# Research on Library Accurate Recommendation System Based on Big Data Technology and User Portrait

Lujie Duan

{miaomiaofangkun@163.com}

Library, Shandong University of Science and Technology, Qingdao, 266590, Shandong, China

**Abstract:** As an information sharing center, university library covers almost all languages and various carriers of information, so how to effectively use information resources and give full play to the service role of the library is very important. Nowadays, with the rapid development of the Internet, big data and cloud computing, the undifferentiated recommended content of the existing library management system has been unable to meet the diverse and personalized needs of users; for a large number of user data accumulated in the library management system over the years, the value of data needs to be excavated. Therefore, the use of data technology for innovation has become a new driving force for the development and transformation of libraries. User portrait is a kind of data analysis tool which can process the data related to users, extract the user feature vector, and then get the user feature model to display the user panorama intuitively. This paper applies the user portrait to the library field, constructs the portrait model through data analysis, and uses the portrait to effectively predict user preferences, user interests and user behavior, which can be used as a basis to achieve accurate recommendation of library books and meet the personalized needs of readers.

**Keywords:** Big Data Technology, User Portrait; Library, Accurate Recommendation System

## 1. Introduction

Nowadays, in the context of the information age, the in-depth development of artificial intelligence technology of big data and machine in-depth learning has promoted changes in the way of knowledge discovery and acquisition, and knowledge technology has driven library services to be personalized, precise and intelligent. The library is one of the important sources for the whole people to obtain information resources, and it is a new challenge for the library construction to construct a library recommendation service system that adapts to the development of the information age and the change of readers' needs. In the era of big data, libraries have the problem of information redundancy.Digital libraries can start from the supply side, introduce intelligent technology into book recommendation and resource construction, and provide readers

ICSETPSD 2023, November 17-18, Coimbatore, India

DOI 10.4108/eai.17-11-2023.2342824

with recommendation services to meet their diverse needs, so as to enhance user experience in an all-round way. Therefore, this paper comprehensively utilizes user portrait technology and collaborative filtering algorithm, through multi-type, real-time and real analysis of massive data, guided by user behavior characteristics and interest attributes, establishes a classification from user interest preferences to book categories, and constructs a precise library recommendation service model for user reading preferences. To meet the personalized needs of users.This paper applies the user portrait to the library field, constructs the portrait model through data analysis, and uses the portrait to effectively predict user preferences, user interests and user behavior, which can be used as a basis to achieve accurate recommendation of library books and meet the personalized needs of readers.

## 2. User portrait Model Construction

User portrait essentially uses all data resources related to users to label and visualize users, so as to produce precise solutions that meet users'needs in combination with corresponding scenarios. University libraries can integrate reader-related data, locate user behavior preferences, deeply mine data characteristics, establish user portrait label system, build user portrait model around label combination characteristics, and realize accurate book recommendation on this basis. The construction of library user portrait model mainly includes two parts: data analysis and processing and portrait mining. Firstly, the basic information of users and user behavior information can be obtained by using web crawler technology, and then a series of data preprocessing work is carried out on the collected data to avoid the influence of noise data on modeling and ensure the integrity of the data. On the basis of collecting and processing data, portrait mining uses machine learning algorithms to deeply mine user demand information, transforms the results into labels depicting user characteristics, and uses visualization technology to present portraits. The construction framework of library user portrait model is shown in Figure 1 below.

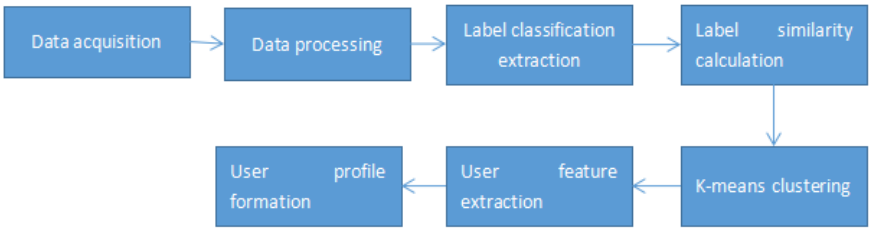


**Fig.1** Construction of Library User Portrait Model

### 2.1 User Data Processing

Library user data is generated in the process of interaction between users and library systems, and the generated data is rich and varied.For the construction of reader user portraits, multi-source data is the basic guarantee for accurately depicting reader

portraits, and the more user-related data are collected, the more accurately the portrait model can reflect the potential characteristics of users[1]. Interactive behavior data is a description of the active behavior of users in the process of using the library, covering the identity of the user login platform, the user's access time, access device support, and then through log crawling and data mining to further obtain the user's access frequency, access URL records, access location and access device parameters. Social behavior data is the data generated in the process of communication, sharing, commenting and leaving messages between users and other users on the library official website, which can be used to further understand the implicit interests of users, capture the potential needs of users, and help administrators better allocate book resources. Therefore, to build a clear portrait of university library readers, we need to consider obtaining data from the following four dimensions, as shown in Table 1 below.

**Table 1.** Dimensions of Reader Data Acquisition

| Serial number | Dimension | Content |
|---|---|---|
| 1 | Natural attributes | Name, student number, major, department |
| 2 | User preferences | Browsing, retrieving, collecting and borrowing records |
| 3 | Interactive behavior | Access time, login IP, access times |
| 4 | Social behavior | Communicate, share, leave comments |

## 2.2 User Portrait Formation

Portrait mining is the core part of personalized recommendation, using data mining technology to effectively analyze a large amount of data, it is necessary to reduce the dimensionality and weight the multi-dimensional data, and use the main information that can describe the user's characteristics as the user label to form the reader's portrait label library. The formation of user portrait is based on the analysis of data, data processing, tag extraction and user clustering, and the visualization technology is used to present the portrait. In this paper, K-means clustering method is used to divide the user group, so that each type of portrait has obvious differences, in order to better explain and describe the characteristics of users. K-means clustering, also known as K-means clustering, is the most widely used clustering algorithm at present. Its core idea is to calculate the mean center value of the samples contained in each cluster value subset, and this mean value will be the representative point of each cluster. The calculation process of the K-means clustering algorithm is shown in Figure 2 below.

On the basis of the user group clustering results, according to the user characteristics, the visual tool word cloud is used to display the user portrait intuitively, which makes the results more readable. To generate the word cloud, first import the data into the image processing module PIL and the scientific counting module numpy, import the third-party library word cloud library of Python, and further use the word cloud. Word Cloud () to generate the word cloud object and configure the relevant parameters. Then.generate () is used to load a piece of text into the word cloud object, and finally.to _ file () is used to output the generated program word cloud object as a visible image file.

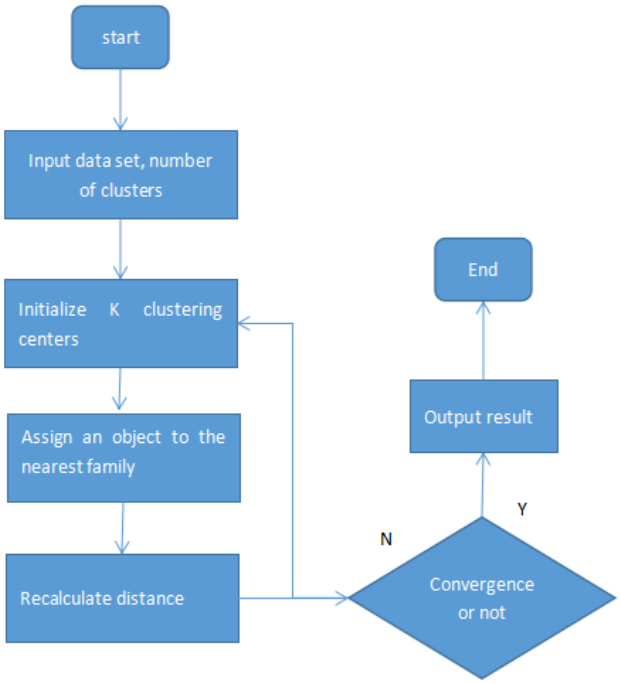


**Fig.2** Calculation steps of K-means clustering algorithm

## 3. Design of Library Accurate Recommendation System Based on User Portrait

Based on the above analysis of user portraits and the interconnection environment of digital libraries, the innovation system constructed in this section is a progressive architecture consisting of data module, recommendation module and user module. The data module is the foundation of the innovation architecture, which is the process of perception, collection and storage of data[2]. The recommendation module is the core of the innovation architecture, which is the process of user data integration, user demand modeling and matching recommendation service; As the top layer of the innovation architecture, the user module is the direct portal platform to provide services for users, and is the process to complete the recommendation task and provide decision support for users. The three modules complement each other and link up with each other to complete the virtuous circle of data from perception, collection, storage, modeling, matching, recommendation and interaction, so as to realize the innovation of precise recommendation service in digital libraries. See Figure 3 below for details.

### 3.1 Data Module

The data module is the basis for the development of accurate recommendation services, and the digital library is equipped with the Internet of Things technology.Time

perception ability, timely identification of objects and data collection, integration of data between various systems, for portraits and recommendations Modules provide basic support and build dynamic links with user modules. Only the collected data can prove the user Data becomes valuable only when it is used to create portraits. Accurate recommendation service of digital library through RFID Frequency identification), GPS (Global Positioning System), network monitor, sensor and other technologies sense and collect the real-time data of the user. The acquisition of such data cannot be perceived by the user itself, and is obtained by the one-way perception of the system. Attention should be paid to the breadth and accuracy of data acquisition. By setting up multiple data acquisition points, digital libraries can ensure the integrity of data acquisition, improve the sensitivity of equipment, and achieve high-precision control of data. The data obtained at the same time are in the In the big data environment, the generation speed is accelerated, the sources are different, the data is high-dimensional and sparse, and the proportion of data structure is low. Unstructured, semi-structured and streaming data have become common types and need to be deconstructed. After a series of standardized processing such as selection, cleaning, detection, conversion and loading of data with different structures, Cloud storage stores structured and ordered data in a large data repository to facilitate the sharing of resources[3]. Using cloud computing for data mining, the valuable knowledge is stored in the knowledge base, and the user data is extracted. It is stored in the user's personal database for portrait analysis.

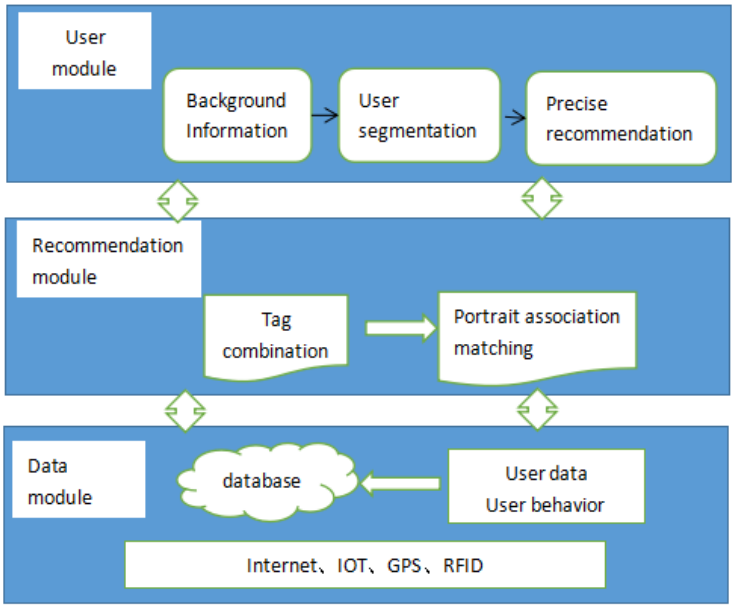


**Fig.3** Digital Library Precision Recommendation Service Innovation System Architecture

### 3.2 Recommended Modules

The recommendation module is an important intermediate layer to undertake the data module and the user module, which uses portraits to analyze people, the association between content and content, as well as between people and content, through accurate extraction and prediction of user preference data, combined with relevant information to establish a demand model for users, to achieve the service process of demand modeling, demand resource matching and recommendation. Through the data layer to collect the user data to obtain the original data of the portrait, with the help of classification and clustering algorithms to extract the user's factual attribute data, such as user's natural attributes, behavioral attributes, preference attributes and so on. Statistical analysis, machine learning and other methods are used to extract features, label user fact dimension data, construct user feature labels, identify users, deeply understand user interests, user activity, user association, user value and other information, and further give weight to labels through user modeling analysis. Make tags valuable while improving the measurement of value to users[4]. Based on supervised learning, regression prediction, linear programming and other ideas, prediction labels are formed to predict user behavior trends, interest levels, churn levels and so on. Based on different business scenarios, the system uses certain logical reasoning and business rules to flexibly combine feature tags and prediction tags to form customized business tags and complete the construction of user portraits in specific scenarios. In order to meet the needs of each user, clustering is used to combine users with common needs to form a group with common needs, usage patterns and easy management, and further understand the usage patterns of users. At the same time, because the user behavior is constantly changing, each user behavior record is monitored in real time, the label is updated, the generation of user preference label granularity level is realized, and the transformation of individual users in different groups is accurately identified.

### 3.3 User Module

User module is the application module of precise recommendation service, which is the platform of direct communication between digital library and users. When users register and log in to the recommendation portal, the system can divide users into different disciplines, different specialties, different ages, different periods, different fields and other user groups according to the collected user background information, and provide different recommendation services for different user groups, such as scientific research users[5]. It can provide recommendations for journal rankings, databases and research reports in the field of scientific research; According to the user's recommendation request, behavior habits and other clear needs of target users, find and locate core users, ordinary users, potential users, etc. Generally speaking, the image granularity of core users is more delicate and accurate, recommend from the core users, combine with the user's interest points to recommend knowledge that users may be interested in, and the service is more targeted; Similar users are clustered according to their interest characteristics, and user interest ecosphere is established to provide services such as user evaluation, similar user recommendation and expert recommendation for users.

## 4. Implementation of Precise Recommendation Service System in Digital Library

### 4.1. Collaborative Filtering Algorithm with User Portrait

If the data has a large scale, it is very time-consuming to find the K nearest neighbors of user U based on the global User or Item in the process of computing KNN in collaborative filtering[6]. At this time, we can first use the clustering method to divide the crowd or items, and then carry out collaborative filtering recommendation in the cluster to which the target or Item belongs, and calculate KNN, which is better than the global collaborative filtering effect, and can solve the cold start problem of new users to a certain extent. Because this idea first recognizes that the interests and hobbies of different groups of people are different and have patterns. The user-based collaborative filtering algorithm pattern is shown in Figure 4 below.

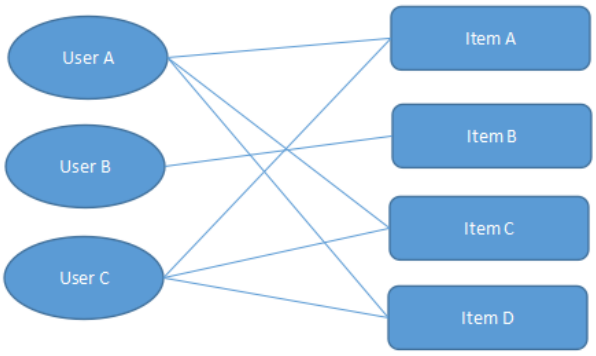


**Fig.4** User-based collaborative filtering algorithm schema

### 4.2. Evaluation Index of Recommendation Results of Collaborative Filtering Algorithm

Perfect recommendation system needs to accurately predict the user's personalized needs and reading preferences. Provide accurate recommendation service for users.

(1)Accuracy

Accuracy is the most common recommendation system evaluation index, the core of which is the ratio of the items that users are really interested in to the total number of recommendation lists. Generally speaking, the recommendation effect is proportional to the accuracy, but sometimes the high accuracy cannot represent a good recommendation algorithm[7]. For example, in the case of imbalance between positive and negative samples, it is meaningless to use the accuracy as the recommendation result of the evaluation index. Let R (u) be the list of recommendations made to the user based on the user's behavior on the training set, and T (u) be the list of the user's behavior on the test set. The specific calculation formula of the accuracy rate is as follows:

$$\Pr ecision = \frac{\sum u \in U \mid R(u) \bigcap T(u) \mid}{\sum u \in U \mid R(u) \mid} \quad (1)$$

(2Recall rate

Recall refers to the ratio of the items that the user is really interested in in the recommendation list to the total number of items that the user is interested in in the test set. The specific calculation formula of recall rate is as follows:

$$\mathrm{Re} call = \frac{\sum u \in U \mid R(u) \bigcap T(u) \mid}{u \in U \mid T(u) \mid} \quad (2)$$

(3)Comprehensive index value

When there is a contradiction between the accuracy and recall indicators, they need to be considered together. The most commonly used method is the composite index F value. The comprehensive evaluation index is the weighted harmonic mean of precision and recall.

### 4.3. Precise Recommendation Service Implementation Process

The precise recommendation method of digital library based on user portraits is aimed at the clustering analysis results of user portraits. Through collaborative filtering algorithm, the similarity of reading type preferences between target users and each type of user groups is calculated, and the books that users similar to target users are interested in are recommended to target users, so as to achieve accurate recommendation to meet the personalized needs of users. Collaborative filtering is one of the most important technical ideas of recommendation system, collaborative filtering algorithm can accurately mine the potential psychological needs of users, and the recommendation effect is targeted and intelligent, so digital library can apply collaborative filtering algorithm to library personalized precise recommendation service[8]. The core content of collaborative filtering is to find users with similar interests, and to predict the interest of the purpose according to the interests of similar users, which can be divided into building user-item matrix, finding users with similar interests and generating recommendations. In this section, the collaborative filtering algorithm is used to realize recommendation based on the establishment of library user portraits. The book recommendation mode based on user portraits is shown in Figure 5.

The user-book type matrix is constructed according to the user, the book type, and whether they like it or not. Firstly, a preference matrix of the user U for the book I is established, and the scoring value is represented by 1 if the library user likes the book I and 0 if the library user does not like the book I[9]. The users were divided into six categories by K-means clustering analysis. Combined with the information about favorite book categories filled in the library user questionnaire, add 1 to the users with the same reading interest preference, and construct the user-book type scoring matrix R as shown in Table 2.

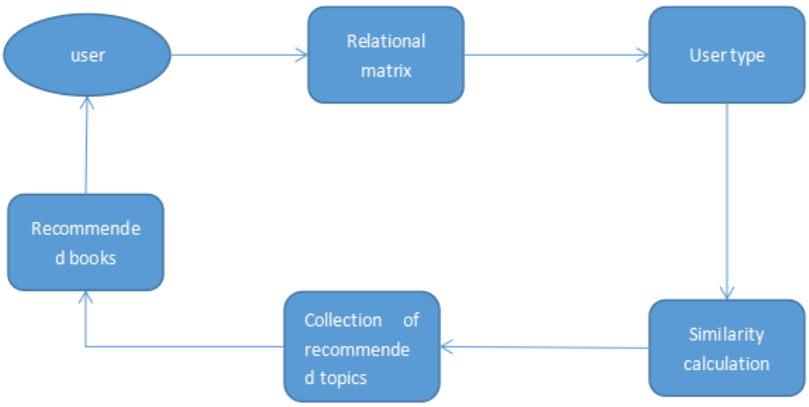


**Fig.5** Digital Library Precision Service Recommendation Process

**Table 2.** User-book Preference Matrix

| | I1 | I3 | I4 | I5 | I6 | I7 | I8 | I9 |
|---|---|---|---|---|---|---|---|---|
| U1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| U2 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 |
| U3 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 |
| U4 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| U5 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| U6 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| U7 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| U8 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 |
| U9 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 |

The calculation of nearest neighbors is also a key step in the recommendation process, and the nearest neighbor set is generally defined by the op-N threshold. In this chapter, the top-N method is used to obtain the result, that is, the matching set of close neighbors of the target user is obtained, and the top N users with the maximum similarity are obtained by arranging the similarity degree between the target user and the combination of the close neighbors. Let Sim (U1, U2) denote the similarity between user U1 and user U2, K (U1) is the reading preference set of user U1, and K (U2) is the reading preference set of user U2. According to the established user-book type matrix, the degree of similarity between two users is calculated by using a very common method of calculating similarity-Pearson correlation coefficient method, and the order of users with greater similarity is obtained.

Before recommending the target user, the neighbor set N of the user U is found and represented by S (u, I), the book types preferred by the user in S are listed, and the book types in each list are quantified through a formula[10]. O as to calculate the preference degree of the user for the book type I, sort the preference degree of the user for each type of book from high to low, and recommend the book with high preference degree to the target user. Where Pui represents the rating prediction of the

user U1 for the i-type book, Ru represents the average rating of the user U1, and Sim represents the degree of similarity between the user U1 and the user U2.

### 4.4. Discussion

This paper mainly designs the precise recommendation system of university library based on big data technology, and the research results of this paper can provide certain reference for the construction of the intelligent library in Chinese universities. But at the same time, this paper also has problems such as the selection of household portrait data and the construction of user portraits. According to the research content and shortcomings of this paper, I will continue to study the accurate recommendation service of digital library for user portraits in the follow-up study and work process.

## 5. Conclusion

To sum up, in today's era of big data, the innovation of technology has led to the continuous development and transformation of digital libraries, at the same time, a large number of complex data has brought burden to users. In this case, how to achieve accurate information recommendation for users has become one of the difficult problems of digital libraries. With the progress of external technology and the change of internal service of digital library, user portrait has become a new perspective to solve the problem.Through the construction of user portrait model, accurate prediction of user needs, refined positioning of user groups, digital library can fully understand the differences between user groups, carry out different recommendations for different users, and meet user needs.

## References


[1] An Yujie, Yan Yuwei, Intelligent retrieval method of library document information based on hidden topic mining[J]. Web Intelligence, 2022, 20(2):93-102.

[2] Adrian St Patrick Duncan, . The intelligent academic library: review of AI projects & potential for Caribbean libraries[J]. Library Hi Tech News, 2022, 39(5):12-15.

[3] Majideh Sanji, Hassan Behzadi, Gisu Gomroki, et al. Chatbot: an intelligent tool for libraries[J]. Library Hi Tech News, 2022, 39(3):17-20.

[4] Subhajit Panda, Rupak Chakravarty, . Adapting intelligent information services in libraries: a case of smart AI chatbots[J]. Library Hi Tech News, 2022, 39(1):12-15.

[5] Chi, Hailing, . Retraction Note to: The characteristics of rainfall in coastal areas and the intelligent library book push system oriented to the Internet of Things[J]. Arabian Journal of Geosciences, 2021, 14(22):1-1.

[6] Neroda T.V., Ivaskiv R.R., Slipchyshyn L.V., et al. Intelligent search engine development in designing the multilevel model of academic library[J]. CEUR Workshop Proceedings, 2021, 2917:128-140.

[7] Zhang Steven E., Nwaila Glen T., Bourdeau Julie E., et al. Deriving big geochemical data from high-resolution remote sensing data via machine learning: Application to a tailing storage facility in the Witwatersrand goldfields[J]. Artificial Intelligence in Geosciences, 2023, 4:9-21.

[8] Barron-Lugo J. Armando, Gonzalez-Compean J.L., Lopez-Arevalo Ivan, et al. Xel: A cloud-agnostic data platform for the design-driven building of high-availability data

science services[J]. Future Generation Computer Systems, 2023, 145:87-103.
[9] Zhang Zicheng, Lin Xinyue, Shan Shaonan, et al. Big data-assisted urban governance: An intelligent real-time monitoring and early warning system for public opinion in government hotline[J]. Future Generation Computer Systems, 2023, 144:90-104.
[10] Dou Hui, Zhang Lei, Zhang Yiwen, et al. TurBO: A cost-efficient configuration-based auto-tuning approach for cluster-based big data frameworks[J]. Journal of Parallel and Distributed Computing, 2023, 177:89-105.