

## *Retraction*

# **Retracted: Research on Fast Recommendation Algorithm of Library Personalized Information Based on Density Clustering**

## **Security and Communication Networks**

Received 26 December 2023; Accepted 26 December 2023; Published 29 December 2023

Copyright © 2023 Security and Communication Networks. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

## **References**

- [1] Y. Shi and Y. Zhu, "Research on Fast Recommendation Algorithm of Library Personalized Information Based on Density Clustering," *Security and Communication Networks*, vol. 2022, Article ID 1169115, 9 pages, 2022.

## Research Article

# Research on Fast Recommendation Algorithm of Library Personalized Information Based on Density Clustering

Yuqing Shi <sup>1</sup> and Yuelong Zhu<sup>2</sup>

<sup>1</sup>Library, Hohai University, Nanjing 210098, Jiangsu, China

<sup>2</sup>College of Computer and Information Engineering, Hohai University, Nanjing 210098, Jiangsu, China

Correspondence should be addressed to Yuqing Shi; shiyuqing@hhu.edu.cn

Received 8 November 2021; Revised 25 December 2021; Accepted 5 January 2022; Published 29 January 2022

Academic Editor: Jian Su

Copyright © 2022 Yuqing Shi and Yuelong Zhu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to improve the accuracy and efficiency of library information recommendation, this paper proposes a fast recommendation algorithm for library personalized information based on density clustering. According to the analysis of the clustering principle, the algorithm achieves the clustering of library information by designing density interval function. Then, the collection priority of library personalized information is judged, and the library personalized information is recommended quickly by designing tags according to the library users' preferences. Experimental results show that the recommendation accuracy and  $F$  value of the proposed algorithm are higher than those of the two traditional algorithms, and its coverage rate is higher and the mean absolute error is lower, indicating that the proposed algorithm effectively achieves the design expectation.

## 1. Introduction

In today's information age, science and technology are changing with each passing day, especially the rapid development of computer information network technology and communication technology, which has brought about great convenience to our daily lives. At the same time, the development of database technology also enables to store a large amount of information about products and users in all walks of life. With the continuous development of information technology and the rapid growth of information, the processing of massive data has become the key to the development of information. In addition to the huge amount of data, there are also different types of data [1].

In the educational field of university libraries, various information technologies have been used to improve their hardware and software conditions. The library updates its collection resources every day, so that library information accumulates a large amount of data in the database. However, hidden in these information is a lot of knowledge worthy of further study by library service workers, such as

the association rules between readers borrowing books. It is found that these rules can realize personalized book recommendation for readers and optimize collection shelf management [2].

The rise of digital library in the 1990s in the United States—this concept was put forward by the academic field and it garnered worldwide attention. Each institution carries on the discussion, the design, and the development one after another, carries on the test for each model, and carries on the in-depth research and the improvement continuously. It is an entity that can collect, process, and mine information. With the development of electronic information and virtualization technology, the realization forms of digital library are also becoming richer [3]. Compared with traditional libraries, it can provide all-day, open, and interactive resource sharing to a large group of users. Another obvious difference is that the data processed by digital libraries is massive and multi-form, and this growth rate is still the development trend. In terms of user experience, it classifies different books, periodicals, and other resources, including academic journals, newspapers, conference papers, and

academic dissertations, which not only optimize the structure of the database but also facilitate users to narrow the query range for quick positioning.

Informationization makes sure that the library's massive collection of books, periodicals, and other resources can be saved into the database system. It not only increases the library's information storage ability but also introduces the database to facilitate the insertion, modification, inquiry, statistics, and other functions, simplifying the operation and improving the efficiency, thus forming the initial information of the library. Then, it obtained the rapid development of the library online network, made the user's operation and the library everyday tasks usher in the era of automation, including the user's information retrieval, the user personal lending operations of procurement, books and periodicals, library cataloging, information storage, the daily management of library management, and other internal business [4].

All the services provided by library work are oriented to the needs of readers, and now the library services provided by readers are constantly improving and personalized information needs, so how to accurately determine the needs of library resources and readers who are interested in book information and how to find readers who chooses books are urgent problems need to be solved at present. The research on this aspect has promoted the transformation from the traditional library service idea to the personalized information service idea.

In reference [5], a library personalized fast recommendation algorithm based on data mining technology is designed. The algorithm first introduces the main methods and organizational structure of data mining, and then improves the Apriori algorithm in the classical association rule mining algorithm to improve the operation efficiency of association rules. Finally, the improved Apriori algorithm is used to carry out association analysis on the historical data of book lending, so as to make personalized recommendations for readers. In reference [6], a library personalized recommendation service algorithm is designed. From the perspective of library personalized recommendation, the algorithm digs the information data of library users, puts forward the operation conditions of personalized service, and designs the algorithm flow of personalized service on this basis. In reference [7], a digital library personalized information dynamic recommendation service algorithm integrating real-time situation is designed. This algorithm analyzes the role of personalized information dynamic recommendation service on digital library from three aspects of discipline classification, resource integration, and user information management. This paper analyzes the demand of personalized dynamic information recommendation service system of digital library integrating real-time situation, constructs the model of personalized dynamic information recommendation system of digital library integrating real-time situation, and probes into the functions of real-time situation transmission layer, collection resources management layer, and personalized information recommendation layer. However, in practical application, the traditional recommendation algorithm is not feasible, and the recommendation accuracy needs to be improved.

Although library recommendation service has developed, it still does not make full use of valuable information in the library system. For example, the professional, interest, knowledge level, and other information in the user's registration information represent the user's interest preference and knowledge level; the users' browsing history records, which contain a large number of users' browsing habits, further can be drawn between the rules of the relationship between books; information such as the time of book storage and page views can reflect the value of books to readers. If we can fully mine this information and use this valuable intrinsic information for recommendations, we can provide a nearly customized service for a variety of user groups.

Therefore, aiming at the shortcomings of traditional recommendation algorithms, this study designed a fast recommendation algorithm for library personalized information based on density clustering. Density clustering examines the connectivity between samples from the perspective of sample density, and expands from connectable samples until the final clustering result is obtained. This algorithm can overcome the situation that  $K$ -means and BIRCH are only suitable for convex sample sets. So, in order to cluster data efficiently, this paper proposes a fast recommendation algorithm of library personalized information based on density clustering.

The main processing steps of this new method can be described as follows: firstly, the data set is divided into basic clusters, and then the closer basic clusters are combined by using the idea of condensed hierarchical clustering to obtain the final cluster division.

The main contributions of this paper can be given as follows:

- (1) This paper is mainly concerned with a fast recommendation algorithm of library personalized information, which is very important but has not been well solved.
- (2) This paper introduces the density clustering method in a practical problem of fast recommendation algorithm of library personalized information.

The rest of this paper is organized as follows: section 2 is the density clustering process design. Section 3 gives the description of fast recommendation of library personalized information based on density clustering. Experiment and result analysis are given in section 4. The last section gives the conclusion.

## 2. Density Clustering Process Design

*2.1. Analysis of Clustering Principle.* Generally, clustering is a method to find similar data objects in the data set. There are certain similarities within the same cluster, but there are obvious differences between the clusters. A data instance in clustering is often referred to as an object. However, in real life, real data often appear as an object, which may not only be a description of a number or a group of numbers but also a collection of attribute descriptions [8].

Object is a better representation of how actual data exist. When the object in clustering appears in the field of

numerical analysis, it is often regarded as a data point; especially in two-dimensional or three-dimensional data space, the intuitive significance of these data points is more obvious, and the process of clustering is the process of finding the area of data aggregation. However, when the size of the data set is very large, it will be difficult to directly observe the clustering region, so we must find an automatic method to calculate the clustering situation.

From the point of view of mathematical definition, the clustering process can be defined as such a set division. It is supposed that there is a data set, including multiple data objects, and it is necessary to find an appropriate distance function to measure the distance between two points. Only when the similarity degree between two objects can be scalar by using this distance function can we find several similarity clustering between them [9].

Generally speaking, there are the following types of clustering objects:

- (1) Interval measure attributes: They can be discrete values or values on a continuous interval. In practice, we need to normalize these values into intervals  $[0, 1]$ .
- (2) Proportion measure attributes: Their value attributes are also numerical categories, but they are different from the linear attributes of interval measure attributes, but nonlinear laws, such as the interval of some exponential features. In practice, these values also need to be standardized. One of the methods is the same idea as above, but the distribution on the unit scale is extremely uneven. Another approach is to perform a function transformation on all data sets in the region. For example, exponential data can be logarithmically transformed and then normalized.
- (3) Symbolic attributes: They belong to the unordered category and are not represented by a single number. These attributes can be a set of states or categories. For example, a person's personal hobbies and include surfing the Internet, reading books, and singing. These data are different from numerical types, and they have no relationship in order. A common special case in this class of attributes that has only two state attributes, such as true and false, is called a Boolean attribute.
- (4) Sequential attributes: They are similar to symbolic attributes, but the most obvious difference is that their attributes also have sequential relations like numeric types. For example, at elementary, intermediate, and advanced levels, the most common approach is to convert these attributes into interval measure attributes and then standardize the data [10].
- (5) Mixed attributes: In the above cases, we assume that the object type in the cluster only occupies one of them. The objects in practical application often occupy several types of algorithms, and the clustering of objects with mixed attributes will greatly increase the complexity of the algorithm.

**2.2. Density Interval Function Design.** Density interval clustering is a kind of multidimensional data processing process constructed by designing distribution function. The density function is mostly generated based on the density coefficient. Compared with other algorithms, density-based clustering method can find clusters of various shapes and sizes in noisy data. According to different clustering indicators, the process will divide and summarize the density grade represented by the object to be clustered, and then judge the density category of the object to be clustered, so as to complete data aggregation processing [11].

Compared with other commonly used clustering methods, the density interval clustering process can not only avoid a lot of calculation, the evaluation results and qualitative analysis not fitting, and other problems, it is less affected by the objective environment, so that the data processing results are more real and reliable.

Therefore, before formally implementing the density interval clustering, this paper first designs the density interval distribution function.

Assume that the data set to be clustered is  $A = \{A_1, A_2, \dots, A_I\}$ , the data density indicator set is  $B = \{B_1, B_2, \dots, B_I\}$ , and the density category of the clustering result is  $H$ . Then, the weight of the  $N$ -th density indicator in the  $M$ -th cluster is  $\omega_{MN}$ , and the distribution function of the  $N$ -th density indicator in the  $M$ -th cluster belonging to the  $H$  density category is  $F_{MN}^H(\omega)$ . If  $F_{MN}^H(1)$ ,  $F_{MN}^H(2)$ , and  $F_{MN}^H(3)$  can represent the turning point of  $F_{MN}^H(\omega)$ , the distribution function can be expressed as follows:

$$F_{MN}^H(\omega) = \begin{cases} 0, & \omega \notin [F_{MN}^H(1)], \\ \frac{\omega - F_{MN}^H(1)}{F_{MN}^H(2) - F_{MN}^H(1)}, & \omega \in [F_{MN}^H(1), F_{MN}^H(2)], \\ 1, & \omega \in [F_{MN}^H(2), F_{MN}^H(3)]. \end{cases} \quad (1)$$

If the distribution function does not have the first two transformation points, it is a lower bound distribution function. If the distribution function does not have a third transformation point, it is an upper bound distribution function [12].

In order to more clearly describe the related uncertainty information of the data processed by clustering, the weight of the  $N$ -th density index in the  $M$ -th clustering was represented by the interval density grade, and it was denoted as  $\omega_{MN}\Phi$ , and  $\omega_{MN}\Phi = [\omega_{MN}^X, \omega_{MN}^Y]$ .

In general, if the interval density grade is too small or too large, it will have a negative impact on the final clustering result. Then, the relationship is shown in the following formula:

$$\omega_{MN}^Y - \omega_{MN}^X > \min(\omega_{MN}(3) - \omega_{MN}(2) - \omega_{MN}(1)). \quad (2)$$

If the data satisfying the relationship shown in Formula (2) are removed, the distribution function of the interval density grade can be expressed as  $F_{MN}^H(\omega\Phi)$ .

**2.3. Density Interval Clustering Process Design.** Based on the above design distribution function, the design density interval clustering process is as follows:

*Step (1).* Let the distribution function of the  $H$ -th density subclass of indicator  $N$  be  $F_{MN}^H(\omega\Phi) = 1$ ;

*Step (2).* Determine the clustering weight of the  $N$ -th density indicator as  $\omega_N$ ;

*Step (3).* In combination with the distribution function and clustering weight, assume that the sample value of object  $M$  about index  $N$  is  $P_{MN}$ , and then calculate the clustering coefficient  $\rho_{MN}^H$  of its density interval, and the process is as follows:

$$\rho_{MN}^H = \sum_{I=1}^N F_{MN}^H \times P_{MN} \times \omega_N. \quad (3)$$

*Step (4).* Judging the category of sample values according to the density interval clustering coefficient, so as to complete the density interval clustering.

### 3. Fast Recommendation of Library Personalized Information Based on Density Clustering

**3.1. Priority Discrimination of Library Personalized Information Collection.** This paper classifies different kinds of library personalized information by using a quadrant map. Firstly, the collection priority set is divided and different collection schemes are formulated. In the process of collection, in order to avoid interference and other special circumstances in the process of library personalized information collection, a flexible collection method is proposed, which quantifies the degree of data fluctuation with double judgment and formulates a stress adjustment scheme.

In order to improve the reliability and accuracy of the library's personalized information collection, it is necessary to adopt the data related to all cycles as far as possible to build a unified data environment and make a comprehensive judgment on the state of information [13]. Therefore, this study adopts the method of four-quadrant graph to judge the collection priority of library personalized information data.

Using fuzzy comprehensive evaluation method for data acquisition, set priority for the process of judgment; according to priority, set the key to evaluate library personalized information and real-time data acquisition, divided into different quadrants; on the basis of the different ways of acquisition scheme, for different categories of data, make the corresponding reasonable judgment; construct a subset of judgment factors, including the importance of library personalized information discrimination and the real-time requirements of data collection; determine the corresponding weight of each factor; determine the judgment set: the first quadrant, the second quadrant, the third quadrant, and the fourth quadrant; after several

professionals studied the evaluation results of various factors, a factor discrimination matrix was obtained as follows:

$$R = \begin{pmatrix} r_{11}, r_{12}, r_{13}, r_{14} \\ r_{21}, r_{22}, r_{23}, r_{24} \end{pmatrix} = \begin{pmatrix} 0.4, 0.1, 0.1, 0.4 \\ 0.4, 0.4, 0.1, 0.1 \end{pmatrix}. \quad (4)$$

Then, the fuzzy comprehensive discrimination is:

$$B = (0.5, 0.5) \otimes \begin{pmatrix} 0.4, 0.1, 0.1, 0.4 \\ 0.4, 0.4, 0.1, 0.1 \end{pmatrix}, \quad (5)$$

where  $\otimes$  is A fuzzy operator.

The discriminant result of data acquisition priority set can be obtained through the above calculation process. According to the result, the personalized information data of the library is collected in real time according to the higher frequency, and the actual working state data and environmental information are collected in real time according to the lower frequency, and the method of flexible collection is proposed.

According to the difference of the actual working state in the process of data collection and the ultimate goal of the collection work, the time interval of data collection and the total amount of data to be collected should be automatically adjusted, and the data collection should be focused, and the degree of data fluctuation should be effectively evaluated.

Acquisition is conducted on the premise of changes in adjacent data. With the changes in data, the system acquisition interval will change correspondingly, and the temporary changes will cause disorderly changes in the acquisition interval, which will consume the system and cause large errors [14]. Therefore, the dual judgment method is used to calculate data fluctuation. Specific judgment methods are as follows:

The first judgment: the standard deviation is used as a quantitative standard to change the degree of data fluctuation. Each density interval consists of  $k$  collected data to form the data variation, which represents the standard deviation of collected data in each density interval. To judge the fluctuation and change of library personalized information data, the following calculation formula is used:

$$f_0^2 = \frac{\sum_{i=1}^k (f_i - \bar{f})^2}{k}, \quad (6)$$

$$f_0 = \sqrt{\frac{\sum_{i=1}^k (f_i - \bar{f})^2}{k}},$$

where  $f_i$  represents the  $i$ -th data collection point within the current density interval, and  $\bar{f}$  represents the mean value of the data collected within the current density interval.

The second judgment: due to the instability of the information acquisition system's state changes, there will be two situations. The first is that the information collection system is affected by external interference, so that the

number of collection systems will temporarily differ greatly from the value of the central data, and then be close to the value of the data center, and the collection interval will remain unchanged. In order to prevent the collection interval from affecting the collection process, the moving average method is used to judge the degree of data fluctuation. The first item of information in each density interval is taken as the standard, and the subsequent changes of each datum cannot affect the data in the whole interval. The specific formula is as follows:

$$f_1 = \sqrt{\frac{\sum_{i=2}^k (f_i - f_1)^2}{k-1}}, \quad (7)$$

$$f_2 = \sqrt{\frac{\sum_{i=2}^k (f_i - f_{i-1})^2}{k-1}}. \quad (8)$$

The second case for state cannot respond to the change of the information acquisition system; the information data in the system will appear as a short wave, after this they will generate a new center value, and the data in the system will gather again around the new center value or so, and maintain a stable state, and the interval of data collection will still be the same. The former values of two adjacent sides are used as the standard to judge whether the data in this area are stable or whether they fluctuate around the central value [15].

The two methods mentioned above are to judge the changes of collected data by analyzing the changes of system data in a certain period. This method has higher work efficiency because it does not need to be considered in the process of judging the adjacent degree of data changes. On the basis of collection efficiency, reducing external interference to the collection degree can effectively improve the accuracy of data collection.

The double judgment method is used to process the data of the information system. Due to the inconsistency of the information system and many uncertain factors, in order to make the library's personalized information collection results more accurate, the adjustment method of the data collection period and state has been adjusted. The maximum value of the information system is  $\sigma_{\max}$ , and the minimum value is  $\sigma_{\min}$ . According to the accuracy requirement of collected data and historical operating status data, the average data variation of original data is selected when determining  $\sigma_{\max}$  and  $\sigma_{\min}$ :

$$\bar{f} = \frac{\sum_{i=2}^n (f_i - f_{i-1})}{n-1}. \quad (9)$$

Calculate the changes of the maximum and minimum values of the original data, respectively, and consider the three reference numbers, respectively, to determine the values of  $\sigma_{\max}$  and  $\sigma_{\min}$ .

According to the quantified value of data fluctuation degree obtained from the above double judgment, the

method of adjusting the interval stress of library personalized information collection is as follows:

- (1) If  $\delta_{\min} \leq f_0 \leq \delta_{\max}$  is met, the collection interval of library personalized information remains unchanged; if not, the execution step (2) will be changed.
- (2) If  $f_0 > \delta_{\max}$ , the collection interval of library personalized information remains unchanged; if not, it is determined whether it is due to the change of the status of the information collection device:
  - (a) If  $\delta_{\min} \leq f_1 \leq \delta_{\max}$  or  $\delta_{\min} \leq f_2 \leq \delta_{\max}$  is met, the collection interval of library personalized information remains unchanged; otherwise, step (B) needs to be performed;
  - (b) If  $f_1 > \delta_{\max}$  and  $f_2 > \delta_{\max}$ , the collection interval is reduced; if the collection interval is not constant, perform Step (3).
    - (a) If  $\delta_{\min} \leq f_1 \leq \delta_{\max}$  or  $\delta_{\min} \leq f_2 \leq \delta_{\max}$  is met, the collection interval of library personalized information remains unchanged. If not, step (B) is performed.
    - (b) If  $f_1 < \delta_{\min}$  and  $f_2 < \delta_{\min}$ , increase the collection interval; if not, the interval stays the same.

According to the stress adjustment scheme of collection interval, if the change of data exceeds the maximum change  $\sigma_{\max}$ , the collection interval of library personalized information should be reduced. The change scale of collection interval is determined by the difference between the change degree of data and  $\sigma_{\max}$ , and the specific formula is as follows:

$$\frac{f^i - \sigma_{\max}}{\sigma_{\max} - \sigma_{\min}} = \frac{T_i - T_{i+1}}{T_{\max} - T_{\min}}, \quad (10)$$

where  $T_{\max}$  indicates the maximum collection interval, and  $T_{\min}$  indicates the minimum collection interval.

**3.2. Design User Preference Models.** In order to obtain the implicit needs and preferences of library users, it is necessary to build a user label model, which usually calculates the user preferences based on event ontology from the frequency weight factor. On the basis of weight, label clustering (TF-IDF) is used to calculate the relationship between library users and book labels.

Let  $U$ ,  $L$ , and  $Z$  represent the collection of library users, labels, and books, respectively.

- (1) When  $u \in U$ ,  $Z_u$  and  $L_u$  are the collection of books and labels labeled by library user  $U$ , respectively;
- (2) When  $l \in L$ ,  $Z_l$  and  $U_l$  are the books, marked by label  $l$  and the set of users who have used label  $l$ , respectively;
- (3) When  $z \in Z$ ,  $L_z$  and  $U_z$  are the labels of book  $z$  and the collection of library users, respectively.  $L_u(l_1u(f_1), l_2u(f_2), \dots, l_mu(f_m))$  represents a tag vector used in the tag set of all library users. The

number of labels is  $m$ , and the  $m$ -th label and label frequency of user  $u$  are  $l_m u$  and  $f_m$ .

Tf-idf formula is used to calculate  $f_m$ , and the library user preference of  $t_m u(f_m)$  is described by the following formula:

$$l_m u(f_m) = L f_u(f_m) \times \text{IDF}_u(f_m). \quad (11)$$

The frequency of label  $l$  is denoted by  $L f_u(f_l)$ , and the formula is as follows:

$$L f_u(f_l) = \frac{\text{UserCount}(u, l)}{\sum_{k \in Y_l} \text{UserCount}(u, k)}, \quad (12)$$

where the number of label  $l$  labeled by library user  $u$  is count  $(u, l)$ , the user labeled by book  $z$  labeled by label  $l$  is represented by  $\text{UserCount}(l, z)$ , a label labeled by user is represented by  $k$ , and all the labels collected and used by all users are represented by the following formula:

$$\text{IDF}_u(f_l) = \log \frac{N}{n_l}, \quad (13)$$

where all library users are  $N$ , and all the labels collected and used are  $n_l$ . Thus, the relationship between users and book labels can be obtained as follows:

$$l_m u(f_m) = \frac{\text{UserCount}(u, l)}{\sum_{k \in U_l} \text{UserCount}(u, k)} \times \log \frac{N}{n_l}. \quad (14)$$

The frequency of each label can also be expressed as follows:

$$\text{relate}(l, z) = \frac{\text{UserCount}(l, z)}{\sum_{q \in Z_l} \text{UserCount}(l, z)}. \quad (15)$$

A certain book marked with label  $l$  is  $q$ , and the value of  $\text{relate}(l, z)$  directly affects users' borrowing priority of this book. The larger the value is, the more users of the same book marked with label  $L$ , and the more consistent the relationship between label  $l$  and this book.

**3.3. Quick Recommendation of Library Personalized Information.** Library personalized information quickly recommended process is designed in this paper through the establishment of book tag event ontology and user preference model, by calculating the link between the library users and books tag, and classifying books labels to all users, and then classifying good books tag and library user preferences match, with higher similarity of book resources recommended to the user.

The more times a book is tagged, the more it is borrowed by users. The more times a book is flagged, the more times users will check it out. When a user selects a book, he searches for the similarity between the user and the system and the category label of the book and finds out the users with the same hobbies as other users through event (user) ontology and priority calculation. The similarity of label and hobby uses  $N$  kinds of books to recommend to each user [16]. In this way, users can get high similarity of book recommendation results, which are matched with labels and preferences.

Label vector of library users can be expressed as follows:

$$L_u = l_i u(f_i) \times l_i u(\text{time}_i). \quad (16)$$

The following formula is used to represent the book label vector:

$$L_z = l_j z(\text{relate}(l_j, z)). \quad (17)$$

Cosine similarity is calculated for all books and user labels represented by vector, and the similarity between the label of book  $z_j$  in the same category and the user label is as follows:

$$\text{sim}_{ij}(u_i, z_j) = \frac{\vec{u}_i^* \times \vec{z}_j^*}{\|\vec{u}_i^*\| \|\vec{z}_j^*\|}. \quad (18)$$

In order to make books and users have the same preference, a preset threshold  $e$  is needed to set  $e < \text{sim}_{ij}(u_i, z_j)$ . To make books as pre-selected recommendations, the similarity between user labels and book vectors should be calculated first, and the value obtained is greater than the threshold. When the similarity of all the pre-selected recommendations is calculated, all the pre-selected recommended books are arranged in the order of similarity value from large to small to form the TOP recommendation. In this way, the information that all users see when browsing the website is the book information that they are interested in.

## 4. Experiment and Result Analysis

**4.1. Experimental Scheme.** In order to verify the practical application performance of the library personalized information fast recommendation algorithm designed above based on density clustering, the following simulation experiment is designed.

Taking a digital library as the experimental object, the book resource data input in it is taken as the basic data. The library has a total of 1,040,325 pieces of data, including 62,139 pieces of user data, 373,362 pieces of book data, and 607,824 pieces of book borrowing records. Some data were selected as the test set of the experiment, including 500 user data, 1000 book data, and 3000 book borrowing records.

For experimental comparison, the traditional fast personalized recommendation algorithm of library based on data mining technology (algorithm of Reference [5]) and the dynamic recommendation service algorithm of personalized information of digital library incorporating real-time situation (algorithm of Reference [7]) are taken as the comparison methods.

**4.2. Performance Indicators.** Suppose the test data set is  $X$ , the user set is  $Y$ , and  $A(y)$  is the resource set that user  $y$  is interested in. Let a book column of length  $N$  be recommended by the system to user  $y$ , denoted as  $R(y)$ . Get  $R(y)$  from the user behavior record in the data set. The accuracy of recommendation results, F-measure value, coverage rate,

and mean absolute error were taken as our evaluation indexes for the experimental test.

- (1) Accuracy. Accuracy refers to the ratio of the number of recommendations adopted by users to the total number of recommendations:

$$\text{Precision} = \frac{\sum_{y \in Y} (R(y) \cap A(y))}{\sum_{y \in Y} |R(y)|}, \quad (19)$$

where  $(R(y) \cap A(y))$  is the number set selected by users in the recommendation results.

- (2) F-measure value ( $F$  value).  $F$  value is used to represent the feasibility of the recommendation algorithm. The higher the  $F$  value is, the higher the feasibility of the recommendation algorithm is. The formula of  $F$  value is as follows:

$$F = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (20)$$

where Recall is the recall rate of recommended results.

- (3) Coverage. Coverage can reflect whether recommendations are widely distributed in the data set. Coverage rate is the percentage of the total number of recommended resources to the total number of resources in the training set. The recommended chance is proportional to the value of overlying probability, and the coverage ratio is calculated as follows:

$$C = \frac{|H_{y \in Y} R(y)|}{|X|}. \quad (21)$$

- (4) Mean Absolute Error (MAE) is a commonly used standard to evaluate the excellent performance of the recommendation algorithms. Since there is no score for the recommendation results of the three recommendation algorithms in the experiment, there is only a problem of the quality of the recommendation results. Therefore, assume that the predicted recommendation result is + or -. If the recommended result is the same as the actual resource selected by the user, the recommended result is +; if the recommended result is different from the actual resource selected by the user, the recommended result is -. MAE index calculation formula is as follows:

$$\text{MAE} = \frac{\sum A(y) - (R(y) \cap A(y))}{R(y)}. \quad (22)$$

When  $(R(y) \cap A(y)) = +$ , the output result of the recommendation algorithm is the same as the book resources selected by the user. When  $(R(y) \cap A(y)) = -$ , the output result of the recommendation algorithm differs greatly from the book resources selected by the user.

TABLE 1: The accuracy of the three recommendation algorithms and parameter results of  $F$  value.

Number of experiments/times	Numerical	
	Precision	$F$ -measure
Algorithm of reference [5]		
10	0.5170	0.6628
20	0.5444	0.6402
30	0.5957	0.6823
40	0.6014	0.6886
50	0.6101	0.6984
60	0.6105	0.6901
70	0.6144	0.6940
80	0.6046	0.6865
90	0.6159	0.6942
100	0.6163	0.6960
Algorithm of reference [7]		
Precision		
10	0.5267	0.5174
20	0.5342	0.5324
30	0.5564	0.5957
40	0.5536	0.6310
50	0.5542	0.6360
60	0.5564	0.6416
70	0.5564	0.6413
80	0.5575	0.6424
90	0.5681	0.6488
100	0.5693	0.6495
Algorithm of this paper		
Precision		
10	0.8455	0.8710
20	0.8474	0.8702
30	0.8522	0.8715
40	0.8543	0.8785
50	0.8544	0.8656
60	0.8568	0.8646
70	0.8567	0.8630
80	0.8571	0.8634
90	0.8660	0.8666
100	0.8676	0.8659

### 4.3. Experimental Results and Analysis

4.3.1. Comparison between Accuracy of Recommended Results and  $F$  Value. In the experiment, 607,824 library lending records were divided into 10 groups on average, and 10 experiments were designed to calculate the average value of each index, and the average value was used as the final evaluation index of the three algorithms to test the feasibility of the different algorithms.

Table 1 is used to show the accuracy and  $F$ -value test results of the different algorithms.

According to the results shown in Table 1, it can be seen that with the increase of the number of experiments, the accuracy and  $F$  value of the recommendation results of the three algorithms will change. The accuracy of the algorithm of Reference [5] and algorithm of Reference [7] are 0.6163 and 0.5693, respectively. The maximum accuracy of the algorithm of this paper is 0.8676. The maximum  $F$  value of the algorithm of Reference [5], the algorithm of Reference [7], and the algorithm of this paper are 0.6960, 0.6495 and

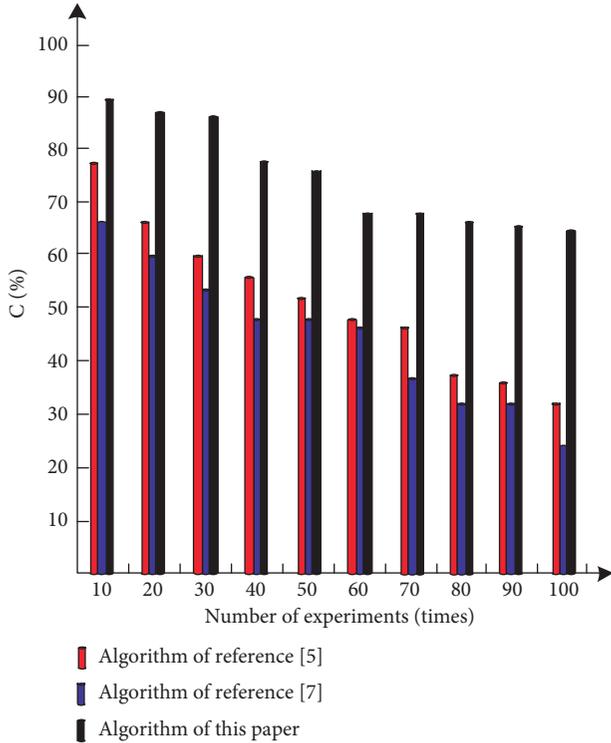


FIGURE 1: Coverage comparison diagram of the different algorithms.

0.8659, respectively. According to the above results, the accuracy and  $F$  value of the algorithm of this paper are both higher than those of the two comparison algorithms, indicating that the algorithm presented in this paper is more feasible.

**4.3.2. Coverage Ratio.** Based on the clear feasibility of the algorithm of this paper, the application performance of the algorithm of this paper is further tested. With coverage as an indicator, the application effects of the algorithm of Reference [5], the algorithm of Reference [7], and the algorithm of this paper are compared, and the comparison results are shown in Figure 1.

As can be seen from the results of Figure 1, with the increase of the number of experiments, the coverage of the three algorithms presents a declining trend. After the number of experiments is more than 60, the coverage data curve of the algorithm of this paper is gradually stable, while the two comparison algorithms do not show a steady trend of data, and the coverage of the algorithm of this paper is higher than the two comparison algorithms.

**4.3.3. MAE Value Contrast.** Based on the clear coverage of the algorithm of this paper, further test the application performance of the algorithm of this paper. MAE value is used as an indicator to compare the application effects of the algorithm of Reference [5], the algorithm of Reference [7], and the algorithm of this paper. The comparison results are shown in Figure 2.

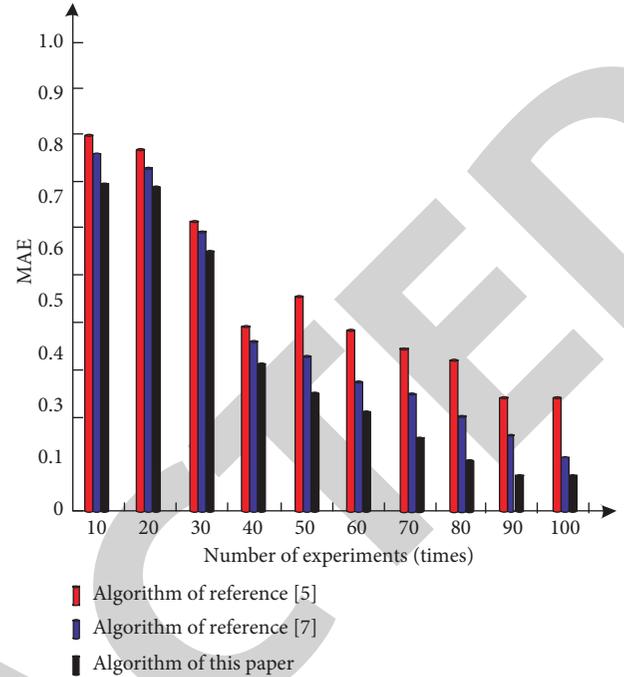


FIGURE 2: Comparison of MAE values of the different algorithms.

As can be seen from the results of Figure 2, the MAE values of the three algorithms also show a downward trend with the increase of the number of experiments. When the number of tests is less than 40, the MAE values of the three algorithms are all in a declining state, and the difference is not obvious. When the number of tests is more than 40, the MAE value of the algorithm of this paper is obviously lower than that of the two comparison algorithms.

Based on the above results, it can be seen that the recommendation results obtained by the algorithm of this paper are more likely to be adopted by users, and the gap between the predicted recommendation results and the actual resource selection results is smaller, indicating that the said algorithm has better quality of recommendation results.

## 5. Conclusion

In order to improve the performance accuracy and efficiency of the traditional library information recommendation algorithm, this paper proposes a fast recommendation algorithm for library personalized information based on density clustering. According to the analysis of the clustering principle, the algorithm achieves the clustering of library information by designing the density interval function. Then, the collection priority of library personalized information is judged, and it is recommended quickly by designing tags according to the library users' preferences. Experimental results show that the recommendation accuracy and  $F$  value of the proposed algorithm are much better than those of the two traditional algorithms, and its coverage rate is higher and the mean absolute error is lower, indicating that the proposed algorithm effectively achieves the design expectation.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 51079040/E090101 and in part by the 2021 Jiangsu Provincial Library Big Data Research Project 2021YYYYJ22.

## References

- [1] J. Zhou, "Integration mode construction of LAM digital resource based on linked data," *Library*, vol. 5, no. 1, pp. 70–75, 2019.
- [2] X. Hang, Y. Zhang, and F. Yang, "Research on document integrated retrieval method of digital library based on distributed structure," *Electronic Design Engineering*, vol. 28, no. 12, pp. 35–38+43, 2020.
- [3] Z. Yin, "Design of shared encryption platform for library electronic resources based on fuzzy clustering," *Bulletin of Science and Technology*, vol. 34, no. 12, pp. 173–177, 2018.
- [4] Y. Liu, X. Jiang, and Y. Duan, "Block chain technology: promoting the digital resource construction of university library in big data era," *Journal of Library and Information Sciences in Agriculture*, vol. 32, no. 6, pp. 15–22, 2020.
- [5] Q. Wang, "Research on library personalized fast-recommendation algorithm based on data mining technology," *Modern Electronics Technique*, vol. 42, no. 5, pp. 149–151+156, 2019.
- [6] Q. Ou, "Design of personalized recommendation service system for library under the background of smart campus," *Journal of Library Science*, vol. 42, no. 10, pp. 93–96, 2020.
- [7] L. Huang, "Research on personalized information dynamic recommendation Service of Digital library integrating real-time situation," *Journal of Library Science*, vol. 42, no. 11, pp. 75–79, 2020.
- [8] S. Laohakiat and V. Sa-Ing, "An incremental density-based clustering framework using fuzzy local clustering," *Information Sciences*, vol. 547, pp. 404–426, 2021.
- [9] Z. Taskin and U. Al, "Natural language processing applications in library and information science," *Online Information Review*, vol. 43, no. 4, pp. 676–690, 2019.
- [10] M. Mousavi, H. Khotanlou, A. A. Bakar, and M. Vakilian, "Varying density method for data stream clustering," *Applied Soft Computing*, vol. 97, Article ID 106797, 2020.
- [11] T. K. Hai, "Development of a scale for data quality assessment in automated library systems[J]," *Library & Information Science Research*, vol. 41, no. 1, pp. 78–84, 2019.
- [12] H. Y. Jeong, "Multi criteria based personalized recommendation service using analytical hierarchy process for airbnb," *The Journal of Supercomputing*, vol. 77, no. 4, pp. 1–19, 2021.
- [13] Z. Cui, X. Xu, F. Xue et al., "Personalized recommendation system based on collaborative filtering for IoT scenarios," *IEEE Transactions on Services Computing*, vol. 13, no. 4, pp. 685–695, 2020.
- [14] E. Baracchini, L. Benussi, S. Bianco et al., "A density-based clustering algorithm for the CYGNO data analysis," *Journal of Instrumentation*, vol. 15, no. 12, pp. 120–123, 2020.
- [15] C. Wang, R. Huang, J. Li, and J. Chen, "Towards better information services: a framework for immigrant information needs and library services," *Library & Information Science Research*, vol. 42, no. 1, pp. 101–109, 2020.
- [16] L. Qian, "Automatic collection and simulation of personalized book information in digital library," *Computer Simulation*, vol. 37, no. 1, pp. 390–393, 2020.