



Classification and Retrieval Method of Library Book Information Based on Data Mining

Xing Zhang^(✉)

The Tourism College of Changchun University, Changchun 130000, China
xingxing11852@163.com

Abstract. The current classified retrieval of book information is mainly based on the key words of books. When there are interdisciplinary contents in books, it is easy to ignore the correlation between book information, resulting in a sharp decline in the retrieval accuracy. In order to improve the quality of library information retrieval service, aiming at the problems of the above traditional classification retrieval methods, this paper studies the library book information classification retrieval method based on data mining. After preprocessing the book information, the Apriori algorithm of data mining is used to mine the deep management of information, and the decision tree algorithm is used to determine the classification of information. Using the concepts in book information, this paper constructs the concept map of book information and builds the information retrieval framework. Use Markov network model to realize book information classification and retrieval. In the test, the retrieval accuracy of the proposed retrieval methods is higher than 92%, and the retrieval performance is significantly improved. Through the research and application of this method, the transformation from the traditional library service concept to the personalized information service concept is promoted.

Keywords: Data mining · Library · Book information · Classified search · Apriori algorithm · Decision tree

1 Introduction

As a comprehensive storage center of social information resources, the library not only has the function of collecting and preserving archives, books and materials and continuing social memory, but also should actively serve all aspects of society, such as economy, politics, science and culture, and actively participate in the process of integration and sharing of social information sources. Libraries have also begun to apply various information technologies to improve their hardware and software conditions. Driven by the development of library digital technology, the data scale generated by book borrowing is becoming larger and larger, the data types and items are becoming more and more complex, and the information update speed is fast. The related resources have an explosive growth trend, but the rich and diverse book resources have also brought some problems to people searching books. At the same time, library users and managers have higher and

higher requirements for book classification and retrieval information, which is reflected in its stronger pertinence, higher real-time, and more diversified types of needs [1]. The library will update its collection resources every day. Readers use all kinds of resources every day, which makes a large amount of data and information accumulated in the library database. In these information, there are many knowledge worthy of in-depth study by library service workers, such as the association rules between readers' borrowed books. It is found that these rules can realize personalized book recommendation to readers, And optimize the management of bookshelves in the collection. When readers borrow, they need to actively go to the library collection rack to find the books they need or enter the library electronic information system to search for download resources [2]. The traditional management of library book information is to input the year, subject, title and author of the book into the database as the classification basis and search label. Readers need a clear search label to search the book information. For some users, the selection of search words is also a problem to be solved in book retrieval. In book retrieval, the selection of search words requires certain skills. If you want to quickly select the appropriate search words, you need to have certain professional knowledge [3]. For those who have received professional training, although they have good classification and subject indexing ability, they sometimes can't accurately distinguish what they need in the chaotic ocean of information.

Modern libraries use digital technology to provide retrieval services for readers, which has become the primary way for readers to contact the library. How to build a good retrieval system has also become one of the research hotspots of library technology. Using the retrieval system, users can retrieve all the bibliographies in the library and obtain useful information. Information mining technology has attracted more and more attention, and has become a hot spot in the current computer field. Its research focus has gradually shifted from discovery methods to system application, and pay attention to the integration of multiple discovery strategies and technologies and the mutual penetration between multiple disciplines [4]. Nowadays, the application fields of data mining are more and more extensive, from early commercial applications to scientific research, e-commerce, product control, financial industry, education and teaching and other fields. With many successful applications, the emergence of data mining technology and its wide and successful application in the commercial field provide a direction for the research of Library in the development of personalized service [5]. Using data mining technology to mine and analyze the massive borrowing information stored in the library database, we can find the hidden Book association rules and readers' borrowing preferences and habits, and improve the quality of library book management service. In order to meet the higher and higher demand standards of library readers, based on the current construction of Digital Library and the advantages of data mining technology in content information data processing, this paper will study the library book information classification and retrieval method based on data mining, so as to provide library decision-making and information reference in the collection, facilitate readers' knowledge learning and improve borrowing efficiency.

2 Research on Library Book Information Classification and Retrieval Method Based on Data Mining

2.1 Library Book Information Preprocessing

Library book information can be regarded as text information. When processing book information, text preprocessing is first needed. The storage formats of corpus text in are various, including various markers, labels, pictures, numbers and English characters. These noises have no effect on text classification, and some high-frequency useless words will even affect the effect of classification. Different from English, Chinese text has no clear blank between words, and Chinese classification is more complex than English. Therefore, the classification of Chinese information needs a Chinese segmentation system to segment the text. The purpose is to separate the words through spaces and format the text information of library books.

In order to quickly match the Chinese character string, it is necessary to convert the text into the form of Chinese character hash table. The book text information is encoded in double bytes according to GBK code, and the overall coding range is 8140-fefe. The first byte is between 81-fe and the last byte is between 40-fe. Eliminate xx7f one line, and 23940 code points in total. It includes 21886 Chinese characters and graphic symbols, including 21003 Chinese characters (including radicals and components) and 883 graphic symbols. Therefore, the selected hash function value is the first byte of the internal code of Chinese characters. Hash table (also known as hash table) is a data structure that can be accessed directly according to key value. That is, it accesses records by mapping key values to a location in the table, so as to speed up the search. This mapping function is called a hash function, and the array storing records is called a hash table. The data structure of hash table can be defined as follows [6]:

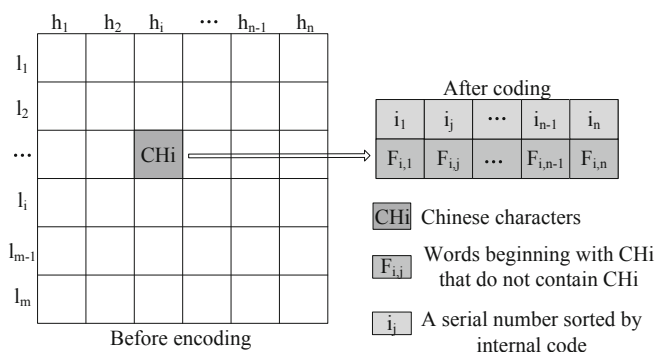


Fig. 1. Schematic diagram of hash table data structure

In Fig. 1 above, h_i is the high byte of the internal code after the book text information is encoded; l_i is the low byte of the internal code after the book text information is encoded. The subscript of the two-dimensional array is used to represent the address of the hash function. Because the inner code of each text information character is different, there is no problem of address conflict. Libraries usually store book information

in the library management system database. The data tables related to book information classification and retrieval in the database mainly include: Book Basic information table, book borrowing information table, reader information table, and Book copy table. These data tables store the original information such as book name, book classification, reader information and book borrowing, which are the basis of data mining for readers. Preprocessing the data before data mining can improve the mining efficiency.

Firstly, delete the text marks, such as the author of the book, publishing time, address source and other marks, and present the text as a single word combination to form a consistent text format; Secondly, the text is processed by Chinese word segmentation. Most text mining research usually involves words or sentences. In order to further process these words, they must be separated word by word. Therefore, at this stage, because it cannot represent any category, all words in the sentence will be segmented, which can simplify the calculation process in the next step; Finally, stop words, redundant words and duplicate words are deleted. There are usually many methods to deal with incomplete data such as missing attribute values in records. The main methods to summarize these methods are as follows: the first method is to directly delete the records containing the missing attribute data values when the proportion of missing attribute data values to be processed in the field is relatively small, However, when the proportion of attribute missing data value in the field exceeds a certain degree, the accuracy of mining results will be reduced, resulting in very poor performance; The second method is to fill in the value manually for those missing attributes. However, when there are many missing attribute values in a large data set, if this method is still used for processing, it will take older time and the feasibility is very poor; The third method is to use a predetermined possible value to fill in the missing attributes. This method is relatively simple in operation, but if it is widely used, it may cause large errors in the results obtained by mining; The fourth method is to calculate the average value of the attribute value in other object records with the same characteristics for the missing attribute, and then use the average value to fill in the missing attribute [7].

After the above preliminary processing of the book information text, the book text information features are extracted. In this paper, mutual information degree is used to characterize the co-occurrence degree between text information class and secondary information. Mutual information measurement is derived from information theory, which provides a formal method for modeling the mutual information between features and classes. According to the co-occurrence degree of class c and word w , the pointwise mutual information IW between word w and class c is defined. The expected co-occurrence of class i and word w on the basis of mutual independence is given by $p_i E(w)$, and the real co-occurrence is given by $P(w)E(w)$. In practice, the value of $P(w)E(w)$ may be much larger or smaller than $p_i E(w)$, depending on the degree of correlation between Class c and word w . Mutual information is defined according to the ratio between these two values, and its calculation formula is as follows [8]:

$$I_i(w) = \lg\left(\frac{p_i(w)}{P_i}\right) \quad (1)$$

Among them, $p_i(w)$ represents the normalized probability value of the global distribution of book information in different classes; P is the global probability of book

information text. When $I_i(w)$, WORD w is positively correlated with class i , and when $I_i(w) < 0$, WORD w is negatively correlated with class i . Note that $I_i(w)$ is specific to a specific class i . The overall mutual information needs to be calculated as a function of word w mutual information of different classes. These are defined by using the average and maximum values of $I_i(w)$ on different classes.

The use of global probability ensures that the Gini index more accurately reflects the differences between classes in the case of class distribution deviation in the whole text set. In addition, information gain is introduced to further describe the characteristics of book information. Assuming that the book information document contains the word w , let P_i be the global probability of class i and P_i be the probability of class i . $E(w)$ is the score of the text containing w .

$$G(w) = - \sum_{i=1}^k P_i * \lg(P_i) + E(w) \sum_{i=1}^k P_i * \lg(P_i) \quad (2)$$

The greater the value of information gain $G(w)$, the greater the discrimination of word w . For a text corpus containing n documents, d words and k classes, the complexity of information gain calculation is $O\{n, d, k\}$. After preprocessing the library book information according to the above process, use the data mining technology to process the book information and obtain the deep information relationship.

2.2 Information Data Classification and Mining Processing

Libraries have basically realized automatic management, and the information data of book resources can be found in the automatic management system. The main problem of association rule mining is to find the correlation between items hidden in large databases, so as to help people do well in management and make more accurate decisions. This paper uses the improved association algorithm to mine the library book information. Apriori algorithm decomposes association rule mining into two sub problems [9]:

- (1) Find all frequent sets in transaction database d that meet the minimum support minsup;
- (2) Using frequent sets to generate all association rules that meet the minimum confidence minconf;

Finding frequent sets is the core problem of association rule discovery. An itemset containing K items is called a itemset. If the itemset meets the minimum support, the itemset is called frequent itemset, which is called frequent itemset for short.

It can be found from the steps of Apriori algorithm that a large number of candidate sets may be generated when generating candidate sets. When there are 10000 frequent sets with length 1, the number of candidate sets with length 2 will exceed 10 million. At the same time, due to the need to calculate the support of itemset, the transaction database will be scanned repeatedly, which will increase the computational overhead. Therefore, this paper improves the Apriori algorithm according to the following process.

- 1) Read in the compressed transaction set (book information after preprocessing);

- 2) Scan the transaction set to find each type of frequent single item;
- 3) Connect all kinds of frequent attribute items and frequent book classification items into 2 a candidate frequency complex item set, $k = 2$.
- 4) Check the k -candidate frequent itemset and record its support and the support of the predecessor. The connection condition of frequent itemsets is that the first n item is the attribute item of readers, and the content of the attribute item of readers is different, and the last item is the same book classification item.
- 5) Output the frequent k -frequent itemsets whose confidence and support meet the requirements. The confidence is the support divided by the support of the previous part.
- 6) $k + 1$ -candidate frequent itemsets are obtained by connecting k -frequent itemsets. Pruning can reduce the number of frequently connected itemsets and improve the efficiency of program operation. The following is the rule of pruning connection:
 - a) If the last item category of frequent itemsets A and B is different, it cannot be connected.
 - b) If there are different items belonging to the same attribute category, they cannot be connected.
 - c) Frequent itemsets cannot be connected to themselves.
 - d) If conf is used to represent the antecedent support, when $\min(A. \text{conf}, B. \text{CONF})/\text{minimum support} < \text{minimum confidence}$, A and B cannot be connected.

Rule $A \rightarrow B$ has credibility. C is the percentage of A -item set and B -item set relative to A -item set, which is the conditional probability $P(B|A)$, i.e.

$$C(A \rightarrow B) = P(B|A) = \frac{|AB|}{|A|} \quad (3)$$

where $|A|$ represents the number of transactions that contain itemset A in the database.

- e) Other situations can be connected.
- 7) $k + +$ if the number of candidate frequent itemsets generated is not 0, turn to 4), otherwise it ends.

The implementation of association rule mining is to find all the strong association rules in the transaction database, and the strong association rule $A \rightarrow B$ is the rule that meets the minimum support and minimum confidence at the same time, then the corresponding item set $A \rightarrow B$ also meets the minimum support, that is, the frequent item set, and the confidence of the rule can be calculated from the support of frequent item sets A and $A \rightarrow B$. After using the improved Apriori algorithm to analyze the correlation between library book information, the decision tree algorithm is used to determine the classification category of book information.

This paper uses ID3 decision tree algorithm to determine the classification of book information. ID3 algorithm is a classification technology of decision tree. Its purpose is to select the best attributes as nodes, and take the constructed decision tree as the simplest state or close to the simplest state. The optimal node is determined by the entropy generated by its node, and its calculation method is as follows. If the objects of

a certain object set D belong to j different categories, the entropy $S(D)$ of this object set is:

$$S(D) = \sum C_j \ln C_j \quad (4)$$

where, D object set; j is the number of categories, C_j is the total number of objects belonging to category j and the total number of items of D .

Next, select an attribute X_j as the node of the decision tree, establish m child nodes under this node, and assign all objects originally belonging to the node to the appropriate child nodes. The attribute X_j values of objects assigned to the same child node must be equal. According to entropy and information gain, the execution steps of ID3 algorithm are described as follows [10–12]:

- 1) First, let the root node of the decision tree be D , and all objects belong to the object set of D .
- 2) If all objects in D belong to the same category, define node D as this category and stop, otherwise continue to step 3;
- 3) Calculate the entropy of all objects belonging to D ;
- 4) From the root node to the current node, if there is an attribute X_j that has not been a node, the D object set is divided by X_j , and the entropy and information gain of some decision trees are calculated respectively;
- 5) The candidate attribute with the maximum information gain is selected as the classification attribute of D node;
- 6) Create sub nodes under node D , which are respectively d_1, d_2, \dots, d_m (assuming that multiple attribute values are selected as classification attributes), and assign all object sets in D to appropriate sub nodes.

After the classification of library book information is determined by the decision tree algorithm, the book information retrieval framework is designed according to the book concept map.

2.3 Design of Library Information Retrieval Architecture

When searching for books of a certain subject, users do not have a clear understanding of the information of the subject they want to search. Especially for users who are beginning to learn a certain subject, because they do not understand the concept points and knowledge structure of the subject and are still vague about the subject knowledge, it may take a long time to search the subject books by themselves, and the relevant books retrieved may not be comprehensive, Users still need some help from the outside world to learn the subject well and understand the scope of knowledge contained in the subject. At the same time, in the process of book retrieval, users may also face the problem of difficult selection of search words. We need to consider how to select search words. Knowledge nodes, i.e. concept words, can be regarded as the retrieval words of book retrieval, and there is a certain dependency between concepts, which constitutes the knowledge network of books, and the concept map can clearly show the concept information. Concept map is a graphical method that uses nodes to represent

concepts and lines to represent relationships between concepts. The theoretical basis of concept map is Ausubel’s learning theory. The construction of knowledge begins with the observation and understanding of things through existing concepts. Learning is to build a conceptual network and constantly add new content to the network. In order to make learning meaningful, individual learners must associate new knowledge with the concepts they have learned.

The library information retrieval framework designed in this paper is mainly designed from the perspective of users and librarians. It can be divided into book retrieval module and concept map management module. The book retrieval module also includes concept map module. The book retrieval module is mainly aimed at users. After users input the search words, the information retrieval model provides the concept map of relevant disciplines according to the search words, and retrieves the relevant books according to the index. Users can view these books; The concept map module included in the book retrieval module has the function of uploading concept nodes. Users can supplement concept nodes here. The concept map management module is mainly aimed at librarians. In this module, librarians filter the concept information nodes uploaded by users and update the original concept map.

In this concept map, users can view and understand some discipline branches under the retrieved discipline and the conceptual knowledge points of the discipline, and have a more complete understanding of the retrieved books. In this way, they can also find other required books related to it to help users further study. For some users who need to consider the selection of retrieval words, the presentation of concept map can also help them solve this problem. The conceptual diagram is the “center - surrounding” structure diagram. Figure 2 shows the book information retrieval framework based on concept map.

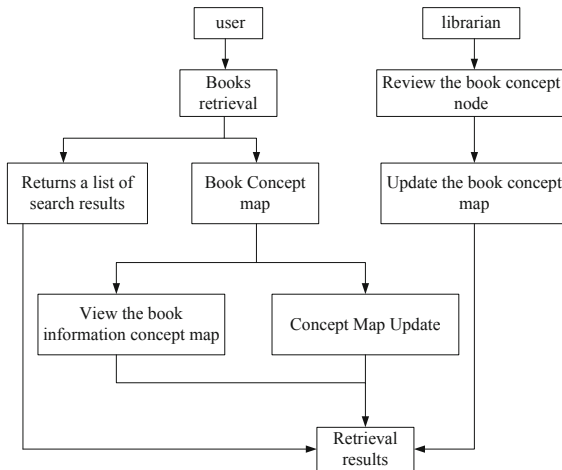


Fig. 2. Book information retrieval framework based on concept map

The central position of the concept map is the discipline name of a discipline. The first layer connected around is the first level concept node, the first level concept node is

connected downward as the second level concept node, the second level concept node is connected downward as the third level concept node, and so on. Level 2 concept nodes are included in level 1 concept nodes, level 3 concept nodes are included in Level 2 concept nodes, and concept nodes of the same level are not included in each other. After using the concept map to design the framework of book retrieval, the book information retrieval is realized on the basis of book information classification.

2.4 Realize Classified Retrieval of Book Information

This paper uses Markov network model to retrieve library book information. In the research and practice of information retrieval, there is the uncertainty of retrieval, that is, the exploratory and fuzziness of retrieval. In the behavior of information retrieval, the information needs of users are fuzzy. Usually, the information requirement submitted by the user is a short sentence or a shorter search term. Therefore, most users find that it is difficult to form an information query that can achieve the purpose of effective retrieval. Users may retrieve some information related to their query, but it is impossible to retrieve all relevant information, and the recall rate is always limited. Especially in the case of synonymy or polysemy, the retrieval efficiency is not ideal under the simple matching model between query words and words in book information. However, if multiple book information describes the same topic, there must be some connections between these book information, and some topic words will appear among them to describe a topic or some topics. Through the book information, the initial retrieval results can be corrected to avoid the poor retrieval effect caused by the simple matching of query words.

Given query q , the document set needs to calculate a final retrieval score of each document t_j and query q , record it as $p_f(t_j|q)$, and sort the retrieval results according to this score. The initial retrieval score of each document without document rearrangement is recorded as $f_o(t_j|q)$. Make full use of the information contained in the document group and $f_o(t_j|q)$ to extract the initial score of the document group. A correction score is recorded as $f_n(t_j|q)$. Finally, integrate the initial retrieval score and correction score, calculate the final retrieval score $p_f(t_j|q)$ of the document according to the following formula, and rearrange the document in ascending order of $p_f(t_j|q)$.

$$p_f(t_j|q) = \xi f_o(t_j|q) + (1 - \xi) f_n(t_j|q) \quad (5)$$

The retrieval mode used for book information in this paper is full-text retrieval. The specific process is as follows: the index is established by scanning each word in the data set, and the index database is established by the index of each word; The user inputs the search words of book retrieval at the input port, and then preprocesses the semantic of the search words input by the user; After completing the semantic query expansion, match the similarity between the index words in the index database and the semantic expansion words of query expansion; Sort the output results and return the sorted search results to users, so as to complete book information retrieval. So far, the research on library book information classification and retrieval method based on data mining has been completed.

3 Method Test

The above studies the library book information classification and retrieval method based on data mining. This section will test the effectiveness of this method. Through the classified retrieval of book information, we can better understand the advantages and disadvantages of the model, find out what factors lead to the poor performance of the method, and facilitate the follow-up to provide the basis for the improvement and adjustment of the retrieval method. Finally, we can comprehensively, quickly and accurately retrieve the required book information and improve the service quality of the retrieval method.

3.1 Test Preparation

In order to test the performance of the library book information classification and retrieval method based on data mining, experiments are carried out in a comparative way. When users use the book retrieval model for retrieval, the retrieved Book results are generally sorted according to the correlation between the book and the user's query information. In general, the retrieved results are ranked from high to low according to the correlation. The book in the front is the book with the highest correlation with the user's search information, which is also the result that best meets the needs of users. Therefore, for the book information retrieval method, the accuracy of the retrieved Book results needs to be evaluated in the process of use.

The original corpus data used in this paper is the complete book of 3000 humanities and social sciences books downloaded from the database. Because some original corpora are picture documents, and the model can not recognize them, it is also necessary to use OCR technology to convert the books into TXT file format required by the experiment, so as to obtain the initial experimental corpus data. OCR technology is a kind of character recognition technology. It can recognize and extract the characters on the picture and convert them into computer characters.

Under the environment of MATLAB software, it is configured on PC: CPU is 3.6 ghz, memory is 8 GB, and a book information classification and retrieval platform is built. The classification effect of book retrieval information is an important basis for testing the classification and retrieval effect of book information. The information retrieval method based on SOM neural network and the information retrieval method based on Top-k query algorithm are selected as the comparison group to compare with the book information classification retrieval method proposed in this paper. By comparing the accuracy and false detection rate of book information retrieval methods, this paper tests the effectiveness of book information retrieval methods.

3.2 Test Result

Table 1 below shows the comparison of accuracy and false detection rate of classified retrieval of book information of single subject type using three information retrieval methods.

Through the analysis of the data in Table 1, it can be seen that when retrieving the book information of a single subject, the error detection rate of the information retrieval

Table 1. Comparison of retrieval accuracy and false detection rate of information classification retrieval methods/%

Group Accuracy	Information retrieval method based on data mining		Information retrieval method based on SOM neural network		Information retrieval method based on Top-k query algorithm	
	Noise factor	Accuracy	Noise factor	Accuracy	Noise factor	Accuracy
1	94.3	1.39	83.0	2.48	83.2	3.19
2	96.5	2.01	84.1	2.35	82.4	3.45
3	95.7	1.38	82.2	2.63	81.5	3.35
4	97.2	1.85	82.6	2.57	81.8	3.08
5	94.1	1.48	84.3	2.56	80.7	3.24
6	95.4	1.33	85.4	2.34	81.1	3.37
7	96.6	1.87	85.7	2.61	82.3	3.55
8	96.5	1.31	82.5	2.39	83.5	3.49
9	95.8	1.73	84.1	2.65	80.0	3.52
10	96.0	1.86	86.6	2.48	81.6	3.26

method based on data mining is less than 2.1%, and the lowest error detection rate of the two retrieval methods is 2.34% and 3.08%. The information retrieval method based on data mining is far lower than the lowest error detection rate of the other two retrieval methods. The retrieval accuracy of this method is over 94%, and the book retrieval effect is obviously better than the other two retrieval methods. Compared with the two traditional methods, the library book information classification retrieval method based on data mining proposed in this paper has higher retrieval accuracy.

Table 2 below shows the comparison of the accuracy and false detection rate of classified retrieval of book information with interdisciplinary by using three information retrieval methods.

By analyzing the data in Table 2, it can be seen that the book information retrieval is disturbed by interdisciplinary disciplines, resulting in the decline of the retrieval accuracy of retrieval methods and the rise of false detection rate to a certain extent. However, the retrieval accuracy of the information retrieval method based on data mining is still higher than that of the other two information retrieval methods. Summarizing the above test and analysis contents, we can see that the retrieval accuracy of the library book information classification retrieval method based on data mining proposed in this paper is higher than 92%, and the retrieval accuracy is significantly improved. The retrieval accuracy of the two traditional methods is lower than 84% and 80% respectively. Compared with the two traditional methods, the retrieval accuracy of the library book information classification retrieval method based on data mining proposed in this paper is higher.

Table 2. Comparison of retrieval accuracy and false detection rate of information classification retrieval methods/%

Group accuracy	Information retrieval method based on Data Mining		Information retrieval method based on SOM neural network		Information retrieval method based on Top-k query algorithm	
	Noise factor	Accuracy	Noise factor	Accuracy	Noise factor	Accuracy
1	92.0	2.23	80.1	3.34	78.3	3.74
2	93.6	1.91	81.3	3.49	75.1	3.95
3	93.3	2.17	82.9	2.92	77.7	3.76
4	95.4	1.78	82.5	3.38	79.2	3.71
5	93.1	2.05	82.7	2.94	76.5	3.67
6	95.8	2.13	79.6	3.15	79.9	3.32
7	93.9	1.84	79.6	2.93	76.8	4.10
8	95.5	1.85	79.3	3.29	78.1	3.35
9	93.7	1.97	83.9	3.11	77.4	4.08
10	93.2	1.96	82.2	3.36	78.6	4.07

4 Conclusion

With the development of information technology, digital library is more and more valued by the country. The content of digital library is more and more abundant, and the number of books is growing rapidly, which makes the structure of library management system become complex and the content distribution is also very extensive. The information needs and forms of the library for users are becoming more and more diversified. In the field of book retrieval, it is the most fundamental to provide readers with retrieval results that can meet their query needs to the greatest extent, which requires the design of good retrieval methods, and to ensure the quality of retrieval results and the efficiency of retrieval speed. Book intelligent retrieval system is not a simple combination of traditional retrieval system and computer network technology, but needs to comprehensively consider readers' personalized materials and other potential information resources to help readers find the resources they want. How to accurately classify and retrieve book information from library book resources and improve the utilization rate of library resources is a difficult problem to be solved in recent stage. Aiming at this research, this paper puts forward a library book information classification and retrieval method based on data mining. After preprocessing the book information, Apriori algorithm in data mining is used to deeply manage the information, and decision tree algorithm is used to classify the information. The concept map of book information is constructed by using the concepts in book information, and the book information retrieval framework is constructed. Using Markov network model to realize the classification and retrieval of book information. The test results of classified retrieval method show that the accuracy of book information retrieval is significantly improved when the

proposed method is applied. Through the research and application of this method, it promotes the transformation from the traditional library service concept to the personalized information service concept. In the future development process, according to the personalized characteristics and personal needs of readers, we will provide more convenient book information retrieval and push services, and constantly improve the speed of book information retrieval.

References

1. Zhao, W.: Design and implementation of book retrieval system based on barcode recognition technology. *Mod. Electron. Tech.* **42**(17), 124–128 (2019)
2. Hu, X., Lu, C., Qi, B.: Classification model of internet public opinion information based on SOM neural network. *J. Ordnance Equip. Eng.* **40**(03), 108–111 (2019)
3. Dong, G., Xia, W.: Fast retrieval method for self-integrated information of books based on top-k query algorithm. *J. Jilin Univ. (Sci. Edn.)* **58**(03), 666–670 (2020)
4. Fan, H..M.: Simulation of accurate retrieval of massive books in library. *Comput. Simul.* **36**(06), 337–340 (2019)
5. He, M.: Design of library mass information classification and retrieval system based on data mining. *China Comput. Commun.* **32**(11), 153–154 (2020)
6. Zhang, H.-T., Zhang, X.-H., Wei, P., et al.: Research progress on network user information retrieval behavior. *Inf. Sci.* **38**(05), 169–176 (2020)
7. Nan, C.-J.: Research on the application of data mining in library user resource management. *J. Nanyang Inst. Technol.* **12**(05), 125–128 (2020)
8. Chugh, A., Sharma, V.K., Kumar, S., et al.: Spider monkey crow optimization algorithm with deep learning for sentiment classification and information retrieval. *IEEE Access* **9**, 1 (2021)
9. Maher, A., Supreethi, K.P.: COVID-19 diagnostic system using medical image classification and retrieval: a novel method for image analysis. *Comput. J.* **8**, 8 (2021)
10. Jasm, D.A., Hamad, M.M., Alrawi, A.: A survey paper on image mining techniques and classification brain tumor. *J. Phys. Conf. Ser.* **1804**(1), 012110 (8pp) (2021)
11. Kim, D., Kang, H.G., Bae, K., et al.: An artificial intelligence-enabled industry classification and its interpretation. *Internet Res. Electron. Network. Appl. Policy* **32**(2), 406–424 (2022)
12. Karmakar, P., Teng, S.W., Lu, G., et al.: An enhancement to the spatial pyramid matching for image classification and retrieval. *IEEE Access* **8**, 22463–22472 (2020)