

Mining Pattern of Reader Community Using Circulation Log

Qinghua Zhang^{1,*}, Ying Zhang²

¹Beijing Agricultural Vocational College, 102442, China

²Beijing Vocational College of Labor and Social Security, 100029, China

*corresponding author

Keywords: Library; Circulation Logs; Pattern of Reader Community; Data Mining

Abstract: This article introduces the current research situation related to community mining and analyzes the existing relevant technologies and algorithms of community mining, the author selects the Apriori algorithm that is suitable for mining of reader community. This article focuses on the application of these algorithms to get the technology model and implementation framework of mining pattern of reader community, proposes the overall technical framework of mining pattern of reader community, and analyzes the implementation process of mining pattern of reader community.

1. Introduction

Mining pattern of reader community based on library circulation log is to mine the data of library circulation log, find different patterns of reader community, predict the reader's reading tendency, and can be used to guide the library's collection construction, which finally can promote the implementation of personalized book recommendations for reader community [1].

2. Library Circulation Logs and Mining of Reader Community

2.1. Library Circulation Logs

Library circulation log has existed for decades, data mining has existed for 20 years, and the reader community has also been long-standing [2]. The combination of library circulation log, data mining and the reader community is based on the library circulation log to mine the reader community, which will play a positive role in the work of the library.

2.2. Definition of Data Mining

There is currently no uniform definition of data mining. The most common definition of it is the process of extracting information and knowledge that people do not know in advance but are potentially useful from large, incomplete, and vague practical application data.

2.3. Meaning of Reader Community

This article proposes to divide the reader community according to the specific situation of borrowing books. Such reader community has certain rules in the distribution of readers' majors, departments, borrowing time, etc [3].

2.4. Level, Methods and Processes of Mining Pattern of Reader Community

Readers borrows different types of books, keep books for different time, and have different habits, so there are different reader communities. (See Figure 1)

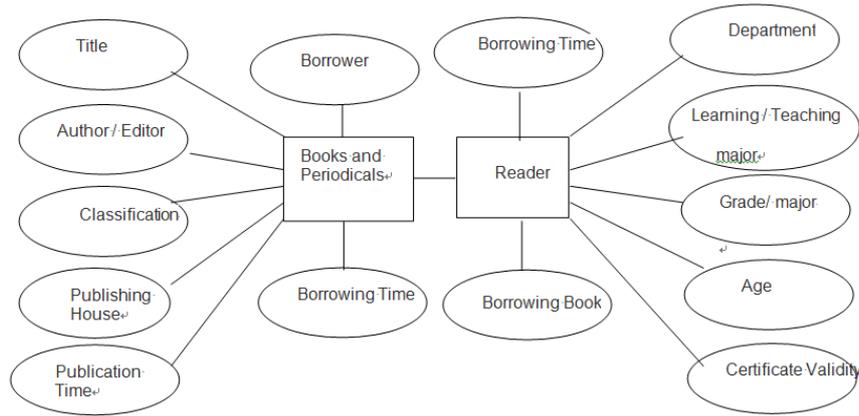


Figure 1 Formation of reader community

(1) The Level of Mining Pattern of Reader Community

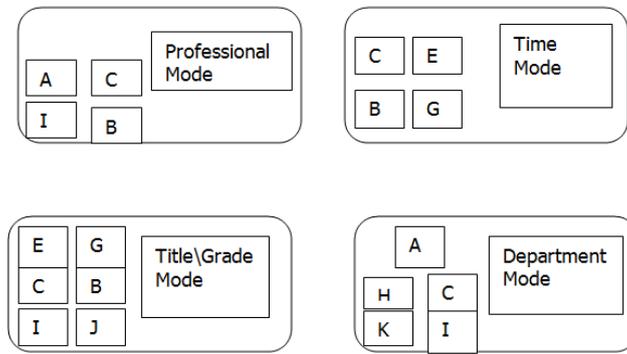


Figure 2

The indicator of the reader community: readers who use more than one type of book of the same kind of books make up the reader community [4]. The pattern of reader communities can be divided into small reader community and large reader community according to the degree of density. There are more similarities in small reader community, and the similarities in large reader community are relatively small.

(2) The Method of Mining Pattern of Reader Community

Mining pattern of reader community is mainly through the analysis of circulation logs, excavating the distribution patterns of different reader communities in majors, professional titles, ages, borrowing time, etc.

2.5. Process of Mining Pattern of Reader Community

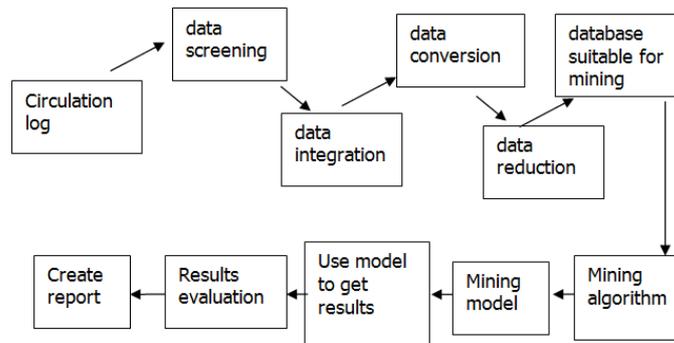


Figure 3

(1) To collect and pre-process of data of circulation logs;

- (2) To extract the features of books and periodicals from the pre-processed database;
- (3) To excavate different reader communities according to the characteristics of the extracted books and periodicals by means of association rule algorithms;
- (4) To analyze the distribution patterns of the reader communities in the fields of majors and departments.
- (5) To direct the library's collection construction, publicity and utilization according to the patterns of reader communities.

3. Technical Framework of Mining Pattern of Reader Community

To mine the patterns of reader communities, we should consider three aspects (see Figure 3): 1. How to extract the characteristics of borrowed books and periodicals; 2. How to construct the patterns of reader communities according to the characteristics of books and periodicals; 3. How to excavate different patterns of reader communities [5]. The technical research of mining reader communities can be considered based on three levels (see Figure 3): 1 Based on data level, what technology can be used for data preparation; 2. Based on the function level, which algorithm can be used for mining pattern of reader communities; 3 Classification of reader communities.

3.1. Applicable Technical Analysis of Mining Pattern of Reader Communities based on the Library Circulation Log

In order to use the library circulation log to excavate the pattern of reader communities, it is necessary to use database technology, data mining software, community mining algorithm, data mining interface and other technologies.

3.2. Overall Technical Framework of Mining Pattern of Reader Community

The overall technical framework of mining pattern of reader community consists of three parts: data preparation module, pattern mining module and result analysis module. The overall technical framework is shown in figure 3-1.

(1) The data preparation module shows that a application system of mining pattern of reader community must be able to handle various types of log data.

(2) The pattern mining module shows that a application system of mining pattern of reader community needs to use the corresponding community mining algorithm to excavate the pattern of reader community.

(3) The result analysis module shows that a application system of mining pattern of reader community can provide corresponding information services for supporting decision for the mined pattern of reader community.

4. Experimental Research on Mining Pattern of Reader Community

In order to study the feasibility of mining pattern of reader community, based on the previous research, the author carried out the application experiment of mining pattern of reader community in the real environment.

4.1. Technical Route of the Experiment

To mine the data of circulation log with the system of mining pattern of reader community, three aspects must be considered:(1) How the system transfers the preprocessing data; (2) How to construct the data mining algorithm of patterns selection; (3) How the system is invoked with other application systems of library.

4.2. Steps of the Experiment

(1) Data preparation

The author has encountered some difficulties in the selection of the data of library circulation log. The circulation log provided by the system can not extract the characteristics of books and

periodicals and readers, and the author needs to extract three databases for merger operations.

(2) Selection of data

The reader's information and book's information in the library circulation log usually output the reader's record number and the bibliographic's record number. The detailed information of the reader and the book can not be found. Therefore, we need to derive the circulation log, the reader list, and the book list and merge them into detailed circulation logs to realize the community mining of the circulation data.

(3) Pre-processing of data

First, we need to improve the information of majors and professional titles of the faculty and staff, and the information of authors of books and periodicals, classification and other information, and delete the borrowing information for the purpose of changing the collection location and the records that readers borrow for relatives.

(4) Data conversion

The Excel format files of "Circulation Log", "Reader Data" and "Bibliographic Data" are respectively imported into "SQL Server 2000" and data consolidation is performed using the "Create View" method.

4.3. Design and Implementation of the System of Mining Pattern of Reader Community

(1) Construction of System Platform

This article verifies the pattern of reader community and uses WEKA, the development software of data mining, Java language, development environment of NetBeans and other technical routes to build a system platform. The key points of the construction process are as follows:

(2) Installation of software

(a) Installation of database software of SQL Server database

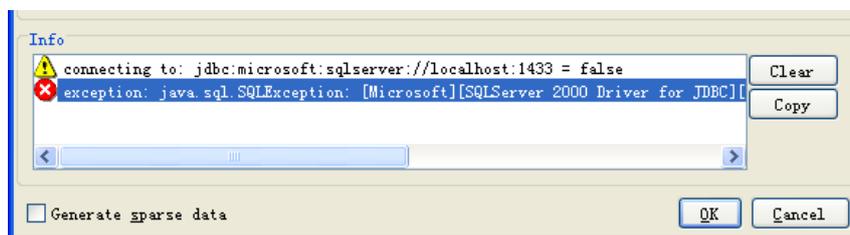


Figure 4 Failed to connect to the database (login failed)

(b) Installation of the environment of JDK and the development environment of NetBeans.

(c) Installation software of WEKA

It should be noted that after installation, you need to modify the files in the weka.jar package.

(3) Testing of Development System Platform of WEKA

Import the data used into the SQL Serve 2000 database; Start the program of weka 3.7(with control), using weka for development;

(4) Mining of Reader Community

According to the pattern mining of the reader communities, the author finds that 95 % of the readers of professional books are the same or adjacent majors, while the readers of literary and leisure books are closer in the aspect of age or grade. There is no obvious rule about the distribution of reader communities in time pattern.

(5) Using the results of mining of reader community to recommend new books

In order to accurately and comprehensively capture the reader's reading preferences and make the recommendations of library's new book more targeted, we can refer the results of mining of reader community to make use of the Apriori algorithm to carry out targeted and personalized recommendations of new book.

5. Conclusion and Discussion

Mining pattern of reader community based on library circulation log has certain application

prospects in library's collection construction and resource's promotion and utilization.

(1) Application in the construction of library collections

Mining different patterns of reader community based on library circulation log can accurately predict readers' reading tendencies, then help us to select readers' favorite books, and book and purchase new books in a targeted manner.

(2) Application in Resource's Publicity and Utilization

Mining different patterns of reader community based on library circulation log can help us to recommend different books according to different readers, enhance the publicity effect of resources, and reduce invalid publicity.

(3) Application in Allocation of Human Resource

Mining different patterns of reader community based on library circulation log can predict the core readers of the library, the book needed, and can help us to rationally allocate interviewers, subject librarians and circulation staff to improve the efficiency of human capital.

References

- [1] Li Jingjing, Shao Peiji, Huang Yiwei. (2004). Research on the Status and Development of Data Mining in China. *Journal of Industrial Engineering and Engineering Management*, no. 3.
- [2] Tan Xu. (2007). Analysis of Reader Community in University Libraries and Targeted Services. *Science and Technology Consulting Herald*, no. 10, pp. 250.
- [3] Yan Wangyan. (2007). Feature Classification and Personalized Service of Readers in University Libraries. *Library Science Research*, no. 8, pp. 77-78.
- [4] Zhang Yinkui, Liao Li, Song Jun. (2003). *Principles of Data Mining*. Beijing: Mechanical Industry Press, CITIC Publishing House, pp. 19-20.
- [5] Huang Ping. (2007). Research Status and Prospects of Small World Network. *Intelligence Magazine*, no. 4, pp. 66.