Personalized requirements oriented data mining and implementation for college libraries

Zhao Xiang[1*], Zheng Hao[2]

[1]*Shandong University of Finance and Economics Library, Shandong Jinan 250014, China;*

[2]*School of Business Administration, Shandong University of Finance and Economics, Shandong Jinan 250014*

Received 1 October 2014, www.cmnt.lv

Abstract

By the study of data mining technologies and systematic theory of personalized service, this paper introduces data mining to the college libraries to provide personalized requirements oriented service for readers. It analyses the demand for college library database at the phase of data mining, and explains the necessary for modelling in theory. Then the structure of model is designed. During data mining, we adopt ClassIndex number to establish index distribution trees. We compute the interest distance among the readers according to the depth of ClassIndex numbers of books. Inspired by Cruskal's method, we use minimum spanning tree to establish a weighted undirected liaison graph to perform clustering analysis for the readers. In the association rule mining, by the clustering of readers' borrowing information we find the results are ideal. So we can offer corresponding rules pattern to provide personalized recommendation service for readers.

Keywords: personalized service, data miming, interest, clustering, apriori

1 Introduction

At present, the information stored in college library is increasing fast and the readers become more and more refined, personalized and complicated for the requirements of information and literature types. In order to provide effective management in college library and better service for readers, the personalized information service needs to be developed and this is a current service to be vigorously developed in each college library. Personalized service in college libraries needs information support such as readers' borrowing interest, borrowing history and implicit relationship among the books [1, 2]. However, this information can be obtained by analysing and data mining from daily data of library. Therefore, it shows essential significance to study the application of data mining technology in college library system [3]. Since data mining is an emerging technology in computer field and information processing. It has become one of the most forward research directions in database and information decision.

For the data mining technology in the application of library management, data mining provides decision support for library management and there have been many current related researches. In the United States, many college library staff and some backbone researchers in computer department of have developed some practical implementations. For instance, Bouzid [4] adopts knowledge discovery to establish Arrowsmith software system which is used to process literature resources in database. He obtains great achievements in medical scientific research, which attracts attentions in information science. At present, domestic application researches of data mining in library system are mainly focused by scientific researchers of college. Gong [5] studies data mining of user behaviour. He explores the social relationship of readers and analyses the classification of readers in detail. He also studies the improvement of association rules algorithm and develops corresponding web environmental product. There are also some researchers who focus on theoretical researches in data mining application of digital library [6, 7].

By the study of data mining technology and personalized service system theory, this paper introduces data mining to reader service of college library and perform personalized books recommendation service. After in-depth requirement analysis and borrowing process analysis of college library database, this paper discusses the necessity to establish data mining model in theory and make structure design for the model. At the phase of data mining, we comprehensively applies index category of books in *Chinese Library Classification* to establish distribution tree of books index, and calculate the similarity among books to further obtain the distance reflecting approximate degree of interest preference between two readers. This paper also integrates a Cruskal-based K-means algorithm to perform effective group classification for readers in library, to provide association rules analysis for readers group who have different interest preferences and reading habits. In implementation, this paper takes these data as specific mining objective. It applies data mining technology and studies specific

[*] ***Corresponding author's*** e-mail: haozheng20041@163.com

process of data mining analysis in personalized service of library. The mining analysis has two aspects: one is adopting clustering analysis to effectively classify the readers in library and summarizing the characteristics of different clustering groups; the other is adopting association rules mining based on cluster analysis for further mining. It can acquire readers' association content of library books resources to find out various books with strong association. It can also actively offer relevant information of borrowed books and actively recommend books to readers in multiple ways.

2 Personalized service system of library based on data mining

The personalized service is based on users' information such as usage behaviour, habit, hobby, characteristics and specific requirements. It offers users a kind of service satisfying their information resources and functions of personalized requirements. It comprehensively considers readers' personal characteristics and special information requirement to provide personalized information environment for readers, so it is the most sufficient expression of the first concept. The personalized information service has explicit feedback and implicit feedback, according to active requirement information offered by users [8]. Explicit feedback and implicit feedback are mainly based whether they need the requirements of reader. Explicit feedback and implicit feedback have respective advantages and disadvantages so they need to be combined to be applied in practical personalized service of college library. It means that they not only need close depend on readers' active interest feedback, but they should make use of relevant technologies in data mining. Then it used to analyse the borrowing pattern and to find out their internal characteristics. Therefore, these two methods are combined to provide personalized service.

In our system, we take into account data warehouse, data mining and online analytical process for data mining including readers, resources and access to resources. They aim to discover the relationship between readers' interest and resources for readers to develop their personalized services. The structure of personalized service system in digital library is shown as Figure 1.

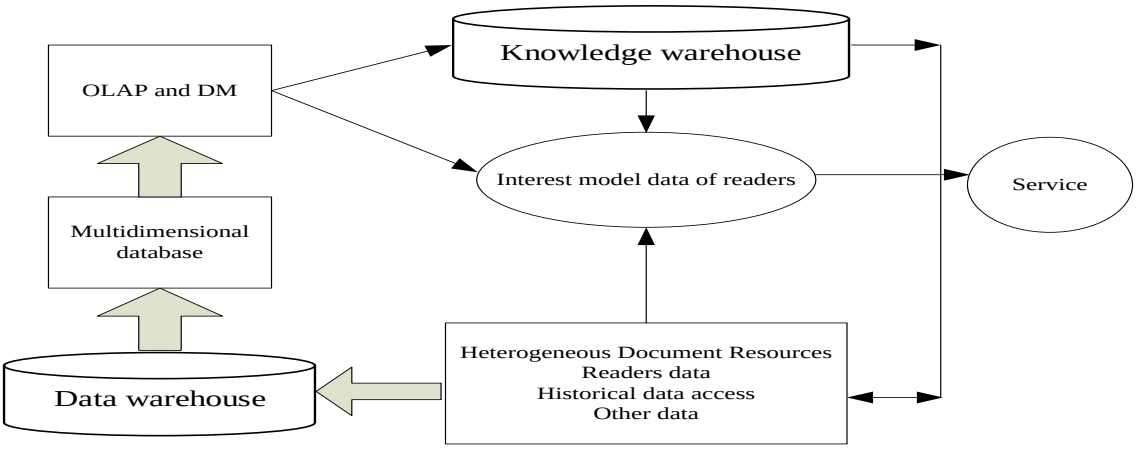


FIGURE 1 Structure of personalized service in digital library

Data mining-based personalized service system in library has two major functions: one is the mining function which is to perform association rules mining and cluster analysis for system data and to seek implicit law; the other is personalized service function which applies mining results to personalized service of the library.

The system has three functional modules:

1) Data processing module. It extracts related data from original library system database, makes pretreatment on some original data and stores the processed data in data mining base.

2) Data mining module. It has obtained required data set from data processing module. On data mining module, it is based on the obtained data set from data processing to apply association rules and cluster analysis algorithm to perform data mining and preserve mining results for preservation.

3) Personalized service module. Personalized recommendation is based on association rules and mining results of cluster analysis. Meanwhile, for readers' requirements, it provides personalized recommendation service for different readers and actively recommends related books for readers. The personalized information push is based on users' customization pushes information via e-mail.

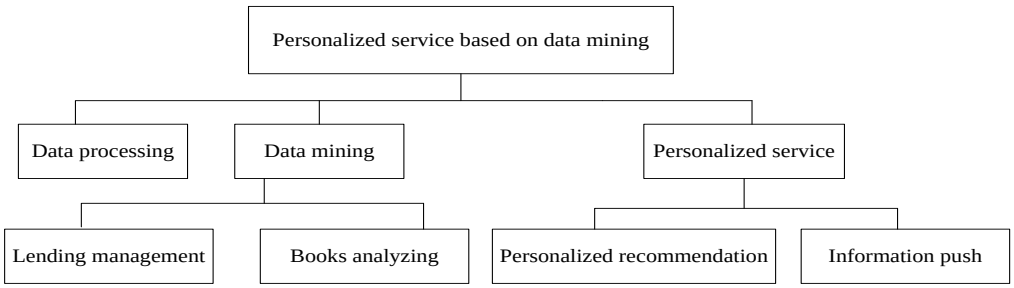


FIGURE 2 Function modules of mining system

3 Mining method

3.1 INTEREST DISTANCE

The reader groups have professional attributes and distinctive identity. They are classified into different groups according to their identities, classes, specialities, departments. These classification attributes are important basis to reflect the demand of readers. The speciality settings of colleges are various and complicated [9]. There also exists intersection and coincidence among the specialites. So the demand of groups may not be reflected in detail or reliably. By the analysis on historical data of borrowers we find it can better reflect the interest and reading habits of readers. Therefore, this paper adopts the method of clustering analysis to specialize the reader groups, based on the borrowing data. So we can clearly get real demand of readers according to the divisions to provide basis for personalized recommendation service.

We compute the similarity among the books according to their classification number. Then the distance between the readers can be acquired. Before this, this paper will classify the indexes of books belonging to each class, according to the classification number. So we establish a book classification index distribution tree. Its classification of root will be more meticulous than top and all the classification number are leaf nodes. The part of classification index distribution tree is shown in the following figure. Root node is 0 layers, and the sequent are the first layer, the second layer, etc.

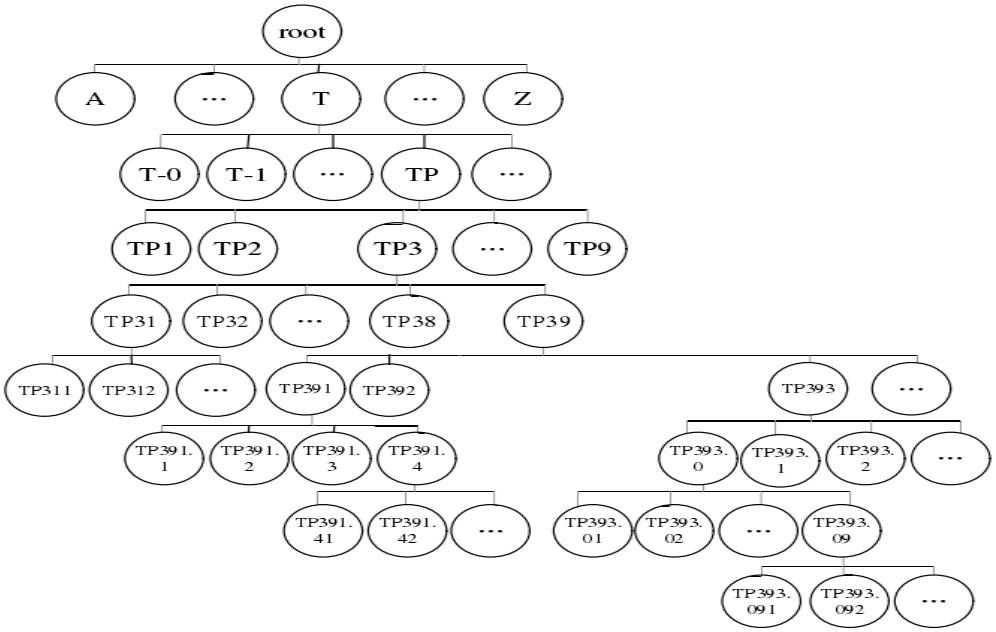


FIGURE 3 Distribution tree of book classification index

According to each data table in database, the lending records of each reader can be acquired by arrangement. It includes the attributes like ReaderId, BarCode, ClassIndex, etc. [10]. We use the classification number of books to compute the similarity of between the books of reader S_1 and S_2. First, when the classification numbers of two users are the same, it denotes that the borrowed books are the same or the books which have the same title. So the lending time should be taken into account when the similarity is computed. Generally the reader will keep one book for some time, the length of which reflects the preference of reader to the book. In this paper it is believed

longer time denotes a higher degree of preference. Then we consider the ration of average lending time for one book and that of all the books, so as to get the book similarity reflecting the interest of readers.

$$Sim(x,y)=\frac{\frac{1}{2}(L_{1x},L_{2y)}}{\frac{1}{m+n}\left(\sum_{j=1}^{m}L_{1j}+\sum_{j}^{n}L_{2j}\right)}. \quad (1)$$

x,y denote two books that have the same ClassIndex; L_{1x},L_{2y} denote the lending time for book x,y for reader S_1 and S_2 ; m and n denote the behaviour times of book borrowing; L_{1j},L_{2j} denote the lending time of S_1 and S_2 . So the denominator denotes average lending time of all books borrowed by S_1 and S_2 .

When the ClassIndex number of borrowed books are not the same, we should find the same parts of the class number first, that is, the nearest common ancestor. It explains the class of two books belonging to. Based on the index distribution tree given above, we search the depth of the class of two class numbers and their nearest common ancestor. Then according to Equation (2) the similarity is computed as

$$Sim(a,b)=\frac{D(Nca(a,b))}{(D(a)+1)+(D(b)+1)-D(Nca(a,b))}, \quad (2)$$

a,b denote the books and $Nca(a,b)$ denotes the nearest common ancestor of two class numbers. $D(a)$ and $D(b)$ are the depths belonging to the ClassIndex distribution tree of a,b . The molecular is the depth of ClassIndex distribution tree of their nearest common ancestor.

TABLE 1 Lending records of S_1 and S_2

ID	Barcode	Index number	Borrowing days
S_1	90978536	H314.3/31	25
S_1	94045571	H319.6/224	20
S_1	96854478	TP313.1/1230	8
S_1	90894134	TP393.08/1987	25
S_2	70033751	TP312/1005	10
S_2	90588740	H314.3/38	35
S_2	90841251	TP393.02/1835	20
S_2	90856234	TP390.41/217	15

Each similarity of book is acquired by Equations (1) and (2). Then we can establish a book similarity matrix. The reader who borrows fewer books is taken as rows, and the other one is taken as lines. For example, we compute the similarities of books listed in Table 1, to acquire the similarity matrix as follows. The attributes of S_1 is taken as rows and the attributes of S_2 is taken as lines.

	90978536	94045571	96854478	90894134
70033751	1.440	0.333	0	0
90588740	0	0	0.778	0.273
90841251	0	0	0.308	0.25
90856234	0.333	1.296	0	0

For the similarity matrix we will find the minimum similarity of each row. Thus we get the book similarity reflecting the same interest of two readers as: $Sim_1=1.440$, $Sim_2=1.296$, $Sim_3=0.778$, $Sim_4=0.8$, $Sim_5 0.714$. $k=5$ means the similarity numbers among the borrowed books of two readers, that is, rows number of matrix. According to Equation (3), the interest distance of two readers can be acquired.

$$D\ D(S_1,S_2)=\frac{1}{1+\sqrt{\sum_{i=1}^{k}Sim_i^2}}=0.31 \quad (3)$$

3.2 IMPROVED METHOD BASED ON CRUSKAL CLUSTERING AND K-MEANS ALGORITHM

Learning from the ideas of Cruskal's minimum spanning tree, we can establish a maximum-cost forest or several loops to optimize K-means clustering algorithm [11]. Then an improved method based on Cruskal and K-means algorithm is proposed in this paper. It will create K cluster and initial clustering centres automatically to overcome the defect that the value of K and initial clustering centre need artificial determination randomly. Since different value of K and initial clustering centre would cause uncertainty of the nearest neighbours. The following example studies the clustering of book similarity and interest distance mentioned above to explain the process of our method:

Algorithm: Readers clustering.

Input: Reader set $R(m)=\{r_1,r_2,...,r_m\}$

Output: The cluster $U_1,U_2,...,U_k$ *and* K *initial clustering centres* $c_1,c_2,...,c_k$.

Step 1: For $\forall u,v\in U$ they are set as vertices and the connections between r,v is (r,v) . We adopt Equations (1) and (2) to compute the distance among the vertices, that is, the weight of edge $w=D(u,v)$. So the readers set and the interest distances are described as a weighted undirected liaison graph $G=(V,E)$. $V=U$ and $E=\{((r,v)w\,|\,r,v)\in U,w=D(r,v)\}$ are the set of vertices and edges. Its initialized status is a disconnected graph $T=U=(V,\{\})$ with m vertices, and each vertex makes a connected component by itself.

Step 2: We select the edges from the weights of E descending. If the vertex of edge lies in different connected components of T , the two components will be connected

as one. Then an edge with the maxima weight is chosen until the vertexes of one edge emerge on the same connected component of F. If it joins T it will cause a maxima loop. These vertices are added to the same set U_i and be deleted in T. This process will be continued until the vertexes of the last edge emerge in the last connected component of F.

Step 3: Repeat step 2 until all the edges construct the forest with the maximum price, that is, all the vertices are distributed to K uncrossed set to form K cluster U_i, $i \in [1, K]$. We compute the mean of U_i to initialize K clustering centres $c_i = \sum_{r \in U_i} \frac{r}{|U_i|}$, where $|U_i|$ is the number of users in U_i.

Step 4: for each user $r \in U$, we adopts above equations to compute the distance to clustering centre. r is reclassified to the clustering U_i nearest to c_i, until all the readers are assigned to the clusters.

Step 5: Recalculate the means of clusters and assign the value to all clustering centres, that is, $c_i = \sum_{i=1}^{K} \sum_{u \in U} |u - c_i|^2$.

Step 6: Repeat step 4 and 5 until the square error criterion function $E = \sum_{i=1}^{K} \sum_{u \in U} |u - c_i|^2$ is convergent, that is, c_i is unchanged.

3.3 ASSOCIATION ANALYSIS

The association analysis focuses on the books borrowed by readers and Apriori algorithm can be used in this paper [12]. On one hand, the association analysis is used for the borrowed books of the whole readers; on the other hand, we can make a conclusion by association analysis of the active readers obtained by clustering. The mining algorithm has two main steps:

A. Get all the frequent sets of things set D satisfying the minimum support degree;

B. Use the frequent sets to create all the association rules satisfying the minimum creditable credibility.

The procedures in detail:

1) Each item is the member of candidate first-item C_1. Scan database D and count the times of occurrence of each item.

2) Determine the frequent first-item set F_1 according to the count of the minimum thing;

3) Execute $F_1 \otimes F_1$ to create candidate second-item C_2 and the frequent second-item F_2 is composed of C_2. Then scan D to compute their support degree;

4) Repeat above operations until the final frequent set is acquired;

5) Output $L = \bigcup_{k=1}^{n} L_k$.

We transform the clustering data into the data for association analysis and import them to file *clementine*, using Apriori algorithm for analysis. The minimum support of association is set first. According to experience, its value $S \in [0.3, 0.6]$. Then we get the association results under the support 0.3, 0.4, 0.5 and 0.6 respectively after relative operation. The direction of input data is also set based on the demand of Apriori. The imported data are taken both as input data and output data, that is, they are treated as bidirectional data. Some data value are specified in advance, for the data may has only one true value or false value, cause error during the process of algorithm. Thus sometimes we should specify the true value or false value manually. The types of data are all specified as "signed". If the value of s is too small, that will lead to many association results to 80% of the users. It cannot find effective pattern and lose expected effect of association rules; if the value of s is too large, that may cause the association rules lose the effect and no operation result. After continuous debug we determine the value of s as 0.5. The associated results are imported to database finally.

4 Implementation

Now, there are 418642 readers' lending records of one college library in 2013 are taken as objective and MSSQLServer 2008 is taken as the basic database [13]. It pretreats the library data first. Then it uses attribute computation method and mining algorithm of previous sections to perform clustering analysis on readers and association mining on books.

By the analysis on books management system database in college libraries, we find that the data tables related to this work are: basic information table of books (Book information), borrowing information table of books (Lending records), readers' information table (Reader information) and books copy table (Book cope). These data tables store original information such as book name, book classification, reader information, lending records, etc. They are the basis of readers to perform clustering analysis and data mining. The principle information and relationship of these tables are shown as Figure 4. Lending data from library is Access file in the form of log so the data which is stored in dispersion monthly as a unit in the library. Then, data introduction and extraction are performed in SQLServer 2008 and data is pretreated before data mining, which can improve the mining efficiency.

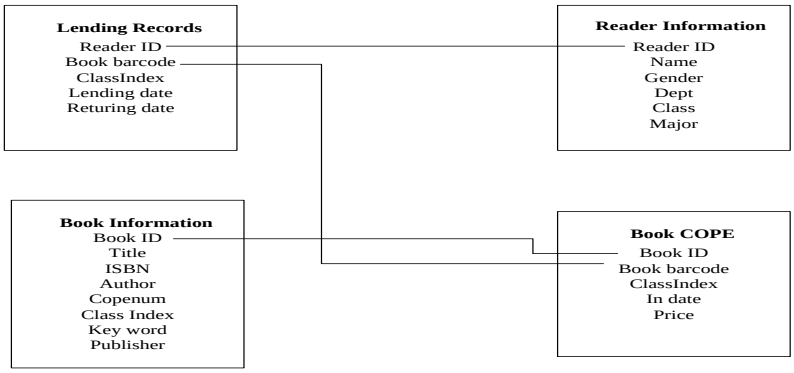


FIGURE 4 Relation graphs of tables

4.1 READERS ORIENTED CLUSTERING ANALYSIS

In the database of college library management system, there are amounts of data information related to readers. The lending record of each reader contains the following information: readers' card number (ReaderId), grade (Grade), type (ReaderType), book information (BarCode, ClassIndex) and borrowing condition of each book, etc. The information of different attribute is used as characteristic research target to perform clustering analysis, so various representative clusters can be obtained. This paper makes cluster analysis from books category, subdivides readers in college library into different groups and performs targeted association mining.

First, needed information is extracted from transaction database and various attribute information situations, which only needs reader ID, book barcode, class index and borrowing time. The consideration in this paper is just each reader's borrowing study on each book.

Secondly, according to previously improved clustering algorithm to cluster the data, after continuous adjustment and optimization during cluster, the finally obtained optimal cluster number is 7 classifications.

Finally, based on obtained clustering group results, the main attributes of data in each cluster are analysed. In order to easily understand specific situation of clustering, a classification profile in Figure 5 is used to analyse readers' interest.

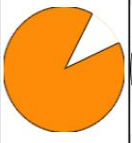
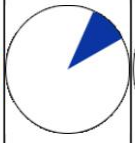
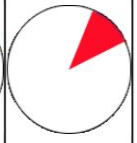
	Clustering A			Clustering B			Clustering C			Necessity
										
TN710-43/W313_TF	0	77	95.06%	0	9	100%	0	9	100%	Unimportant
	1	4	4.94%	1	0	0.00%	1	0	0.00%	
0224-39/447-TF	0	81	100%	0	9	100%	0	9	100%	Unimportant
	1	0	0.00%	1	0	0.00%	1	0	0.00%	
TP312C/733_TF	0	78	96.36%	0	9	100%	0	9	100%	Unimportant
	1	3	3.70%	1	0	0.00%	1	0	0.00%	
TP312VG/525:8_TF	0	75	92.58%	0	9	100%	0	9	100%	Unimportant
	1	6	7.41%	1	0	0.00%	1	0	0.00%	
TN015-43/P791_TF	0	81	95.06%	0	9	100%	0	9	100%	Unimportant
	1	0	4.94%	1	0	0.00%	1	0	0.00%	

FIGURE 5 Clustering profiles

On the whole, reader type is classified into the readers of undergraduates, postgraduates, doctors who are interested in books of industrial technology type (T). This is also the common reading interest of three types of users. Reader interest of each clustering group is simply interpreted and the borrowing tendency in each clustering group is expressed by large ClassIndex in library classification. Readers in group 1 are inclined to books in types of H, I, K and T. The readers in group 2 tend to borrow books in types of B, J, O, Q and T. The readers in group 3 tend to borrow books in types of B, C and P. The readers in group 4 tend to borrow books types of D, N, U, V and T. The readers in group 5 mainly tend to borrow books types of F, G, K, N, R and T. The readers in group 6 mainly tend to borrow books types of A, B, C, D, E, H, I, K, N, O, P, Q, R and S. The readers in group 7 mainly tend to borrow books in types of H, I, J, K, O and Z. The final clustering table is obtained by clustering analysis and the following table shows part of the results:

TABLE 2 Clustering results of readers

Reader ID	Clustering ID	Reader ID	Clustering ID
11031008	C1	10021213	C5
11951071	C4	11091289	C7
11321185	C1	13481103	C3
11341221	C3	16221201	C6
11361223	C6	11871345	C6
10341301	C6	11381401	C7
11361090	C3	16591274	C2
11261405	C6	15341258	C4
10971204	C5	10221130	C5
……	……	……	……

Above mining is used to balance readers' requirement from the perspective of borrowing quantity and it reflects the behaviours characteristics in aspect of library service group. For the second kind of readers who are produced from clustering results, since their borrowing requirement quantity is very large. It can be considered to adjust their maximal volume number of their borrowing circulation books rather than to adopt the same one standard to treat readers with different requirements in unitary form, to provide better service for readers. On the other hand, for those more active readers, we can further mine their borrowing interest to offer more active recommendation service.

4.2 ASSOCIATION ANALYSIS BASED ON BOOK

This mining aims to calculate association degree size among various books by setting threshold value of support and confidence, to discover potential association among the books. If support degree and confidence of books between A and B is larger than given threshold, it shows that the reader who borrows book A is possible to borrow book B. We recommend book B to the reader who borrows book A and this will improve the utilization ratio of books to some extent. The association rule is realized by following steps. The first step is to find the item sets whose support degree are all larger than the minimal support degree and these item sets are called frequent sets. The second step is finding the expected rules originated from the frequency sets based on the first step. By modelling direct mining, 19 association rules are obtained. This paper will choose the record whose support degree is larger than 2% and the confidence is larger than 15%. Parts of the results are shown as Table 3.

TABLE 3 Clustering results of association rules

Antecedent	Succedent	Support	Confidence
TN710-43/W313_TF	TN710-43/M587_TF	2.713	44.455
TN710-43/M587_TF	TN710-43/W313_TF	3.407	32.598
O224-39/447-TF	O22-42/G577_TF	2.886	31.413
O22-42/G577_TF	O224-39/447_TF	2.886	31.413
TN710-43/W313_TF	TN710-43/B146_TF	2.928	29.102
TP368.1/565_TF	TP368.1/651_TF	3.015	24.288
TN710-43/B146_TF	TN710-43/W313_TF	3.704	23.257
TN015-43/P971_TF	TN710-43/B146_TF	2.929	22.061
TP312C/525.8_TF	TP316:81/L793_TF	2.882	17.92
TP312C/525.8_TF	TP316:81-43/M761_TF	2.929	17.647
TN015-43/P791_TF	TN710-43/W313_TF	2.713	17.460
O157:5-43/B925_TF	O224-39/447_TF	2.885	16.419
TP312VG/58_TF	TP312C/733_TF	3.316	15.584
TP312VG/M761_TF	TP312VG/58_TF	3.316	15.584
TP316:81/L793_TF	TP312VG/525:8_TF	3.316	15.279
TN015-43/P791_TF	TN710-43/W313_TF	3.705	15.221

Based on above association mining results, the result is rational and the confidence is only 44.444% to the highest. The reason is that it may be related to readers' characteristics in universities largely. Different majors in universities have corresponding characteristics literature and readers usually borrow books related to their majors, so it is very hard to provide strong association rules from the whole readers of college.

Therefore, based on cluster mining results above, the reader groups which meet the condition "borrowing quantity is larger than or equal to 80 but smaller than or equal to 120" are derived from one group and they are performed association rules mining again, as is shown in Table 4.

TABLE 4 Clustering results of association rules

Antecedent	Succedent	Support	Confidence
Q22-42/G578	Q224-38/4478	3.345	80
TN710-43/W313	TN710-43/B146	2.786	75
TN710-43/M587	TN710-43/B146 TN710-43/W313	2.005	66.667
TN710-43/M587		2.672	62.5
O224-39/447	O24-42/G577	4.438	61.532
TN710-43/M587	TN015-43/P79	3.012	55.589
TN710-43/W313	TN015-43/P791	5.350	51
TN710-43/W313	TN710-43/M587	2.879	50
TN710-43/B146	TN710-43/M587 TN710-43/W313	2.786	50

From Table 4 it can be seen that there are totally 8 association rules. One of them is that the reader has possibility of 80.0% to continue borrowing book O224-

39/447 after he borrows O24-42/G577. In addition, a rule including strong correlation of three books is discovered. Its confidence reaches 66.667% and book TN710-43/M587 is borrowed. Meanwhile, there is possibility of 66.667% to borrow books TN710-43/B146 and TN710-43/W313. From the results we can see that the result is ideal when the groups with similar characteristics are mined. On the basis of all reading records, the correlation among borrowing interest of different professional literature for their respective majors is very strong. When readers borrow one of those books, we will recommend the rest kinds of books to them. When the readers access the library system, it can be taken as page recommendation, offline e-mail or short message recommendation. Meanwhile, it provides reference information for purchasing books of managers.

5 Conclusions

By means of clustering analysis on readers, the readers groups with different interests, hobbies and reading habits are acquired. Then, the lending information in different groups will perform association rules mining to acquire correlation information of various books and to develop personalized books recommendation. In addition, the implementation process of data mining technology is used to design the model in personalized information recommendation system. Then, the recommendation system model for readers' personalized information in college library is established. The data pretreatment module and mining module are also studied in detail. This paper describes the process including dataset acquisition, readers' clustering analysis and association rules mining and analysis the mining results. It is found that the correlation of borrowed books resources forms book recommendation pattern with different borrowing preferences, reading habits, etc. So it can be used to develop personalized information recommendation service for each reader.

Acknowledgments

1) Support by National Social Science Fund: Study on the growth strategy of local producer service enterprise which is embedded in the "global manufacturing and service chaos modular network" (13CGL042). 2) Support by Ministry of Education, Humanities and social science research projects: the relations between network capacity, relational embeddedness and local outsourcing enterprise innovation performance under the background of Multi-National Corporation outsourcing their service modules. (12YJC630316)

References

[1] Zhong M, Wang M, Lu R 2005 Research on user's interest of web personalized information service *Journal of Computational Information Systems* **1**(4) 959-65
[2] Xu S S, Cao J 2011 Personalized service recommendation algorithm in service-oriented environment *Computer Integrated Manufacturing Systems CIMS* **17**(11) 2526-31
[3] Niu G 2009 A Review of Studies of the Application of Data Mining in Library *Journal of Modern Information* **29**(1) 128-33
[4] Bouzid Y, Harroud H, Berrado A 2009 A context-aware platform to support mobile users with personalized services *Proceedings of International Conference on Wireless Information Networks and Systems Milan Italy* 153-8
[5] Gong S 2012 Learning user interest model for content-based filtering in personalized recommendation system *International Journal of Digital Content Technology and its Applications* **6**(11) 155-62
[6] Dingyu 2011 Quantitative analysis of research papers on library personalized service in China in 2000-2010 *Proceedings of International Conference on E-Business and E-Government Shanghai China* 6150-3
[7] Taylor K, Corrall S 2007 Personalized service – Changing the role of the government librarian *Journal of Information Science* **33**(3) 298-314
[8] Cao J, Liu M, Cao L 2013 Research on digital library personalized information service model based on agent model *Journal of Digital Information Management* **11**(6) 418-22
[9] Zhang Q S, Wang X Y 2013 Research of personalized information service based on association rules *Advanced Materials Research* 760 1800-3
[10] Yang H 2013 Personalized web service recommendation based on user interest *Applied Mechanics and Materials* **411**(14) 98-101
[11] Cruskal R, Di Gacomo M 2003 Personalized and collaborative digital library capabilities: Responding to the changing nature of scientific research *Science and Technology Libraries* **24**(1) 135-52
[12] Qi M 2007 Application of the RPW method in the assembly line balance of water pump factory *Journal of Guangxi University of Technology* **18**(4) 79-82
[13] He B, Wang Y 2006 Web Personalized Information Recommendation System Based on Data Mining *Computer Engineering and Applications* **3**(1) 178-86

Authors

Xiang Zhao, 1978.05, Jinan County, Shandong Province, P.R. China.

Current position, grades: the associate research librarian of library, Shandong University of Finance and Economics, China.
University studies: MSc of management in Educational Economics and Management from China West Normal University in China.
Scientific interest: library information system, library data mining.
Publications: 15 papers.
Experience: teaching experience of 6 years, 4 scientific research projects.

Hao Zheng, 1974.03, Jinan County, Shandong Province, P.R. China.

Current position, grades: the professor of School of business administration, Shandong University of Finance and Economics, China.
University studies: PhD in business management from Shandong University in China.
Scientific interest: marketing engineering, data mining.
Publications: 30 papers.
Experience: teaching experience of 13 years, 10 scientific research projects.