Data-driven Digital Library User Persona Model Construction Postprint

Authors: Xu Pengcheng, Bi Qiang, Zhang Han, Mou Dongmei

Date: 2023-07-26T00:00:00+00:00

Abstract

[Purpose/Significance] To mine the hidden value behind user data, comprehensively understand user needs, construct user profile models, and provide new impetus for digital libraries to achieve precision services. [Method/Process] This study analyzes the connotation and characteristics of user profiles in digital libraries, examines the data sources and collection/processing procedures of user profiles, proposes a main driven path for user profiles under a data-driven approach: datafication → tagging → association → visualization, and constructs a multi-dimensional, multi-level, and three-dimensional user profile model from the dimensions of nature, interest, and social interaction. [Results/Conclusion] The construction process of the digital library user profile model is elaborated in detail, the framework model of user profiles is designed, and user profiles are applied to precision recommendation, personalized retrieval, precision promotion, and reference decision-making in digital libraries, thereby promoting the upgrading of knowledge services in digital libraries.

Full Text

Preamble

Volume 63, Issue 3, February 2019
ChinaXiv Cooperative Journal

Construction of Digital Library User Profile Model Driven by Data

Xu Pengcheng[1], Bi Qiang[1], Zhang Han[1], Mu Dongmei[2]
[1] School of Management, Jilin University, Changchun 130022
[2] School of Public Health, Jilin University, Changchun 130021

Abstract

[Purpose/Significance] To mine the hidden value behind user data, comprehensively understand user needs, construct a user profile model, and

provide new momentum for digital libraries to achieve precision services. [**Method/Process**] This paper analyzes the connotation and characteristics of digital library user profiles, examines the data sources and collection processes for user profiling, and proposes a data-driven main route of datafication → labeling → association → visualization for user profile construction. From the natural dimension, interest dimension, and social dimension, it constructs a multi-dimensional, multi-level, and three-dimensional user profile model. [**Result/Conclusion**] The paper elaborates on the construction process of the digital library user profile model, designs a framework model for user profiles, and applies user profile data to precision recommendation, personalized retrieval, precision publicity, and reference decision-making in digital libraries to promote the upgrading of knowledge services.

Keywords: data-driven, digital library, user profile
Classification Number: G251
DOI: 10.13266/j.issn.0252-3116.2019.03.004

Scholars both domestically and internationally have used new-generation information technologies to directly characterize and describe the current era as the "Big Data Era," "Intelligent Era," and "Algorithm Era." The knowledge creation and utilization environment that libraries depend on is transitioning from the information age to the "data age." The big data-driven wave of "datafication" has enabled libraries to possess multi-form resources and diversified data, which together constitute a big data environment for digital library services that can be fully integrated and associated [**?**]. Faced with massive digital resources, users cannot quickly and accurately find resources of interest, leaving them trapped in the dilemma of knowledge disorientation. In view of this, digital libraries urgently need to transform from information services to knowledge services, and constructing dynamic user profiles is an effective approach to deeply understand user needs, gain real-time insights into user preferences, achieve service transformation, and resolve the asymmetry between users' precise needs and the extensive service model of digital libraries.

1. Overview of Related Research

1.1 Overview of User Profile Research

The concept of user profile (persona) was first proposed by Alan Cooper, the father of interaction design, who pointed out that a user profile is a user goal model based on a series of real data, representing a virtualization of real users [**?**]. User profiles were first applied in interaction design or product design fields, referring to the delineation of real characteristics of specific service groups. They are an effective tool for characterizing target users and connecting user demands [**?**]. Early user profile models were generally constructed from a fictional perspective, where designers concretized their subjective imagination into target customer profiles to design product prototypes. However, user profiles built from a fictional perspective depend entirely on designers' subjective assumptions and can

easily be misleading [?]. With the development of Internet technology and the advent of the big data era, the connotation and extension of user profiles have changed under data-driven approaches, mainly by depicting user characteristics through data to provide quality services. Therefore, user profiles have another expression: user profile, which is adopted in this paper. Meanwhile, the perspective of constructing user profiles has shifted from fictional perspectives to goal-oriented, role-based, and participatory perspectives, with greater emphasis on collecting user data to support profiling results. C. Teixeira et al. believe that user profiles are models that independently describe user needs, preferences, and interests by extracting personal information sets from massive data pools [?].

The construction of user profiles cannot be separated from algorithmic and technical support. As L. Zeng et al. stated, many studies have used big data technology and data mining algorithms to build user profile models, such as vector space models, topic models, and neural network models [?]. A. Sandra et al. [?] proposed a method for constructing user profiles based on contextual preference rules to more accurately reflect user preference characteristics. A. Farida et al. [?] proposed a method for constructing user profiles from user interactions with search systems using dynamic Bayesian networks. J. Yu et al. [?] proposed a method for constructing and updating user profiles based on semantics and browsing sequences, introducing a memory model from cognitive psychology to ensure the dynamic nature of user profiles. Thus, user profiles have integrated and utilized many mature algorithms and technologies, but their essence is actually modeling through multi-dimensional label combinations [?]. Constructing user profiles by extracting user labels from multi-source user data is the core of user profile model construction.

With the deepening of big data and "Internet Plus" concepts, enterprises increasingly focus on how to use big data to achieve precision services. User profiles can help product managers accurately understand and predict user needs, thereby precisely targeting customer groups for promotion and personalized recommendations, ultimately achieving precision services. Therefore, user profiles have been widely applied in various fields, especially in e-commerce [?], where they serve as the basis for precision marketing and decision-making. Additionally, in tourism [?], finance [?], news [?], social networks [?], and health [?] domains, user profiles also play important roles and have even become breakthrough points for achieving precision services and precision marketing in many fields.

1.2 Research Status of User Profiles in the Library Domain

Research on user profiles in the library domain has become a hot topic. The earliest application of user profiles in foreign libraries appeared in 1985 when the British National Bibliography and Blaise-line investigated users' usage of their services through telephone interviews and personal interviews, forming relevant analyses to optimize services [?]. G. Amato et al. [?] elaborated on the application of user profiles in the digital library domain, arguing that the ultimate goal of information providers is to meet user information needs, and

that personalized services depend on user profiles representing user information needs; therefore, digital libraries need to construct user profile models. G. Amato et al. [?] also studied the structure and mechanism of user profiles in digital library personalized services, detailing key issues such as user profile representation, construction, and updating. With the deepening of user profile research, many foreign scholars have mainly applied user profiles to library information recommendation [?], information filtering [?], and service design [?].

Research on user profiles in the domestic library domain started relatively late, only gradually becoming a research hotspot in the past three years, with relatively few research results mainly concentrated on model construction and practical application. Liu Su [?] and Yang Fan [?] elaborated on user profile construction ideas from specific practical cases and proposed concrete user profile analysis methods. He Juan [?] constructed a library intelligent recommendation system based on user profiles to improve reading promotion efficiency and achieve precision recommendation services. Yin Xiangquan and Xue Huanxue et al. explored university library user profile models and conducted empirical studies, analyzing university students' usage behaviors [?] and optimizing university library subject services [?]. Some scholars have also applied user profiles to knowledge communities [?] and Q&A communities [?] in digital libraries, constructing user profiles from micro-level user data. Fan Xiaoyu et al. [?] fused multi-source data to extract researcher information labels from three dimensions—basic attributes, research preferences, and research relationships—and displayed researcher profiles visually. These studies have made the application scope of user profiles more detailed and in-depth, with considerable practical value.

In summary, research on user profiles in foreign library domains started early and is relatively mature, covering theoretical foundations, model construction, methods and technologies, and practical applications. In contrast, research in domestic library domains needs to be enriched and improved. Therefore, we can draw on mature foreign user profile models and combine them with the development status of Chinese libraries and user group characteristics to construct a comprehensive library user profile model, strengthen the practice and application of user profiles, and better grasp user needs to truly achieve precision services. Additionally, in library user profile research, there is relatively little research on user privacy. After all, constructing user profiles requires large amounts of user data, making privacy issues unignorable. Future research should incorporate user privacy as an important consideration factor.

2. Connotation and Characteristics of Digital Library User Profiles

2.1 Connotation of Digital Library User Profiles

A digital library user profile is a target user model constructed by libraries to deeply understand user characteristics, predict users' real needs, and stim-

ulate users' potential needs. Based on a series of real data, it describes user characteristics, needs, and preferences, representing a virtual depiction of users' information appearance, with the goal of achieving precision services. Real data refers to data that can reflect user characteristics from multiple angles, such as user background, interests, habits, and behaviors [**?**]. Digital libraries should adopt a participatory perspective and use the data-user label mapping method to construct user profiles. The data-user label mapping method is a typical data-driven user profile. The process of digital library user profiling can be represented as datafication → labeling → association → visualization. First, collect user-related data and preprocess it to achieve datafication. Based on users' basic attributes and behavioral data, label user profiles and establish a user label system. Based on user interaction data, establish connections between users and build user relationship graphs to achieve association. Finally, visualize user profiles through visualization tools and apply user profile data to personalized retrieval, precision recommendation, user clustering and precision publicity, and library reference decision-making.

2.2 Characteristics of Digital Library User Profiles

D. Travis believes that user profiles should have seven characteristics: primary research, empathy, realism, singularity, objectives, number, and applicability [**?**]. These seven characteristics were proposed based on the early connotation of user profiles. In the context of the big data era, user profiles should have more contemporary features. The author believes that data-driven digital library user profiles should have iterability, timeliness, differentiation, interactivity, knowledgeability, and clusterability.

2.2.1 Iterability User profiles are the result of labeling, associating, and visualizing user-related data. User data consists of static data and dynamic data. Digital library user static information (such as name, age, reader ID, etc.) mainly involves basic attribute characteristics and is relatively stable. Dynamic information (such as clicks, reading, downloads, etc.) is generated through user interaction with digital libraries and continuously accumulates over time. Each interaction between users and the system may change the user profile results, indicating that user profiles are dynamically changing. Digital library user profiles should have iterability, being able to update user profiles according to changes in user needs and behaviors and adjust service methods accordingly.

2.2.2 Timeliness Affected by numerous factors such as learning plans, cognitive deepening, task adjustments, environmental changes, information overload, and the passage of time, the phenomenon of user interest drift occurs. Even for the same content, users' interest levels may gradually increase or decrease or even disappear. The phenomenon of user interest drift determines that user profiles must have timeliness. Digital library user profiles are three-dimensional depictions of target users within a certain time period, and their accuracy is valid

within the corresponding time period. Therefore, an accurate and dynamic user profile model should be able to track user interest drift in real-time and accurately, respond promptly to changes in user interests, and update user profile results in a timely manner. If user profile results are delayed, their accuracy may be reduced.

2.2.3 Differentiation User profiles are user models with distinctive features, representing a virtual manifestation of real users' characteristics in certain dimensions after datafication and reorganization. The user characteristics depicted in user profiles are not average characteristics of all users but have differentiation and targeted objectivity. The depicted user characteristics generally do not overlap or intersect. All characteristics are unique members of the user profile, each with its own specialization and focus, yet complementary and coexisting, forming the user profile as a collection of different characteristic dimensions. The differentiation of user profiles enables digital libraries to accurately identify behavioral preferences and motivations of different users, pointing the way for further service improvement.

2.2.4 Interactivity Data from the interaction between digital libraries and users is an important component of user profiles. User interaction behavior is mainly reflected in users using digital libraries for different purposes and motivations, resulting in interest-oriented interaction, problem-oriented interaction, social interaction, and benefit-oriented interaction. Through various mobile media, users engage in interactive processes with library systems, reference librarians, and other users at the operational or semantic level with transmission and feedback. In the continuous interaction process between users and digital libraries, massive amounts of user data are continuously generated. Rich user data is the foundation for the feasibility and accuracy of user profiles. User profiles are the result of mining large amounts of user data and visualizing users' information appearance, which may differ from users' real appearance. Therefore, digital libraries need to establish a complete user feedback mechanism during the construction of user profile systems, allowing users to provide feedback on user profile results, which the library can then organize to improve and refine the user profile results.

2.2.5 Knowledgeability Users are both consumers and creators of data resources. As a resource, data has underlying knowledge waiting to be mined, discovered, and utilized. The purpose of digital library user profiles is to mine user data and create knowledge to meet users' fragmented, refined, and personalized knowledge needs. Therefore, digital library user profiles should have powerful application functions to assist in digital libraries' knowledge discovery services. Digital library user profiles are built on the basis of resource profiles. Under data-driven approaches, digital libraries conduct fine-grained segmentation of resources to achieve data fragmentation and form visualized resource profiles through semantic association. Digital libraries focus on user participation and

interaction from the user perspective, associating user needs and behavioral data with resources, making user profiles more applicable and knowledgeable. Additionally, some users in digital libraries are also knowledge creators and experts in various fields. Digital library user profiles should have the capability to establish personal knowledge bases for expert users and academic users and build connections between users to promote knowledge dissemination and sharing, thereby achieving knowledge re-creation and leveraging knowledge value.

2.2.6 Clusterability Although users have differentiation, there are also commonalities between them. It is precisely because of these commonalities that homogeneous groups can be clustered and heterogeneous objects can be distinguished. User-related data hides users' common characteristics behind it. Based on the analysis and mining of user-related data, digital library user profiles can classify and cluster users according to criteria such as usage habits, interest preferences, activity levels, participation degrees, and influence, forming different user groups. This enables hierarchical management of users, making user management more scientific and efficient, and providing personalized services and precision publicity for different user groups.

3. Construction of Digital Library User Profile Model

3.1 Data Collection and Processing

User profiles aim to comprehensively and three-dimensionally depict users, requiring extensive library user data as the foundation. Multi-source user data is the guarantee of user profile accuracy [**?**]. T. P. Guimaraes [**?**] summarized the data sources for user profiles as: users' basic literacy, educational levels, social relationships, work status, location, and time information. T. Lafouge, J. P. Lardy, and N. B. Abdallah [**?**] believed that user characteristic information in information retrieval systems mainly includes two aspects: first, stable factors related to users, such as personal information and behavioral habits; second, variable information in the retrieval environment, such as search goals. Focusing on these two types of characteristic information, data sources for constructing user profile models include five categories: user status, user background, user goals, user cognition of related fields, and user familiarity with the system.

The author believes that digital library user profile data mainly includes user demographic data, user behavioral data, user social data, and other user data. User demographic data mainly includes reader ID, name, gender, age, education, major, occupation, title, and email address. User behavioral data mainly includes login frequency, browsing duration, page scrolling, clicks, jumps, downloads, favorites, copying, gestures on mobile devices (swiping, dragging, zooming) [**?**], borrowing, searching, and consulting. User social data includes likes, shares, comments, discussions, interactions, follows, citations, and collaborations. Other user data includes mobile phone and computer models, operating systems, client versions, network types, and academic achievements.

Digital library user data usually consists of structured data, semi-structured data, and unstructured data. Structured data is relatively easy to collect and convenient for forming user labels, such as basic attribute data that can be collected through user registration information. Semi-structured and unstructured data, however, are vast in volume and constitute the main data for user profiles. For example, user behavioral data such as page browsing, clicks, and downloads are mainly stored in user Web logs and need to be extracted through web crawlers and log mining techniques. Additionally, digital library systems need to embed user page behavior monitoring plugins to collect user behavioral data without affecting normal user usage. User social data is collected through interactions and collaborations between users, establishing connections between user-user and user-expert. User-related data collection requires tools such as web crawlers and Web log mining, and the collected data needs to be preprocessed through cleaning, transformation, reduction, and integration to form effective user profile data.

3.2 Dimensions of Digital Library User Profiles

Zhang Huimin and Xin Xiangyang [**?**] proposed four dimensions for constructing user profiles from an interaction design perspective: natural conditions dimension, value orientation dimension, behavioral habit dimension, and cognitive characteristic dimension. Each dimension contains multiple subdimensions, providing good generality and comprehensiveness. Chen Zhiming and Hu Zhenyun [**?**] constructed a UGC website user profile model with four dimensions: basic attributes, social attributes, interest attributes, and ability attributes. Wang Lingxiao, Shen Zhuo, and Li Yan [**?**] constructed a social Q&A community user profile from four aspects: user qualifications, user participation, answer quality, and user development trends. Hu Yuan and Mao Ning [**?**] established a digital library knowledge community user profile model from three label dimensions: basic reader information, user interests, and user activity. These studies provide valuable guidance and reference for this research.

Based on the attributes and needs of digital library users, the author believes that digital library user profiles should be depicted from three aspects: natural dimension, interest dimension, and social dimension, constructing a three-dimensional multi-level label system. The relationships among the three are: the natural dimension is the foundation of the interest and social dimensions, users' basic attributes are the basis for the formation and change of user interests and the development of social relationships; the interest dimension is the demand of user profiles, reflecting users' personal preferences and needs; the social dimension is the result of the joint action of users' natural and interest dimensions, while also reacting to the interest dimension and influencing users' interest preferences.

3.2.1 Natural Dimension The natural dimension aims to provide the most basic understanding and depiction of users, mainly based on user demographic data. Digital libraries extract basic user data from registration information, supplemented and cross-validated through website guidance, surveys, and third-party providers. Since natural dimension data can easily involve user privacy, the natural dimension labels should be users' most basic attribute data and data affecting users' use of digital libraries. The author believes the natural dimension label system should include name, gender, age, education, major, occupation, title, and city. Natural dimension labels are relatively static, and if changes occur, users should be allowed to modify them themselves.

3.2.2 Interest Dimension The interest dimension is the core dimension of user profiles, aiming to reflect user needs and interests. The principle of the label mapping method is to characterize user interests by labeling users. The construction of interest dimension labels is mainly based on user behavioral data. Users generate massive service logs through interaction with digital libraries, which truly reflect user needs and interest preferences. The establishment of the interest dimension should first analyze text content to construct resource profiles, using word segmentation and stop-word removal to obtain feature words, and utilizing TF-IDF to calculate feature word weights. Then, the LDA-Gibbs model is used to model the text collection, mapping the text corpus to various topic spaces using the statistical characteristics of the text, mining the relationships between different topics and words hidden in the text, obtaining document-topic probability distributions and topic-feature word probability distributions, and constructing resource profile labels, as shown in Figure 1 [Figure 1: see original paper].

Then, user interest identification is conducted. User interest identification refers to using user behavioral data to establish user-verb-object triples, calculating the interest intensity of different interaction types for different document topics. When users interact with content, the system aggregates concepts identified in the content and weights them according to interaction type. The specific method is as follows: let n represent the number of users on the platform, p represent the number of possible interaction types, m represent the number of content items on the platform, and k represent the number of topics identified in the content. The user interest topic matrix is shown in Formula (1):

$$UC_{n*k} = \sum w_v * UA_v^{n*m} * DC^{m*k} \quad \text{Formula (1)}$$

Where UC_{n*k} is the matrix of user interest topics, so UC_{ij} is the relevance of topic C_j to user U_i; w_v is the weight of specific interaction type v, indicating the degree of interest of a specific user action in the content; UA_v^{n*m} represents the user-content interaction type v matrix; UA_v^{ij} represents the number of times user U_i completes interaction type v with content item d_j; DC^{m*k} contains the content-topic distribution, so DC^{fr} is the relevance of topic C_r to content item

d_f.

Formula (1) applies to the initial system configuration. However, user interactions are dynamic, and the system continuously records user behavioral data while user interests are constantly changing. Therefore, in each user-content interaction, the user interest model should be updated in real-time, as shown in Formula (2):

$$UC_{after}^{1*k} = UC_{before}^{1*k} + w_v * UA_v^{1*m} * DC^{m*k} \quad \text{Formula (2)}$$

Where UC_{before}^{1*k} represents the user-interest topic matrix before this interaction, and UC_{after}^{1*k} represents the user-interest topic matrix after this interaction. Based on the user interest topic matrix, the user interest dimension topic-feature word label system is obtained and sorted according to weight.

3.2.3 Social Dimension The social dimension lies in constructing a user relationship network graph. Digital libraries themselves have characteristics of virtuality, interactivity, openness, and freedom, encouraging users to participate in knowledge exchange, sharing, dissemination, and innovation, thus requiring the establishment of a good knowledge exchange community. Drawing on user-generated content (UGC) communities, which play an important role in knowledge acquisition and transfer between users, users are carriers of knowledge dissemination and sharing while utilizing knowledge. Through interactive behaviors such as following, consulting, discussing, sharing, citing, being cited, and collaborating, good social relationships are formed. Analyzing user group attributes and using relationships such as citations and collaborations to depict associations among individual users in terms of social connections, hobbies, and research interests reveals indices such as contribution and activity visibility among user groups, forming dynamic relationship network graphs of different types and scopes. Importing user IDs and user interaction data into complex network analysis tools constructs user relationship graphs. Through influence analysis, core users are screened out. Based on user participation and activity levels, active users, silent users, lost users, and returned users are classified, forming a user hierarchical classification management system. Additionally, through clustering algorithms, similar users are discovered to form interest groups for user recommendation and expert recommendation, establishing interest groups to facilitate communication among users with common interests.

3.3 Digital Library User Profile Model

Due to differences in cognitive approaches and the purposes of user profile construction, scholars at home and abroad have successively proposed different user profile construction methods. B. Rampler advocated using sociological approaches, human-computer interaction approaches, cognitive approaches, and case-based reasoning to construct user profile models [**?**]. Y. Kritikou proposed

three functional layers for user profile modeling: the first layer is the monitoring layer, which collects values of key user indicators by monitoring user behavior; the second layer is the modeling layer, which collects user information and profiles users in an effective way, establishing relationships between users; the third layer is the adaptation layer, which updates and optimizes user profile models based on system feedback [**?**]. S. Henczel divided user profile model construction into six stages: 1) planning to determine the dimensions of user data to be collected; 2) locating existing data and identifying required supplementary data; 3) conducting surveys and collecting materials through field research and interviews; 4) analyzing user behavioral data, identifying user demand characteristics, and clustering users with common demand characteristics; 5) generating user profiles and evaluating them; 6) improving user profiles and perfecting related services in continuous interaction and feedback [**?**].

Drawing on domestic and international digital library user profile models, the author believes that the essence of digital library user profiles is to fully utilize user-related data driven by data, display user needs through datafication → labeling → association → visualization, and apply them to various types of services in digital libraries to achieve precision services. The digital library profile model construction process can be roughly divided into data collection, data preprocessing, data storage, data mining, user profile formation, and visualization and application. The data collection module mainly extracts user registration information, completes basic attribute data through questionnaires or interviews, sets API interfaces for web and mobile ends to record user interaction data, and mines Web logs to collect original user-related data. The data preprocessing module screens and cleans the collected raw data, removes irrelevant data, and uses data processing tools to deduplicate user-related data, remove illegal fields, split fields, merge fields, convert resource data information code tables, and standardize data types, fragmenting and datafying user-related data to form standard user profile data. The data storage module integrates and classifies processed user profile data to form user basic attribute databases, user behavior databases, user interaction databases, and user other databases, storing them in distributed storage systems such as HBase or MongoDB for query and calculation. The data mining module needs to use big data tools such as Hadoop or Spark to mine user profile data in databases, calculate user interest weights, combine resource profiles, establish user label systems, form user label libraries, and achieve user feature labeling. Using network analysis tools such as Ucinet, Gephi, or Pajek, user relationship graphs are established, users are classified and clustered, and connections between users, experts, and resources are established to achieve association. The user profile module visualizes data mining results through histograms, word clouds, multi-dimensional multi-level label systems, and relationship graphs. Meanwhile, user profile results accept user feedback and make adjustments and improvements based on that feedback. The application module applies user profile results to knowledge services in digital libraries, improving information retrieval systems, optimizing retrieval result ranking, and achieving personalized retrieval; using user profiles to depict user

needs and interests, improving recommendation systems to achieve precision recommendation; analyzing user behaviors and group characteristics for precision publicity; and providing reference for digital libraries' decision-making. The digital library user profile framework model constructed in this paper is shown in Figure 2 [Figure 2: see original paper].

4. Applications of Digital Library User Profiles

4.1 Precision Recommendation

Traditional recommendation algorithms, such as content-based recommendation algorithms, make recommendations based on resource content features without fully understanding user interests. Collaborative filtering-based recommendations require user ratings, but users in digital libraries are less proactive, making it difficult to obtain large amounts of rating data, and collaborative filtering algorithms have the disadvantages of sparsity and cold start. Applying user profiles to digital library recommendation systems enables dynamic interest identification and user demand prediction, combined with resource profiles. Through fragmented mining, semantic description, and associative connection of digital collection resources, the deep content features of resources are grasped, achieving accurate and dynamic mapping between user profile labels and deep semantic labels of digital library resources. Combined with correlation features, contextual features, and collaborative features, the recommendation system is comprehensively upgraded, and high-performance recall strategies are designed to improve recommendation accuracy, intelligently providing users with personalized knowledge push services, enhancing knowledge discovery service capabilities, and achieving knowledge sharing and knowledge value-added.

4.2 Personalized Retrieval

Different from traditional literature retrieval methods, the distinctive feature of personalized retrieval based on user profiles lies in establishing relationships between search terms, retrieval results, and user behavioral data by analyzing big data from users' search histories. An intelligent retrieval system for digital libraries is constructed, using user profiles to provide information about user information needs, search behaviors, browsing habits, and browsing themes, customizing search methods for users through big data mining and analysis to provide personalized retrieval services. In terms of retrieval result ranking, based on user profiles' grasp and prediction of user interests and preferred topics, resources that users are most interested in and most valuable to them are ranked at the forefront of retrieval results, improving both retrieval accuracy and user satisfaction and interaction stickiness, better improving users' personalized retrieval experience, and avoiding resource redundancy problems caused by indiscriminate push.

4.3 Precision Publicity

Precision publicity differs from previous widespread publicity methods by being centered on target users, conducting publicity for target user groups and audience groups, with greater emphasis on publicity effectiveness. Based on large amounts of user data, user profiles display users' original appearance multi-dimensionally and three-dimensionally. On this basis, using tools such as audience classification, user clustering, association rules, data mining, and statistical analysis, users are comprehensively tracked and finely divided to form different user groups. For example, based on user usage habits and activity levels, users can be divided into active users, silent users, lost users, and returned users; based on user participation levels and influence, core users, loyal users, ordinary users, and potential users can be identified [**?**], enabling accurate hierarchical user management. Based on user interest characteristics, similar users can be clustered to establish user interest groups, recommending similar users and experts in the interest field to users. Different publicity methods are adopted for different user groups to accurately deliver publicity content to target audiences, avoiding publicity waste and interference to unrelated users and improving publicity effectiveness. During the implementation of precision publicity strategies, digital libraries can examine the effectiveness of publicity services through user interaction and evaluation feedback data and feed these effectiveness data back into the user profile database for timely improvement of publicity strategies.

4.4 Reference Decision-Making

For users, the labeling interpretation of user characteristics in user profiles provides a reference for self-cognition and a basis for decision-making in users' learning, research, development, and progress. For digital libraries, user profiles are used to obtain user characteristic information and dynamic interest needs to guide the service orientation and development trends of libraries, providing references for digital libraries' resource procurement, regulation formulation, and service provision. By analyzing users' needs for different types of resources in digital libraries through user profiles, the allocation ratio of traditional, semi-new, and brand-new resources in the collection can be adjusted. More importantly, through understandable user labels and identifiable user interests, user needs become increasingly concrete. The fundamental goal of digital library development is to continuously adjust its service content and types to improve user satisfaction. The introduction of user profiles provides functional orientation and service basis for digital libraries.

In the data-driven environment, as a resource, data is increasingly valued by all walks of life to enable data empowerment and realize data dividends. User data is an important manifestation of users' increasingly fragmented, diversified, and refined needs. How to mine user data, comprehensively understand user needs, and leverage data value is the key to achieving precision library services. Constructing comprehensive and three-dimensional user profiles based on user data provides new ideas for digital library knowledge service innovation. For users,

their diversified and personalized needs are met; for digital libraries, precision services are achieved through analysis of user characteristics and grasp of dynamic interests, improving user satisfaction. With the goal of win-win, this paper analyzes the connotation and characteristics of digital library user profiles, believes that data-driven user profiles should be designed around the main route of datafication → labeling → association → visualization, and constructs a multi-dimensional, multi-level, and three-dimensional user profile model from the natural, interest, and social dimensions. It designs a framework model for digital library user profiles and elaborates on the specific applications of user profiles in digital library precision recommendation, personalized retrieval, precision publicity, and reference decision-making, helping to break the shackles of information islands in the big data blue ocean, promoting the innovation and optimization of digital library knowledge services, and providing new momentum for digital library transformation and upgrading, enhancing knowledge discovery service capabilities, and achieving precision services.

References

[1] Bi Qiang, Yan Jing, Li Jie. New Trends and Requirements for Digital Library Service Transformation in the Big Data Era [J]. Information Studies: Theory & Application, 2017, 40(12): 12-16, 5.

[2] Cooper. The Inmates Are Running the Asylum [M]. Beijing: Publishing House of Electronics Industry, 2006: 10.

[3] Qi Cong, Wu Jun. Research on the Concept Origin and Application Scenarios of User Profiles [J]. Journal of Chongqing Jiaotong University (Social Sciences Edition), 2017, 17(5): 82-87.

[4] MARSHALL R, COOK S, MITCHELL V, et al. Design and evaluation: end users, user datasets and personas [J]. Applied Ergonomics, 2015, 46(B): 311-317.

[5] TEIXEIRA C, PINTO J S, MARTINS J A. User profiles in organizational environments [J]. Campus-Wide Information Systems, 2015, 25(25): 329-332.

[6] ZENG L, ZHANG Y, QIU R G. Adaptive user profiling in enhancing RSS-based information services [C]//IEEE International Conference on Service Operations and Logistics, and Informatics. Philadelphia: IEEE, 2007: 1-5.

[7] AMOS D, DI ALLO M S, DIOP C T, et al. Contextual preference mining for user profile construction [J]. Information Systems, 2015, 49(C): 182-199.

[8] ACHEMOUKH F, AHMED-OUAMER R. Representation and evolution of user profile in information retrieval based on Bayesian approach [C]//International Symposium on Methodologies for Intelligent Systems. Cham: Springer International Publishing, 2014: 486-492.

[9] YU J, LIU F, ZHAO H. Building user profile based on concept and relation for web personalized services [C]//International Conference on Innovation and Information Management. Singapore: ICIIM, 2012(36): 165-172.

[10] Chen Tianyuan. Empirical Study on Constructing University Mobile Library User Profiles [J]. Library and Information Service, 2018, 62(7): 38-46.

[11] AL-SHAMRI M Y H. User profiling approaches for demographic recommender systems [J]. Knowledge-Based Systems, 2016(100): 175-187.

[12] Fu Xiaofei. Research and Application of Mobile Advertising Recommendation Technology Based on User Profiles [D]. Chengdu: University of Electronic Science and Technology of China, 2017.

[13] Shan Xiaohong, Zhang Xiaoyue, Liu Xiaoyan. Research on User Profiles Based on Online Reviews: A Case Study of Ctrip Hotels [J]. Information Studies: Theory & Application, 2018, 41(4): 99-104, 149.

[14] Zhao Shuguang. High-Conversion Social Media User Profiles: A Deep Interview Study of 500 Users [J]. Modern Communication (Journal of Communication University of China), 2014, 36(6): 115-120.

[15] SHIRUDES B, KOLHE S R. Measuring similarity between user profile and library book [C]//International Conference on Information Systems and Computer Networks. Mathura: IEEE, 2014: 50-54.

[16] Ma Feicheng, Zhou Liqin. Knowledge Management and Services for Smart Health [J]. Journal of Library Science in China, 2018(5): 1-15.

[17] BISHOP J, LEWIS P R. BLAISE-LINE and the British National Bibliography: profiles of users and uses [J]. Journal of Librarianship & Information Science, 1985(2): 119-136.

[18] AMATO G, STRACCIA U. User profile modeling and applications to digital libraries [C]//European Conference on Research and Advanced Technology for Digital Libraries. Berlin: Springer, 1999: 184-197.

[19] ZHES L, DONG N Z, TAO S H, et al. Study on the user profile of personalized service in digital library [J]. Journal of Beijing Institute of Technology, 2005, 25(1): 58-62.

[20] MAO J, LU K, LI G, et al. Profiling users with tag networks in diffusion-based personalized recommendation [J]. Journal of Information Science, 2016, 42(5): 711-722.

[21] SHARMA D, KAUR S. Neural network classification for user profile learning over digital library recommendation engine [J]. Indian Journal of Science & Technology, 2016, 9(33): 1-7.

[22] ZAUGG H. Using persona descriptions to inform library space design [M]//The Future of Library Space. Bradford: Emerald Group Publishing Limited, 2016: 335-358.

[23] Liu Su. Discussion on User Profiles in Digital Library Knowledge Discovery Systems: Taking Tianjin Library as an Example [J]. Library Theory and Practice, 2017(6): 103-106.

[24] Yang Fan. Library Big Data Based on Profile Analysis: Taking the National Library Big Data Project as an Example [J]. Library Tribune, 2019(2): 1-8.

[25] He Juan. Application Research on Personalized Book Recommendation Based on Combination of Individual and Group User Profiles [J]. Library and Information Service, 2018, 62(15): 31-40.

[26] Yin Xiangquan, Li Shuning, Gong Jianhua. Analysis of University Library Study Room User Behavior Based on System Logs [J]. Library and Information Service, 2018(13): 67-71, 82.

[27] Xue Huanxue. Process of Creating User Profiles for University Library Subject Services [J]. Library Science Research, 2017(20): 16-20.

[28] Hu Yuan, Mao Ning. Construction of User Model for Digital Library Knowledge Community Based on User Profiles [J]. Library Theory and Practice, 2017(4): 82-85, 97.

[29] Wang Lingxiao, Shen Zhuo, Li Yan. Construction of User Profiles in Social Q&A Communities [J]. Information Theory and Practice, 2018, 41(1): 129-134.

[30] Fan Xiaoyu, Dou Yongxiang, Zhao Pengwei, et al. Research on Constructing Researcher Profiles by Fusing Multi-Source Data [J]. Library and Information Service, 2018, 62(7): 38-46.

[31] Chen Huixiang, Shao Bo. Research Status and Implications of User Profiles in Foreign Library Domain [J]. Library Science Research, 2017(2): 1-8.

[32] TRAVIS D. E-Commerce usability: tools and techniques to perfect the online experience [M]. Oxford: Routledge, 2002.

[33] FARSEEV A, NIE L, AKBARI M, et al. Harvesting multiple sources for user profile learning: a big data study [C]//Proceedings of the 5th ACM on International Conference on Multimedia Retrieval. Shanghai: ACM, 2015: 235-242.

[34] GUIMARAES T P. Perfil de usuários de biblioteca governamental: O Caso do ministério da saúde [J]. Perspectiva em ciência da informação, 2007(3): 96-115.

[35] LAFOUGE T, LARDY J P, ABDALLAH N B. Improving information retrieval by combining user profile and document segmentation [J]. Information Processing & Management, 1996, 32(3): 305-315.

[36] Wang Qiangbing, Zhang Chengzhi. Design and Implementation of User Profile Construction System Integrating Content and User Gesture Behavior [J]. Data Analysis and Knowledge Discovery, 2017, 1(2): 80-86.

[37] Zhang Huimin, Xin Xiangyang. Four Dimensions for Constructing Dynamic User Profiles [J]. Industrial Design, 2018(4): 59-61.

[38] Chen Zhiming, Hu Zhenyun. Research on Website User Profiles [J]. Computer Systems & Applications, 2017, 26(1): 24-30.

[39] RUMPLER B. A study of the impact of the user profile in documentary systems [J]. Online Information Review, 2001, 25(6): 359-365.

[40] KRITIKOU Y, DEMESTICHAS P, ADAMOPOULOU E, et al. User profile modeling in the context of web-based learning management systems [J]. Journal of Network & Computer Applications, 2008, 31(4): 603-627.

[41] HENCZEL S. Creating user profiles to improve information quality [J]. Online, 2004, 28(3): 30-33.

[42] Zeng Jianxun. Precision Services Require User Profiles [J]. Digital Library Forum, 2017(12): 1.

Author Contributions:

Xu Pengcheng: Proposed the research proposition and wrote the paper;
Bi Qiang: Proposed the research ideas and revised the paper;
Zhang Han: Wrote the English abstract and revised the paper;
Mu Dongmei: Refined the paper's ideas and revised the paper.

Note: Figure translations are in progress. See original paper for figures.