

# Research on the Relationship between Grades and Borrowing Log Based on Data Mining

Sanhong Deng, Jiahang Qin, and Hao Wang

**Abstract**—The traditional library data mining that generally uses only a single data source is likely to lose a number of important factors in the mining results. This paper attempts to overcome the drawback by utilizing multiple data sources, specifically, the library circulation information and grade information of the students from two departments. The use of association rule algorithm to explore the relationships in data preprocessing finds some circulation rules about literature books and language books, etc. Through analyses we draw a conclusion from the perspective of providing a useful reference for management, which is very helpful to guide the students' reading behavior.

**Index Terms**—CHAID, data mining, grade information, library circulation

## I. INTRODUCTION

Data mining technology, as the key to the modernization of libraries, is playing a crucial role in the construction and development of digital libraries as well as in the provision of personalized and knowledge-based services. Data mining in library resource construction is mainly for improving the pertinence of literature resource construction by mining the regular pattern of literature usage, whereas in service improvement it is for pushing the information of interest to users efficiently according to resources available and users' individual interests found based on existing data. Niu points out that although library data mining has been explored enthusiastically in China with a reasonable amount of relative papers published, high quality papers are rarely found when the content is taken into account. Besides, the scholar distribution is unbalanced since authors mostly from the university library system; only tame attempt was done without further exploration within this field for only one author posted more than one papers. All these points suggest that library data mining research in China is still in its infancy, which calls for further in-depth study [1]. Sun et al. insist that the development of various internet technologies makes personalized service possible in university digital libraries and proposed corresponding feasible strategies as well as key technologies by investigating personalized service status of university libraries [2]. Zhang et al. sum up the development trend of personalized information service

of digital libraries that includes user-centered design principles being established; resources and services becoming more dynamic, open, and easier to be reassemble; integrated system of personalized client service being built for users by dynamically filtering and customizing resources; the target of function transferring from information management to knowledge management [3]. Yang points out that personalized information services in China is not optimistic due to the following problems: the library personalized information service is still in an exploratory stage with small-scale and limited number of applications available; systems are simply designed with low application level; the application of simple personalized information service form is fruitless. The three countermeasures proposed were building user-centered service to improve the customer relationship, improving basic information service as well as focusing on information integration and interactivity, selecting the right entry point and gradually expand the personalized information service [4].

Personalized service is emphasized as an important means of improving the quality of library services and information utilization efficiency as well as enabling digital libraries to meet the diverse needs of users. Zhang & Leng insist that services and users are highly targeted and personalized under the Web environment, besides, the ways of four levels to achieve Web personalized service are: external links, resource integration, scenarios focus system and ontology system [5]. Li insist that the development trend of library personalized service can be illustrated in the following aspects: the focus of personalized service of digital libraries will switch to user-oriented organization of information resources from the perspective of resource construction; the system's ease of use and convenience is the inevitable outcome of the development of personalized service from the perspective of user experience; personalized service will be integrated with the users' daily work, study, and life, therefore, more room for development would be possible by improving resource utilization[6]. Chu points out that the role of libraries in society and the services provided are changing rapidly under the new information environment with network as the principal part and libraries must adapt to this change. Therefore user-centric and innovative information organization will bring about new library experience to users [7]. "User-centric" service value of digital libraries has been fully manifested [8]. Due to data ownership issues, the data source of most existing research on library personalized data mining is limited to the library itself without utilizing other data resources for complex data analysis, e.g., Huang only provided a model [9], Liu made a valuable attempt with insufficient analysis [10]. This paper uses the student score data from the teaching sector

Manuscript received July 20, 2012; revised September 1, 2012. This work was supported by the National High Technology Research and Development Program of China (863 Program), No. 2011AA01A206

Sanhong Deng and Hao Wang are with the Information Management School, Nanjing University, Nanjing, 210093, China (e-mail: sanhong@nju.edu.cn; ywhaowang@nju.edu.cn).

Jiahang Qin is with the Library, Nanjing University of Finance and Economics, Nanjing, 210046, China (e-mail: qjh@njue.edu.cn).

combined with the library borrowing log to analyze the borrowing behavior difference among different student groups classified according to scores, aiming at assisting students reading in good order on campus and improving the effect of personalized service.

II. ANALYSIS OF THE RELATIONSHIP BETWEEN STUDENT SCORES AND BORROWING HABITS

The principal part of college learning is self-study through consulting the relevant books in the library. Analysis based on the borrowing records combined with student scores can explore and tell the inner link, contributing to the analysis in view of students' activities in order to guide students in developing good borrowing behavior and make better use of the libraries and books.

We analyze the scores of students majored in Information Management, grade 2006. The first step is pretreatment:

- 1) Since we consider compulsory courses only, the non-compulsory ones are excluded.
- 2) Classification of student scores: more than 90 points for "excellent", 80-90 for "good", 70-80 for "medium", 60-70 for "passing", less than 60 for "poor".

The compulsory courses are divided into five major categories: the first category is computer and information management professional courses including Information Management of Modern, Information Security and Secrecy, Information Storage and Retrieval, Object-Oriented Programming (JAVA), Management Information System, Program Design, Practical Technology of Computer Network, Fundamentals of Computer Applications, Data Structure, Database Management System Applications, Database Theory, Information Analysis and Prediction, Discipline Introduction, Professional Practice. The second category is English language courses including College English 1, College English 2, College English 3, and College English 4. The third category is Humanity and Social courses including College Chinese, Basic Principles of Marxism, Military Theory and Mao Zetong Thought, Deng Xiaoping Theory and "Three Represents" Important Thought, Ideological and Moral Cultivation and Legal Basis, Practical Writing, The Outline of Chinese Modern History. The fourth category is math courses including Probability Theory and Mathematical Statistics, Probability and Statistics (Engineering), Calculus 1, Calculus 2, Linear Algebra. The fifth category is management and economy courses including Principles of Management, Macroeconomics, Microeconomics, Political Economy (Capital).

The excellent rate of a student is the number of subjects graded equal or better than "good" divided by the total number of courses in each category and can be classified into four levels according to the following rules:

Based on the combination of the student score table and the library borrowing table, we analyze 309,623 records using CHAID (Chi-square Automatic Interaction Detector) algorithm for classification to reveal the popular books in each student group of different excellent rate level of each category. Various kinds of books are ranked according to the borrowed number and the top 6 categories are selected:

TABLE I: GRADE LEVEL

Excellent Rate	Level
$\geq 0.0625 < 0.25$	D
$\geq 0.25 < 0.5$	C
$\geq 0.5 < 0.75$	B
$\geq 0.75 \leq 1$	A

TABLE I: CATEGORIES OF MOST BORROWED, GRADE 2006, DEPT. OF IM

CLC	Times of borrowing	CLC Name
TP	2252	Robotization & Computer
I2	1672	Chinese Literature
H3	833	Common Foreign Language
C9	612	Sociology
F2	358	Economic Plan & Management
O1	344	Mathematics

The above six categories of books are used as output variables, whereas course classification and excellent rate level are used as input variables, CHAID is used for classification mining. The following mining results reveal the category of books that are borrowed most by the students with excellent rate level "A" and the number of books in this category borrowed by the students with excellent rate level "D", which can be used in mining the relationship between test scores and borrowing patterns.

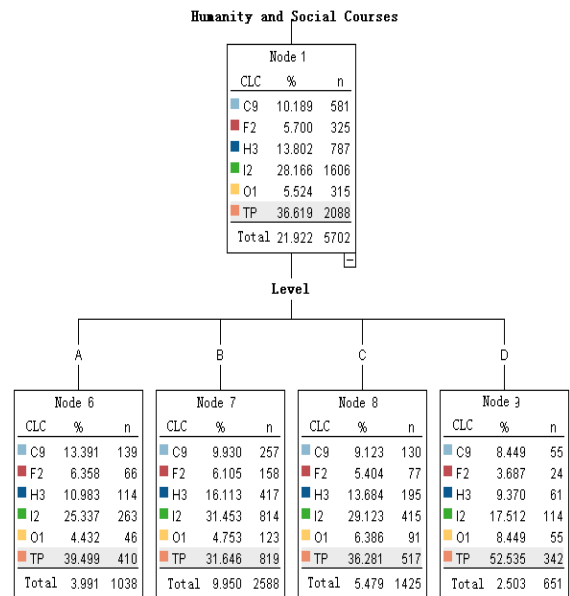


Fig. 1. Relation between humanity courses grade and borrowing behavior.

In Fig. 1, we take courses in Humanity for example, students of level "A" borrowed C9 sociology books for 139 times with the probability of 13.3%, while students of level "D" borrowed sociology books for 55 times with the probability of 8.449%, which is much lower than the former. The relationship between borrowing behavior and student scores of courses belong to other categories is shown in Fig. 2- Fig. 5 below.

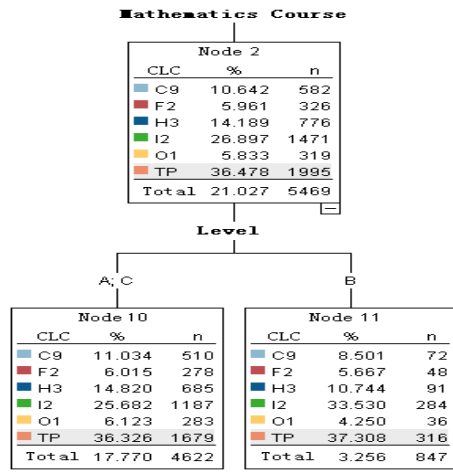


Fig. 2. Relation between mathematics courses grade and borrowing behavior.

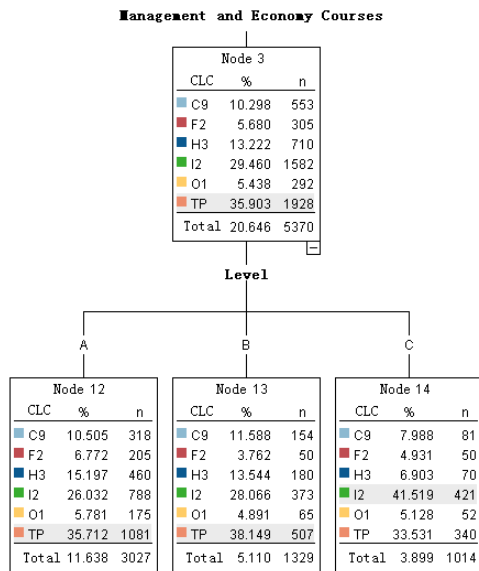


Fig. 3. Relation between management and economy courses grade and borrowing behavior.

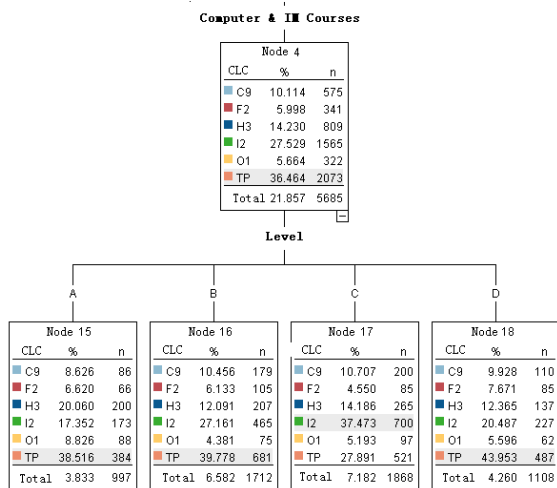


Fig. 4. Relation between computer and IM courses grade and borrowing behavior.

According to Fig. 2 –Fig. 5, the relationship between the scores and borrowing behavior of students in the Department of Information Management can be illustrated as follows:

- 1) No significant difference in borrowing behavior among students with different scores in math related courses;

- 2) In the category of management and economy courses, excellent students borrowed more books in foreign languages and books in the field of management and economy while students of level “C” borrowed more literature books;
- 3) In computer courses, excellent students chose more books in foreign languages while other students borrowed more literature books;
- 4) In foreign language courses, excellent students borrowed more engineering and technical books while other students borrowed more literature books.

As a reference, we also chose the relevant data of students from the Department of Finance for analysis with similar data preprocessing:

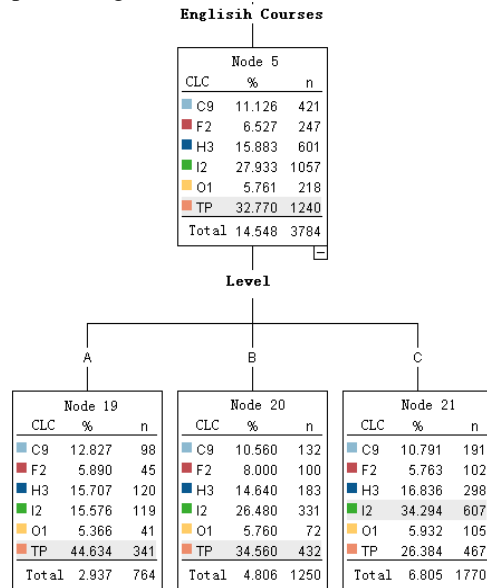


Fig. 5. Relation between english courses grade and borrowing behavior.

The classification of the courses is as follows: the first category is financial courses including Discipline Introduction, International Financial, Financial Accounting, Securities Investment, Corporate Finance, Marketing, Finance, Financial Risk Management, Economic Law, Commercial Bank Management, Professional Practice, Econometrics, Accounting, Financial Markets, Money and Banking; the second category is English language courses including College English 1, College English 2, College English 3, College English 4 ; the third category is Humanity and Social courses including Ideological and Moral Cultivation and Legal Basis, College Chinese, Mao Zedong Thought, Deng Xiaoping Theory and “Three Represents” Important Thought, The Outline of Chinese Modern History, Basic Principles of Marxism, Military Theory; the fourth category is computer technology courses including Database Management System Applications, The Basis of Computer Applications; the fifth category is math courses including Statistics, Linear Algebra, Calculus 1, Probability Theory and Mathematical Statistics, Calculus 2; the sixth category is management and economics courses including Probability and Statistics (Engineering), Linear algebra. The sixth category is management and economy courses including Contemporary Chinese Social Economy, Microeconomics, International Economics, Macroeconomics, Principles of Management, Political Economy (Capital).

The categories of most borrowed books are listed as follows:

TABLE III: CATEGORIES OF MOST BORROWED, GRADE 2006, DEPT. OF FINANCE

CLC	Times of borrowing	CLC Name
I2	3225	Chinese Literature
F8	1979	Finance
H3	1794	Common Foreign Language
F2	768	Economic Plan and Management
K8	663	Biography
F0	467	Economic
O1	449	Mathematics

A total of 253,633 records are involved in the mining and the CHAID mining results of course scores and borrowing behavior can be seen in the Fig. 6 and Fig. 11 below:

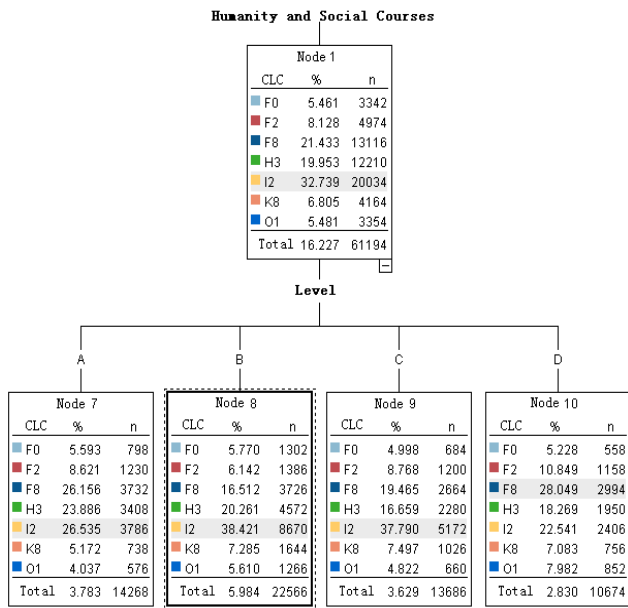


Fig. 6. Relation between humanity courses grade and borrowing behavior.

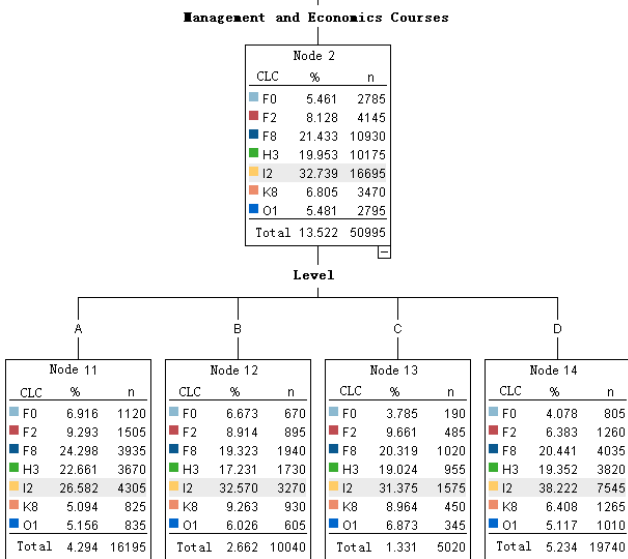


Fig. 7. Relation between management and economics courses grade and borrowing behavior.

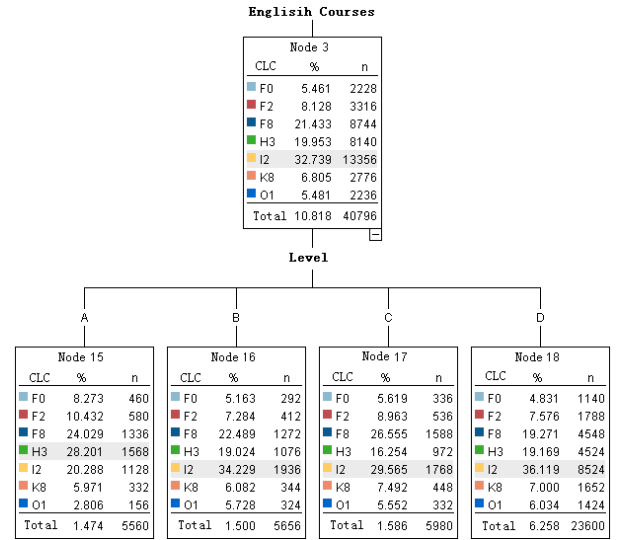


Fig. 8. Relation between English courses grade and borrowing behavior.

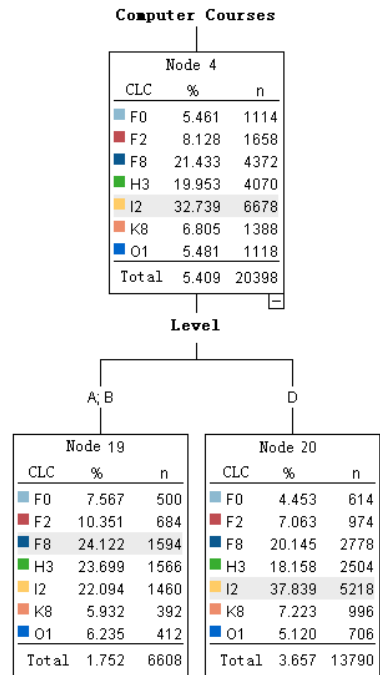


Fig. 9. Relation between computer courses grade and borrowing behavior.

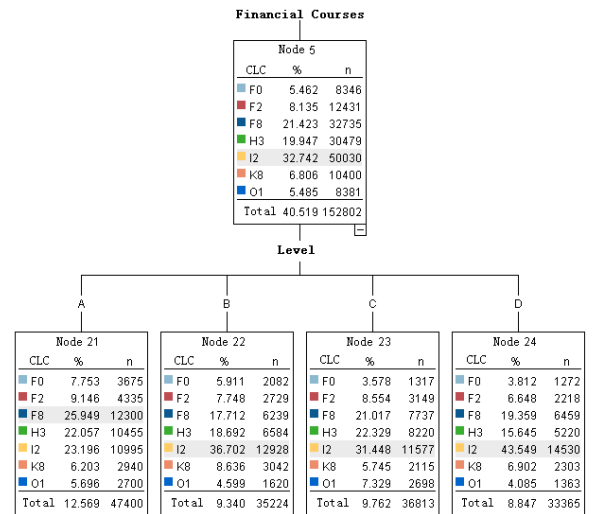


Fig. 10. Relation between financial courses grade and borrowing behavior.

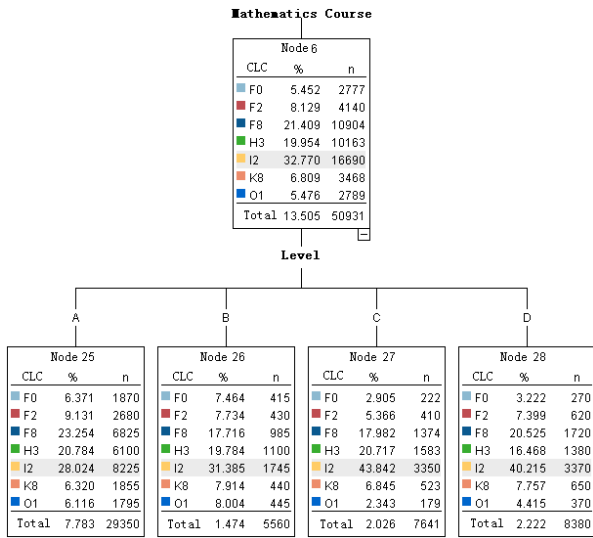


Fig. 11. Relation between mathematics courses grade and borrowing behavior.

The relationship between the course scores and borrowing behavior of students in the Department of Finance can be illustrated as follows:

- 1) In Humanity and Social Science courses, the proportion of books in foreign languages borrowed by excellent students is obviously higher than that borrowed by other students;
- 2) In the category of management and economy courses, excellent students borrowed more books in foreign languages and books in the field of Finance, whereas other students borrowed more literature books;
- 3) In English language courses, excellent students borrowed more books in foreign languages, whereas other students borrowed more literature books;
- 4) In computer courses, the average students borrowed more literature books;
- 5) In Financial professional courses, excellent students borrowed more financial books while other students borrowed more literature books;
- 6) In math related courses, excellent students borrowed more financial books while other students borrowed more literature books.

### III. CONCLUSION

Combined with the basic phenomena in the Department of Information Management, we found the typical phenomena that the average students borrowed more literature books, which taking up much of the time for studying. Besides, since they were not so interested in professional books, insufficient amount of reading affected the improvement of their course results. The borrowing

behavior of these students should be paid more attention to by timely reminding and proper guidance.

### REFERENCE

- [1] G. Y. Niu. "A review of studies of the application of data mining in library," *Journal of Modern Information*, no. 1, pp. 128-130, 2009.
- [2] S. X. Sun, H. Y. Li, and F. C. Wang. "The key technical research about the individual service in college digital library," *Journal of Xichang College*, no. 2, pp. 48-53, 2010.
- [3] L. Zhang and T. Sun. "The evolution and development of the digital library personalized information services," *Library and Information Service*, no. 1, pp. 41-44, 2005.
- [4] B. H. Yang. "Research on the individualization service in digital library," *Modern Information*, no. 7, pp. 94-95, 2006.
- [5] S. L. Zhang and F. H. Leng. "Acquisition of personalized information and realization of personalized services in the web environment," *Journal of Library Science in China*, no. 4, pp. 77-81, 2007.
- [6] Y. H. Li. "The developing trends of personalized service of digital libraries," *Library Tribune*, no. 8, pp. 109-112, 2008.
- [7] J. L. Chu. "Research on user service & service innovation of new information times," *Journal of Library Science in China*, no. 3, pp. 68-70, 2008.
- [8] R. P. Xie. "User service of digital library in ubiquitous knowledge environments," *Researches in Library Science*, no. 10, pp. 35-37, 2008.
- [9] S. D. Huang and Q. M. Chen. "Research on data students grade mining model based on the borrow book information in the library," *Modern Computer*, no. 10, pp. 196-198, 2008.
- [10] L. Liu. "Research on the relationship between grades and borrowing books based on data mining," *Journal of North China Institute of Science and Technology*, no. 10, pp. 117-120, 2010.



**A. P. Sanhong Deng** is from Hubei Province, China. Birthdate: Aug, 1975. PhD of Information Sci. Nanjing University, China, 2003. Work as associate professor in Nanjing University, and research interests on data-mining and knowledge management.



**Dr. Jiahang Qin** is from Jiangsu Province, China. Birthdate: May, 1964. PhD of Information Sci. Nanjing University, China, 2012. Work as research librarian in library of Nanjing University of Finance and Economics and research interests on digital library.



**A. P. Hao Wang** is from Zhejiang Province, China. Birthdate: July, 1981. PhD of Information Sci. Nanjing University, China, 2008. Work as associate professor in Nanjing University, and research interests on data-mining and ontology.