ISSN : 0974 - 7435 Volume 10 Issue 13

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(13), 2014 [7563-7569]

Research of library management system based on data mining

Lingxing Yang
Qujing Normal University, Qujing Yunnan 655011, (CHINA)

ABSTRACT

In this study, data mining method based on the use of regression analysis to approach from the Natural Sciences Subjects (N), such as eight books to analyze Fudan University's data distribution ratio of the loan amount of all kinds of books, and then the library's books management system optimized so that the number of books in the library is more accurate and targeted distribution, making the construction of great help for the science library. Through data mining, discovery of mathematical sciences and chemical (O) within the library and industrial technology (T), which the number of books to maintain the absolute superiority, appropriately increasing the number of Aerospace's (V) books year by year and updating rate, remained unchanged the purchased number of books of natural Science Subjects (N), the demand for mathematical sciences and Chemistry (O) class books need to do the survey in advance.

KEYWORDS

Data mining; Library; Management system; The assignment of amount of books; Rational allocation of the number of books.

INTRODUCTION

In recent years, businesses are very concerned about "data mining" of this emerging concept, many people need to own a large amount of data into the appropriate information and relevant knowledge. Data mining has been in the telecommunications, securities, banking and many other industries has been an extremely wide range of applications.

With the rapid development of education, each university vigorously accelerates the construction of the library, and optimizing the library management system which is one of the core issues. Among the many books the library management system, there will be a lot of statistical data are generated on every day, these data must be treated reasonably, to be able to provide strong support for the library management system, and for the entire book is also a contributing. Many hidden in complex data information management systems can not directly be found in books; leading too many readers of the information can not be predicted. This study suggests that the current number of small library management system mainly exists some problems, such as retrieves slow, inefficient, less librarians, heavy workload, hard for books' statistical work and the problems can not be updated timely. In order to solve these problems, taking advantage of data mining technology, large amounts of data can be generated in-depth analysis of the library, and then analyzed to provide effective protection for their readers' information. So that the library management system has been optimized. In the present study, data mining in library management system has been indispensable. Digital Library is a digitized strong guarantee on promoting modern education, and digital library is also to achieve digital information carriers, it will be a revolutionary improvement, and will cause the revolutionary change of traditional book.

Literature[1][2] are proposing what is based on data mining, using a very scientific method to establish a very viable model for library management, so that in terms of collections and libraries interview[3] [4] also has a number of attempts to improve the personalization services; In addition, the library can also be based on visual analysis of the needs of readers. Using data mining techniques that can further the library's information resources to improve and optimize services, especially in decision-making in an interview with the book, dramatically improving library services to readers. Literature[5] brings together all kinds of different readers to retrieve the database's data, taking advantage of relevant expertise to determine their property and correlation, and thus the associated data mining is using the principle of these data to quantify, and further analysis, the final preliminary study by empirical data mining techniques and methods, which are used in the construction of Library resource Center.

The purpose of this study is the use of data mining methods, analyzing the borrow information of eight categories' books in one library, and then getting useful information to help optimize the allocation of the number of different types of books and library management system.

THE INTRODUCTION OF DATA MINING

Data mining is from a large number of heterogeneous data through dealing、screening process to obtain useful information. W. J. Frawley's definition of data mining is given by the public recognition - "data mining is from a large number of incomplete, noisy, fuzzy and random practical application data extracted knowledge of interest." A process of knowledge discovery data mining process essentially by a large number of heterogeneous database data filtering dealing,、validation and modeling process, thereby enabling information and knowledge to be able to become an important basis for decision support、guidelines. According to the structure of the main object of study, the data mining is divided into data mining、 text mining,、Web mining three categories in general. In all walks of life frequently used algorithm consists mainly of decision trees, Bayesian, etc., these algorithms have been widely recognized, not only has some practical significance, but also has some common sense in a way.

With the rapid development of education, library management database management system has also been widely used, but led to a library management system contains a lot of data, but very little useful information plight. Hence the need for a new approach to help reader's rational use of large amounts of data libraries, useful information can generally get better service for readers; moreover, emerging data mining techniques happen to be satisfied with this requirement.

For a long period of time, some of the management of the library is to be completed by artificial, low efficiency does not say, but also wasted a lot of manpower and financial resources. So the use of information technology means you can avoid this phenomenon, not only can improve efficiency, but also can serve more readers. The applications of Library computer allows us to see a reduction in the number of employees、but also books、 library、 fewer people queuing. Later, the emergence of a bar code with the continued development of new technologies, it appears to be more convenient for libraries, making efficient management information system has been gotten an unprecedented increase, because the bar code with the book is basically one to one, not only can reduce the error rate of staff, but also making the work efficiency and having been a relatively large increase, while preventing the loss of books.

With the continuous advance, consulting continuous development of cultural information and the era as a new field slowly got people's attention and concern, but also the emergence of the phenomenon of the knowledge economy, knowledge can innovate the wealth, therefore libraries as the knowledge hall has become a very important place gradually as consulting services. However, in some of the relatively large size of the library spaces, because some objective reasons, such as greater mobility of staff、supervision become more difficult and so often there will be a long line readers or books stolen phenomenon. So how to use some of the new technologies and ways to make the library management is an important issue

faced by the parties. Currently, if the RFID technology、 data mining technology and computer software technology combination effectively, it not only can avoid a lot of manpower、 material and financial resources. You can also book-related resources for effective control and management, so that you can expand the number of Library's project services.

Using data mining technology can improve the efficiency of working in a large extent, and can provide efficient services for library management and lending.

(1) Data mining techniques can aggregate some of the data and information within the system library information, processed and analyzed, forming the formation in order, it also can provide timely information for decision-making related to the management staff.

(2) Data mining technology can effectively related objects multidimensional analysis and comparison. It is not only improve the credibility of the decision-making, but also be reasonably related allocation and utilization of resource, and then reaching the optimal allocation of resources associated libraries capacity.

(3) Data mining technology can identify certain rules and models based on past some of the data, and then make their own judgments and predict ground. Basically realized the semi-intelligent form of requirements for the construction of digital library has great significance.

Classification of data mining technology

If you look for a certain classification method based on actual data, according to the classification of the actual data classification induction, to find the essential characteristics of various types, which is called the classification analysis. Typical classification analysis includes decision trees,、neural networks etc. These classification algorithms have been well verified in practice, so the application is very common.

Based on a certain type of data classification feature called clustering, clustering of data can make the difference in the same class as small as possible, in the contrast; there is a big gap between the different categories. If by some technology to find association patterns between data, the technology is called correlation method. The prediction of these methods can deal with the general trend analysis、filtering white noise and so on. For some special instances or data, error analysis is required to use, the process of a derived class can be analyzed with special properties, and then distinguish the difference from the other categories.

Regression of analysis

The regression of analysis is an important method of mathematical statistics, which depart from the sample data to determine two or more variables, the two variables are interdependent. If you are using the regression of analysis to analyze the reliability of a variable, then you can find the main affecting factors, and also can be used to predict the amount of the other according to their numerical size. Multiple regression is to solve the dependent variable and two or more independent variables of the regression, the dependent variable can be expressed in the form of a combination of independent variables; It is to describe that how the dependent variable y represents the independent variables $x_1, x_2 \ldots\ldots x_k$ and the equations of errors ε, involving k (k stands for numbers) independent variables Multiple regression model can be expressed as the following equation:

$$E(y) = b_0 + b_1 x_1 + b_2 x_2 + \cdots\cdots + b_k x_k$$

DATA MODELING

The introduction of model overview

In order to deal with the accuracy and integrity of the data, making the library get reasonable results, the practical problems and data should be set up and analyzed by the data model firstly in this study. The two components are basic data maintenance and borrow data analysis constituting the entire data model, aimed at analyzing the publication of news publishing industry and various types of science and engineering books (in this study included a total of eight categories)

We can deal with the borrow data of Chinese Polytechnic books over the years by taking advantage of the ways of mathematical modeling. Through some the operations of math, we can make the library to analyze whether the borrow data of a class of Polytechnic for readers reasonable or not.

The constitution of data model

(a) Basic data maintenance

The flow of data list and news publishing industry data maintenance is the basis for data maintenance of two parts:

1) an explanation of the flow of data list is as follows: for the purpose of analyzing and providing the support of historical data base, making more convenient query for interviews with staff to process the detailed overview of the basic data. the flow of data list contains borrow, buy、appointments、renewals and other data, and even the number of faculties and students can also be provided. 2), published by the Press and Publication Administration of Press and Publication Industry data maintenance. The number of books are published in the news publishing industry is reasonably important reference for forecasting next year's expenses, the module is designed to maintain the various disciplines publishing books and the first edition of the proportions of the share and other data in the past.

(b)The fluctuation model of borrowing data

From the time dimension that we can build borrowing data fluctuation model , and then analyze the borrow data of the same subject in the past to the extent of fluctuations and trends, judging the reasonableness of the year borrow data from the time dimension.

In order to analyze the volatility of science and engineering books to borrow data, this study uses probability theory and the theory of mathematical statistic to process mathematical model, which is called the fluctuation model. Based on the various disciplines of borrow data in the past, getting the corresponding expectations and standard deviation, and then get the fluctuation range of "O mathematical sciences and chemistry," such as eight Chinese Polytechnic book lending data, it is facilitating to judge the corresponding data reasonably and scientifically for librarians. TABLE 1 gives the paper of Chinese science and engineering "O mathematical sciences and chemistry" borrow books data.

From the TABLE 1, in 2005, the categories of "O mathematical sciences and chemistry" borrow books data exceeds the upper limit, while in 2009 and 2010 "O mathematical sciences and chemistry" borrow books data below the lower limit, so library management system will recommend librarians take appropriate measures to address and resolve.

TABLE1: Fudan University Library "Mathematical Sciences and Chemistry" and the fluctuation analysis of borrow books calendar data

Year	borrow data	average value	Standard deviation	upper control limit	lower control limit	normal ange
2005	61444	54460	6036	60496	48423	beyond the upper limit
2006	57835	54460	6036	60496	48423	normal ange
2007	58134	54460	6036	60496	48423	normal ange
2008	55080	54460	6036	60496	48423	normal ange
2009	47735	54460	6036	60496	48423	beyond the lower limit
2010	46529	54460	6036	60496	48423	beyond the lower limit

MODEL VALIDATION AND RESULTS PROCESSING

This study is based on the real data at Fudan University Library from 2005 to 2010, the number of faculty and students at Fudan University, publication of news publishing industry as a data source in 2005-2010, through the establishment of appropriate mathematical models, digging out the useful information from a number of practical data, doing a research on the distribution of the number of books in the library management system, and then optimizing the number of books and improving the effectiveness of library management.

The sources of data

This paper collected the following types of data:

(1) An interview with the flow of real data at Fudan University Library from 2005 to 2010, including increasing the amount of books in previous years, loan amount of books, the amount of the reservation and renewals and other data. Fudan University library system through integrated A-LEPH500 the database side for statistical programming statements, and then obtains the corresponding data;

(2) The number of teachers and students may borrow books in Chinese science and engineering faculties, from the "Yearbook of Fudan University in 2010"; (3) Basic overview of the national press and publication industry in 2005-2010, the data from the General Administration of Press and Publication released document.

The establishment of model and the management of data

TABLE 1 is the Chinese science and engineering books' borrow data at Fudan University from 2005 to 2010. The list of data were processed by filtering, mapping, weight setting, the normalization process and several steps such as multiple regression analysis in TABLE 2. In accordance with the requirements of the "Chinese Library Classification" level subject's classification, and cleansing the circulation data for the calendar year, we can draw the classification level of science and engineering data, as shown in TABLE 1 (results accurate to four decimal places bit). Book's borrow data contains renew and reservation data in TABLE 1, and will renew and make an appointment with weights 1: 2 into circulation data for calculation.

Through the Natural Sciences Subjects (N), Mathematical Sciences and Chemistry (O), astronomy, earth science (P), biological sciences (Q), Industrial Technology (T), Transportation (U), aerospace (V), environmental science, safety science (X) and other eight categories of books from 2005 to 2010 in the number of six-year loan analysis, get the data normalized as shown in TABLE 3. Figure 1 shows aerospace (V) to the number varies of borrow books with the change

curve of the year. Figure 2 shows the Natural Science Subjects (N) number of borrowed books of the year with the change curve.

TABLE 1: Fudan University Library Chinese science and engineering paper books lending data in 2005-2010 (Unit: List)

Categories number	2005	2006	2007	2008	2009	2010
Natural Sciences Subjects (N)	1,508	1,785	1,693	1,576	1,413	1,461
Mathematical Sciences and Chemistry (O)	61,444	57,835	58,134	55,080	47,735	46,529
Astronomy, earth science (P)	606	527	785	861	873	791
Biological sciences (Q)	12,593	14,593	15,292	15,219	13,369	13,673
Industrial Technology (T)	59,775	59,458	55,365	49,670	44,853	42,165
Transportation (U)	46	45	49	54	62	66
Aerospace (V)	1,494	1,344	1,991	2,156	1,985	1,740
Tatal	137,523	135,642	133,403	124,774	110,471	106,692

TABLE 2 : Chinese Science and Engineering, Fudan University, Fudan University Library 2005-2010 borrow books on paper data (normalized)

Categories number	2005	2006	2007	2008	2009	2010
Natural Sciences Subjects (N)	0.0110	0.0132	0.0127	0.0126	0.0128	0.0137
Mathematical Sciences and Chemistry (O)	0.4468	0.4264	0.4358	0.4414	0.4321	0.4361
Astronomy, earth science (P)	0.0044	0.0039	0.0059	0.0069	0.0079	0.0074
Biological sciences (Q)	0.0916	0.1076	0.1146	0.1220	0.1210	0.1282
Industrial Technology (T)	0.4347	0.4383	0.4150	0.3981	0.4060	0.3952
Transportation (U)	0.0109	0.0099	0.0149	0.0173	0.0180	0.0163
Aerospace (V)	1	1	1	1	1	1

Several conclusions can be found in TABLE 2 and Figure 1, Figure 2, Figure 3:

(1) Mathematical Sciences and Chemistry (O) and the Industrial Technology (T) types of books has always topped the list of the top two in six years to borrow, and aerospace (V) to borrow books always at the bottom standings; In other word, the professional setting and people's interests in reading are focusing on the first two categories in this school; (2) aerospace (V) showing the increase number of yearly lending trends, indicating that the school teachers and students interested in aerospace aviation yearly; (3) Other categories of books have fluctuations in the amount of loan within six years, so the need for further analysis and processing; (4) Natural Science Subjects (N) books of six years remained basically unchanged, there is no trends of increase or rise and in a relatively stable state; (5) mathematical sciences and Chemistry (O) with books of the year fluctuated trend that does not have a fixed increase or decrease trend, and the trend can not be predicted in a random state.

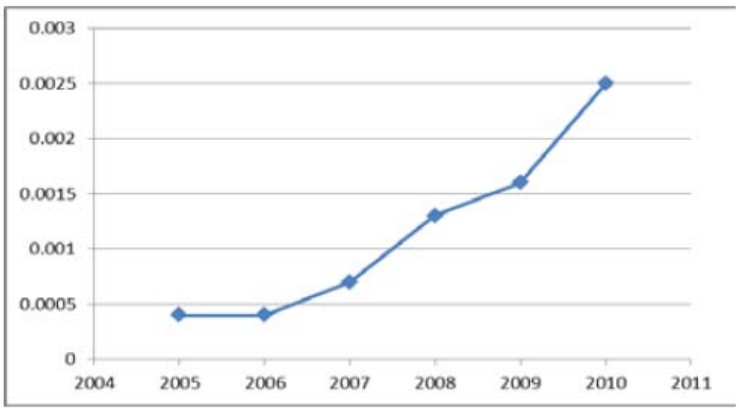

Figure1: Aerospace (V) number of borrowed books of the year with the change curve.

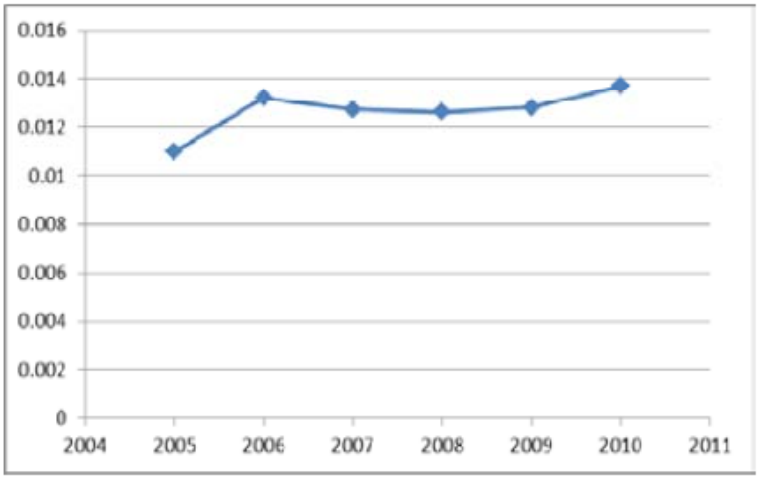

Figure 2: Natural Science Subjects (N) numbers of borrowed books of the year with the change curve

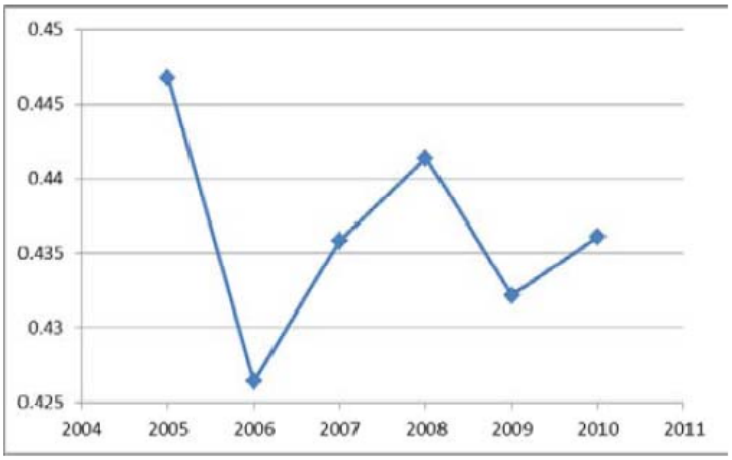

Figure 3: mathematical sciences and chemical (O) numbers of borrowed books of the year with the change curve

But through data mining, we can make a few suggestions for library management system:

1) Always keep the number of Mathematical Sciences and Chemistry (O) and the Industrial Technology (T) Books in absolute advantage, the updating speed should be adapted to the reader's borrow speed; (2) Annually appropriate increase in Aerospace (V) type the number of books and updates speed; (3) For the Natural Sciences Subjects (N) books, we can hold the number of the current configuration, and not deliberately to increase or decrease the number of its purchase, the update rate can maintain its original value; (4) For the mathematical sciences and Chemistry (O) purchase of books, library need do surveys and research for teachers and students' demands and then determine the appropriate amount to purchase and update speed, instead of haphazard, it will lead to a lot of waste of resources.

CONCLUSION

This paper used regression analysis, analysis of trends to borrow a number of Fudan University Natural Science Subjects (N), such as eight books over the year from massive borrowing data. Taking advantage of data mining methods to optimize the library book management system and improve the distribution of the number of books'accuracy and relevance, in favor of the construction of the library science. And after the actual inspection, we can find that the method is indeed feasible, mathematical sciences and chemical (O) and the Industrial Technology (T) books amount to maintain absolute superiority, gradually appropriate increase the number of books in Aerospace (V) type and update speed, keeping unchanged. SCIENCE (N) number of books purchased, the demand for mathematical sciences and Chemistry (O) class books to do ahead of demand surveys are in line with the actual proposal.

REFERENCES

[1] Chuanming Yang; The Application and Research of which is based on Mobile Agent Data Mining in Digital Library, Information Systems, **31(3)**, 436-439 **(2008)**.

[2] Honghui Ma, Ping Zhang; The Optimization and Application of Books Circulated Information of Multilevel Association Rules' Mining Method. Library and Information Service, **52 (7)**, 94-97 **(2008)**.

[3] Yuanzhang Zhang Readers' Demand of Visual Analysis is based on WEKA Click Stream Data. Library Forum, **27(4)**, 67-69 **(2007)**.

[4] Bin Chen, Li Meng; Data Mining and Chinese Medicine College Library Reference Services, Medical Informatics, **10** 1800-1802 **(2007)**.

[5] Yongjin Yu; The Application of Data Mining Books Interview, Changchun Institute of Technology (Social Science Edition), **9(4)**, 27-29 **(2008)**.

[6] Jilong Zhang; Shenqin Yin; Xiangyang Long The Interview Finance Allocation model of Chinese Science and Engineering paper book is based on Data Mining, Journal of Academic Libraries 2013-03, **(2013)**.

[7] Wei Tang; Ruixue Tian Empirical Study based on Data Mining Utilization Books - A Case Study of Liaoning University Library, Agricultural Information Sciences in 2012-11, **(2012)**.

[8] Chaoyang Wang; The Design and Implementation of adaptive Web sites is based on data mining, Xi'an University of Electronic Science and Technology Master's thesis, 2009-03 **(2009)**.