

Study of User Behavior Data Mining and Applications in University Libraries

--Taking the Central China Normal University Library as an example

Rong Xu ^a, Yi Yang

Department of Library; Central China Normal University, Wuhan, China

^axurong@mail.ccnu.edu.cn

Abstract. Through analysis of various data types generated from the library of Central China Normal University, such as admission data, borrowing data, digital library utilization data, and consulting service data, we explore the correlation information, value and improvement of library daily work and service. Finally, we propose the application mechanism and strategy of user behavior data of university library, which mainly includes the improvement of the organization system of big data collection, the evaluation of the data scale and usability, positioning and locating the user's big data personalized service demand, and guaranteeing the regularity and effectiveness of the work of each layer in the organization system.

Keywords: University Library, User, Behavior Data, Data Correlation.

1. The Type of Big Data in Central China Normal University Library

1.1 Admission Data

The library access control system requires users to enter the library to swipe cards in order to enter, which provides the library's admission data. This part of the data is currently controlled by the Information Office. From this, we can obtain the number of people entering the library, which can be accurate to each time period and counted by people.

1.2 The data Generated by the Library Huiwen Management System.

1.2.1 Borrowing Data

Borrowing data is the largest and most core data generated in a library management system. In 2015, the library lent a total of 337,842 books. In addition, the Huiwen system also keeps a record of borrowing of all books. Currently, there are more than 700,000 books available for borrowing in the library. Borrowing information includes: reader's name, card number, college, call number, title, publishing house, publication year, borrowing time, and time of entry. In other words, it contains all the reader information and book information, and there is a certain correlation between these data. Borrowing is the main function of the library and the main body of the library user behavior. Therefore, this part of the data is most worthy of value analysis and mining. For the utilization of this part of the data, the library is mainly used for the analysis of readers' reading behavior, and has not been further explored.

1.2.2 Reservation Data

In the Huiwen system, readers can directly make reservations for borrowed books, and the subscription data and borrowing data are the same, including all reader information and book information. Excavation analysis can also be carried out from the reader group, book type, and time. Generally speaking, this part of the data is not large, but it is of great significance for the library to purchase resources and effectively process the books.

1.2.3 Book Overdue Data

The book overdue record is also a very interesting part of the data generated by the Huiwen system. At present, the use of the library is limited to the sum of the overdue fee and the statistical overdue

fee. This part of the data has not been collected, analyzed and utilized. Judging from the actual library work situation, the overdue situation is in fact widespread. Analysis can be carried out from the overdue reader group, the overdue time, and the overdue book itself, thus it can provide a more scientific basis for the library's book lending management.

1.3 Digital Library Utilization Data Generated During the Process

1.3.1 Library Website Traffic

On the library website are two statistics for "Today's Visits" and "Total Visits". Looking at these data alone is of little significance. We need to monitor the time period, the monthly visits and annual visits. Through the data comparison, we can find the corresponding pattern and feature.

1.3.2 The Number of Views and Downloads of Digital Resource

The number of digital resource views and downloads is the most important and largest data generated in the digital library, and it is equally important to borrow data. From the perspective of resource utilization, digital resources are increasingly becoming the favorite and most popular form of resources for teachers and students. Digital resources are expensive and worth of hundreds of thousands of dollars. Therefore, the analysis of the use of digital resources is of great significance for the evaluation and ordering of digital resources. Digital resource browsing and downloading are the most direct manifestation of the use of digital resources. In view of this, the library attaches great importance to the statistics of digital resource browsing and downloading, but this part of the analysis is limited to the use of cost discussion, and does not involve the value of the resource content itself and the analysis of reader behavior.

Since the technology of the library on its own cannot count these data, the current data source mainly depends on the digital resource provider. Depending on the supplier, we have the several problems: First, the authenticity of the data is open to question; second, the data cannot be timely feedbacked; third, the detailed information of the data cannot be obtained. Our library is also actively seeking ways to obtain this part of the data.

1.3.3 Special Room Utilization Data

In addition to the acquisition of resources, library user behavior also includes the use of special room. The admission data is the use of the library's overall space. Besides, the library also controls the use of some special spaces. At present, our library has separate control of special space: the ladder book bar on the basement floor and research study room. There is a separate access control system to the ladder book bar on the basement floor, where readers can swipe in. A separate appointment system is used for the research and study room. Teachers and students can make appointments through their universal-campus-cards. This part of the data was not collected and analyzed.

1.4 Reference Service Data

1.4.1 Online Help Desk

The library currently provides an online help desk where readers can consult with librarians at any time. The online help desk asks the user to enter relevant information: title, name, mobile phone, email address and consultation content. The name, mobile phone, and email address belong to the user's personal information, which are only used by the administrator when they are dealing with the inquiry and problem. The details cannot be seen on the message board. The reader's personal information is of little value in the analysis, but through the analysis of the consultation content and consultation time, it is easy to find related problems at work.

1.4.2 Document Delivery and Interlibrary Loan

The document delivery and utilization system we use is the CALIS and CASHL document delivery system. The reader needs to register and submit on the system. It also has a relatively large amount of data, and can fully reflect the school's document sharing. At present, there are special documentary transfer librarians who sort out this part of the data every year. The main purpose is to understand the reader's needs, the amount of document transmission, and compare it among the national university

libraries. The National Association of University Libraries also attaches great importance to the development of document delivery, and publishes data on a regular basis.

Inter-library loan is used among a small number of registering readers, mainly limited to universities in Wuhan. It can reflect the library resource sharing of university libraries in Wuhan.

1.5 Equipment Utilization Data

It mainly includes self-service copy printers, e-book borrowers, and search machines. Currently, this part of the data has not been sorted out.

1.6 Weibo Utilization Data

Weibo is an important window for library image and resource service promotion. The major university libraries also value the importance of the management of Weibo. The data of Weibo is different from the above digital comparison, and is more like an analysis of the use behavior. At present, this part of the data has not been taken seriously and organized.

1.7 Other Data that Cannot be Collected

1.7.1 Data in the Reading Room

The same as the borrowing data, reading data in the reading room is of great significance for understanding readers' reading tendency and scientific ordering documents, and can reflect the utilization of books as a whole. When users come to the library, whether they study or read books, is an objective starting point for library management. At present, there is no effective way to collect the data of the library usage.

1.8 Information System Training Data

Our library has more than 50 information system trainings each year, and it trains more than 7,000 readers. A detailed analysis of the number of trainees, training time and the correlation with the training content is of great significance for us to improve the information search system education. At present, this part of the data has not been utilized and organized. From the practical point of view, there is no effective way.

2. User behavior Data Value Correlation and Application

Data Name	Data Detail	Data Correlation	Data Application
Admission Data	Readers's universal-campus-card information, admission time	Frequency of admission and Reader type	1. Library reception of readers 2. Effective service response
Book lending	Reader card borrowing information, book borrowing information, borrow time	Reader borrowing Pattern Book utilization	1. Reflecting the overall reading situation of teachers and students, and providing targeted guidance on reading 2. Resource utilization assessment to provide decisions for subsequent purchase 3. Effective configuration for book circulation
Book reservation	User reservation information, book reservation information, appointment time	Feature of book reservation and timing	1. Effective configuration for book circulation 2. Provide a basis for purchase

Book Overdue	Overdue reader information, overdue book information, duration of overdue	Pattern of Overdue	Optimization of book management regulations
Website traffic	Website traffic Access IP address	Website utilization	Website function optimization and maintenance plan development
Digital resource browsing and download	Browsing and downloading traffic, download time, IP address	Digital resource utilization	1. Evaluation of the use of digital resources in order to provide a basis for subsequent subscriptions 2. Understand user usage behavior and provide targeted services
Special room utilization	Reader's reservation information, reservation time, reservation volume	Space utilization	1. Effective space management 2. room alternation
Equipment utilization	Charge, utilization, user information	Status of Equipment utilization	1. Equipment management 2. Equipment placement
Online help desk	Consultation volume, time, content	Consulting behavior	1. understand reader's demands and thoughts in time 2. Service optimization and adjustment
Document delivery and interlibrary loan	User information, total amount of delivery, incoming and outgoing volume, and delivery of document information	Library resource sharing	1. Optimization and promotion of the document delivery work by itself 2. Reader service 3. Increase resource satisfaction rate
Use of Weibo	Posting volume, click volume, response volume, response content, fan volume	Weibo user behavior	1. Weibo publicity work optimization
Information System Training	Number of trainees, training content, time, audience information	Audience and content relevance	1. Training time, content optimization 2. Understand reader needs
Reading in the reading room	Reader information, reading volume, reading book information	Reading behavior of readers, book utilization	1. A comprehensive understanding of readers' reading status 2. Book utilization analysis 3. Management and space optimization of reading room

3. The Application Mechanism and Strategy of User Behavior Data in University Libraries.

1. Improve the organization of big data collection. the entire architecture consists of five parts: acquisition, storage, organization, analysis and policy-making. The acquisition part is the layer of obtaining data, which mainly consists of equipment, tools, systems and manual statistics, including data center system monitors, video monitors, R FID devices, digital cameras, network monitors and sensors. It is responsible for collecting the data related to system operations of data centers and the big data of readers' reading behavior. The storage part refers primarily to the storage of structured, semi-structured, and unstructured data collected by the data-collecting layer. Organization refers to the analysis and processing of the structural characteristics of data in the database, as well as

describing the information from the perspective of data organization. The analysis layer uses data statistical analysis, semantic analysis, visual analysis, graphical analysis, predictive analysis and data mining to efficiently and accurately search and discover the special relationships and knowledge hidden in the vast amount of data. Based on the processing results of the analysis layer, the policy-making layer supports and provide assistance to the decision-making in the aspects of the library's system operation management and the reader's big data personalized service. At the same time, the effectiveness of management and service decision results can be evaluated and optimized.

2. Evaluation on data size and availability. First, an assessment of the size, value, availability, and extent of application can be made. The scientific, normative, effective, and economic assessment in the data integration activities. We can assess whether data can meet library management, service, and decision-making needs. Furthermore, the user community needs to evaluate, so that the data is combined with the reader's needs, and then personalized service can be provided. Finally, one can evaluate the efficiency of the big data analysis system of the library, data security, management, and emergency response capabilities.

3. Position the user's big data personalized service needs. The collection of user behavior data is the basis for understanding user needs, but in turn, the investigation and understanding of user needs can make the data release better value. On the one hand, according to the user's big data information collected by the master, the position, content, habits, interests and social relations of the reader's reading activities are collected in real time. On the other hand, the library can determine the reader's personalized feature, locates personalized reading demands, improves service quality and reduces service costs in a timely, fast and accurate manner.

4. Ensure the standardization and effectiveness of the work in each layer of the organizing system. First, define the responsibility and establish a complete data collection, distribution, analysis, and application process. Each data related department proposes requirements, including time and data information. The technicians are responsible for the regular collection of the entire data and submit it to the work management department. The departments involved in each data are responsible for analysis, application, and provide a basis for policy-making. For example, the Central China Normal University Library conducts statistical analysis on the use of digital resources and funding, and provides funding basis for the management of the museum. In this process, analysis is a more important link and the most relevant part of the analyst's personal ability. Therefore, the analysis of various data cannot rely on task work of individual participants, but on the team work, which can form the optimal template for analysis. Therefore, each item of analysis work can be made standardized and also ensure the quality of the analysis.

References

- [1]. Chen Lianfang. Collection, analysis and application of Library users' small data under the big data environment [J]. *National Library Science Journal*, 2016, 25 (03): 69-74.
- [2]. Li Tian. Big Data Concept and Library Big Data [J]. *New Century Library*, 2014 (06): 24-27.
- [3]. Chen Jiayi, Qu Jianfeng and Li Bao. Research on UHF RFID data model specification of University Libraries [J]. *Journal of University Library*, 2014, 32 (05): 103-109.
- [4]. Chen Chen Chen. A Personalized Intelligence Service Model of Library Based on User Behavior Analysis in Big Data Era [J]. *Library Theory and Practice*, 2015 (02): 96-99.
- [5]. Chen Chen Chen. Construction of Personalized Interest Prediction and Discovery Model for Small Data Readers in Library [J]. *Library Forum*, 2017, 37 (05): 98-105.