unscientific and cannot reflect the achievement of user needs.

Given the above dilemma, effective targeted countermeasures are recommended to be adopted to solve it.

3. Accurate Dissemination of Public Welfare Books, the Construction of User Portrait Model and Benign Development Countermeasures in China

(1) Diversification of data collection indicators

Most of the data sources of readers' reading need to rely on the active provision of readers, and if readers do not cooperate, the data collection will be passive. To grasp the initiative of data collection, public interest book reading service providers need to discover more alternative indicators that reflect users' reading needs and dig deeper into alternative sources of alternative indicator collection.

(2) Classification of user tags

User tags that are too sloppy do not accurately meet the real needs of users and may also cause negative emotions in users. Public interest book reading service providers need to make user labels detailed, scientific, and standardized, which is conducive to users finding book information and accurately pushing book information to users. Sequence mining, regression analysis, and other algorithms are used to model the user's historical portrait to ensure that the user's growth law can be dynamically presented. In the stage of personalized service, the identification information of the user tag should be used to analyze the personalized service content to ensure that the individual interests most conducive to the development of the user are effectively served.

(3) Maintain the positive image of public interest book reading service providers

Public interest book reading service providers are non-profit organizations or individuals, and the user data obtained can only be used to provide users with better services. They cannot be used for commercial purposes. Public welfare book disseminators need to take the user satisfaction label as one of the main criteria for service performance evaluation, evaluate the accuracy and utilization rate of their own services through user feedback, motivate practitioners to check and fill in gaps, improve business capabilities, improve service efficiency, and eliminate the "information gap" between public welfare book disseminators and recipients.

(4) Do a good job of a friendly visual interface for feedback

In the user-oriented interface, the feedback function can effectively portray the user, which is very beneficial in pushing the book resources promptly to meet readers' real-time reading needs. The active participation of users can not only make their own portraits more vivid and accurate. It also opens up the last link in constructing user portraits to accurately disseminate public welfare books and form a benign closed loop of user-centered precision services.

4. Conclusion

The rapid development of technologies such as the Internet, big data, the Internet of Things, blockchain, and artificial intelligence has made traffic an important factor in the rise of public welfare book service providers. In the era of information explosion, the fiercest competition among various information service providers is due to users' attention. To increase user traffic, public welfare book service providers must provide users with accurate services, establish user portraits, and refine preferences and needs by mining users' data traces in the reading process to provide users with optimal reading information and improve the user service experience.

References

[1] Wang Zhengyou, Zhang Haidi. Accurate video recommendation based on user portrait in the era of big data[J]. E-commerce,2019(10):62-65.

[2] Wang Xunhong,Ma Jianchun. Research Status and Theme Analysis of User Portrait in the Field of Graphics and Information in China [J]. Library Journal, 2022, 44 (04): 104-111. DOI:10.14037/j.cnki.tsgxk.2022.04.018.

[3]Zeng Hong. Big data user portrait and precision marketing based on microblog[J].Modern Economic Information,2016(16):306-308.

[4] Ceng Xiufeng, Zhou Weiting, Zhang Xiaolong, et al. Research on Library Smart Reference Service Model Based on User Portrait [J]. Journal of Library Science Research, 2021, (02): 86-93+101. DOI:10.15941/j.cnki.issn1001-0424.2021.02.012.

[5] Cheng Quan. Research on digital library information service mode based on user portrait[J].Library Journal,2018(4):68-71.

[6] Zhu Jing, Xu Yiran, Wu Ying, et al. Research on User Portrait of University Library Based on Big Data [J]. Science & Technology Wind, 2021, (31): 74-77. DOI:10.19392/j.cnki.1671-7341.202131025.

[7] Guo Qiang, Liu Jianguo. Modeling and analysis of user behavior in online social network[M].Beijing: Science Press,2017:170-172

[8] Zhao Ping." Research on the promotion path of smart service in university libraries under the background of "double first-class" construction[J].Journal of University Library and Information Science,2019(4):44-47.

[9] Hu Chaojun. Research on personalized intelligent service mode of public library based on knowledge mining[D].Tianjin: Tianjin University of Technology,2019:33-35.

[10]Wang Li,Tan Kaibo,Huang Yun. Problems and Countermeasures of User Portrait in the Application of Library Personalized Service[J]. Digital Library Forum, 2021(10):66-72.