Analysis of User's Behavior on Borrowed Book Record in National Central Library University of Laos

Thongvin SIHACHACK and Lasheng Yu

School of Information Science and Engineering, Central South University, Changsha, Hunan, China
thongvin12@hotmail.com, yulasheng@csu.edu.cn

ABSTRACT

The purpose of this paper is to analyze the student's behavior of borrowing book in order to find out the frequent group of book items using association rule technique based on the data mining technologies. Data mining is a powerful tool for discovering interesting relationships hidden in large data set. Apriori algorithm can be applied on the university's library transaction database to improve resource management to find out the relationship which two groups of books nearby together that user like to borrow. In this paper we collected data set of user's record from Central Library National University of Laos. The number of user's behavior data for training set was 6,615 record of personal log and the analysis is done by WEKA application software. The study shows that the total of 8 rules are with the confidence of 75%, 64%, 61%, 52%, 45%, 37%, 34%, and 23% respectively, the results about user's behaviors can be used to manage the library resource.

Keywords: Data mining, Association rule, user's behavior, Library, Apriori Algorithm.

1   Introduction

Library plays a very important role in any organization as the collection source of knowledge and information for Educational institutions, which made it accessible to a defined community for reference or borrowing. Just as now library always supports on and off line services and library services provide a lot of resources, library's collection may include books such as, e-books, periodicals, newspapers, CDs, document and online database so on, to support knowledge learning for students. Computer and communication technology has been currently used in library service to keep resource and user's information and then to provide the students better services. Using technology to search for useful information from the saved data in order to enhance library service is beneficial. Giving exact resources to the users demonstrates the library's effectiveness in passing knowledge to the users. When we compare the borrowed book behavior of every library, we can find that one user can borrow many books at the same time. This can influence the precision in keeping track of books for future use; therefore, the library might not have enough resources for the users' need. Consequently, it is very important to analyze library users' behavior in order to enhance and oversee Library Management [1].

The Central library was created within the National University of Laos (NUOL), a university in Vientiane, the capital of Laos. Central library is a source of information to students, teachers, professors, researchers as well as general user whose interest aggregates for learning resources and information services to

support learning, teaching, research and publication in the academic community field. And cultural mission of the university is to make the operation more efficient to provide service to most users, the corresponding users' services and the maintenance of information resources and tools available to them.

In the Central library National University of Laos, there are a lot of books which are clarified as many groups of books and there are also many users who borrow those books. To provide more attentive service, the arrangement and display of books in the library can be modified based on the inference out of analysis of the library for instance, when a user wants to borrow some books at the same time the manager in the library will recommend some related books together which will ease the time spent by the user.

Association rule mining which is one of the most popular data mining approaches has been used. Data mining is the analysis of relevant information results and to produce different perspectives to understand more about the user's behavior so as to handle library resources. Besides that, it can also figure out what factors to affect users how to use the library including what categories are related to their interest and how to estimate the future needs [2]. Association rule mining is a procedure which is meant to find frequent patterns, correlations, associations, or causal structures from data sets found in various kinds of databases such as relational databases, transactional databases, and other forms of data repositories. [3]

Various algorithms for association rule mining have been presented over time and the most popular algorithm as we know among them is Apriori algorithm.

Apriori algorithm finds the frequent item set from the sets of items that have minimum support. Subset of a frequent item set must also be a frequent item set. Repetitious finding frequent item sets from 1 to k (k-item set) and use the frequent item set to generate association rule.

We will use association rule data mining technique with Apriori algorithm to improve association rule to find two group of books arranged closely to each other after that we will use WEKA software to process dataset. The remainder of this paper is organized as follows. Section 2 describes related work and data mining. Section 3 discusses about methodology such as association rule mining, Apriori algorithm and program WEKA. Section 4 discusses the Experimental and Resultant, the conclusion is in section 5.

2   Related work

2.1   Data mining.

Data mining is a process to create the knowledge from transactional and large amounts of data stored in database, data warehouse by using statistic procedure and machine learning and training set to get the precise information for future management to be used in decision making, also summarizing the useful information results data analysis and knowledge discovery, which may not be displayed in real time by using a database in the fields such as artificial intelligence that can the information or knowledge research. There are several data mining models, every model include a business equation and results in different forms such as starting from rules then becoming decision trees[4].

3   Methodology

In this paper we use Association rule learning which is a popular method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using some measures of interest, and talking about Apriori algorithm.

3.1 Association rule mining

Association rule mining (ARM) is the one most popular method of data mining for discovering interesting relations between items in the dataset. The concept of strong rules was used by Agarwal et al [5] to find association rules in items sold for large scale transaction data base log by point of borrowed systems in library.

An Association rule can be expressed in the form of, X →Y, Let I = {i_1, i_2, i_3. . .i_m} which is the set of all items available at the store. Where X and Y are two disjoint item sets. X is an antecedent and Y is a consequent, in other words, X implies Y. The main concept of association rules is to examine all possible rules between items and turn them into 'if-then' statements. In this case the 'if' part is X or the antecedent, while the 'then' part is Y or the consequent. There are two basic parameters importance thresholds for measure association rule mining: support and conference [6].

The support of a rule is the probability of the item set{X, Y} that mean the relative of the rule and the confidence of a rule is the conditional probability of Y given X that indicate the exactness of the rule.

$$\text{Support (X, Y)} = \frac{\text{Support count X,Y}}{\text{Total number of transactions in D}} \tag{1}$$

Therefore, confidence is an important measure of the association rules to indicate the strength rules. If the confidence of the association rule is 80%, it means that 80% of the transactions that contain x also contain y, based on users to indicate the specified minimum confidence [7].

$$\text{Confidence (XY)} = \frac{\text{Support (XY)}}{\text{Support (X)}} \tag{2}$$

3.2 Apriori Algorithm

Apriori algorithm is an influential algorithm in the mining of Association rules in a centralized database**.** The main idea of Apriori algorithm is described in the following [8]

	Apriori algorithm
	L_1 = {find_frequent_1-itemsets};
(1)	L_1 = {find_frequent_1-itemsets};
(2)	For (k = 2; L_{k-1} ≠ ∅; k++) do begin
(3)	C_k = apriori_gen (L_{k-1}); // New candidates
(4)	For all transactions t ∈ D do begin
(5)	C_t = subset (C_k, t); // Candidates contained in t
(6)	For all candidates c ∈ C_t do
(7)	c.count ++;
(8)	end
(9)	L_k = {c ∈ C_k	c.count ≥ min_sup}
(10)	End
(11)	Answer = $U_k L_k$;

Figure: 1 Apriori Algorithm

3.3 WEKA

WEKA is full name for Waikato Environment for Knowledge Analysis; the abbreviation of this software also is a unique bird in New Zealand. Interestingly, the main developer of WEKA just comes from New Zealand's University of Waikato. WEKA is a free license for academic use, which does not integer with other systems.

WEKA is completely open software for data mining work provides a unified interface, collects the most classic machine learning algorithm and data preprocessing tools. As a complete knowledge acquisition system, it includes the data preprocessing, classification, clustering, and Association rule, attribute selection, and achieves visualization in a new interactive interface.

3.4 Mining Association rule process

The problem is that the manager in the library does not know user's behavior of borrowing book. It is impossible to predict which group of book any student like to borrow together or are related. Therefore, we collected data set to assess user's behavior record from Central Library University in Laos for find relationship borrowed by design and process data analysis to find association rule.
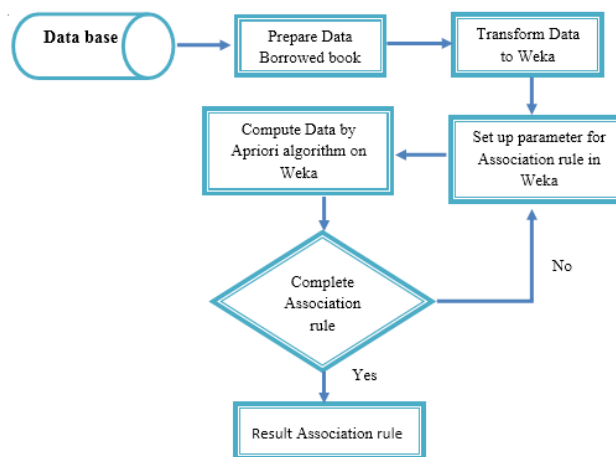

Figure2: Mining association rule process

Figure2: step for preparing data and transforming data to discover association rule by using WEKA software. The core on Figure 2 is computing Data by Apriori algorithm on WEKA. The Apriori algorithm is an influential algorithm for mining frequent item sets. Apriori is a seminal algorithm proposed by R. Agrawal and R. Srikant in 1994 for Boolean association rule. The name of the algorithm is based on the fact the algorithm uses prior knowledge of frequent item set properties.

After transform data from excel to LibofNoul.arff file that can support to WEKA software and arff file was loaded in to WEKA explorer. The associate option in WEKA enables to apply Apriori algorithm on dataset before compute data .We set up Apriori algorithm you can see on figure 4: then WEKA software will compute data for find association rule.

4 Experiment and Analysis

Our paper, we collected data set user's record from Central Library National University in Laos. The number of user's behavior borrowing data was 6,615 the data is composed of personal log, loaning records file from library system as shown in figure 3

List of borrower book						Central library NUOL
ລາຍການຢືມປຶ້ມ						
IDuser	Name	Name of book	Author	IDGroup	Date issue	Date due
ລະຫັດຜູ້ໃຊ້	ຊື່ຜູ້ຢືມ	ຊື່ປຶ້ມ	ຊື່ຜູ້ແຕ່ງ	ໝວດປຶ້ມ	ວັນທີຢືມ	ວັນທີສົ່ງ
S001134	ນ.ທິບພະຈັນ ພົມມະເຫສ	Addvanced spoken English	Jones, daniel	E110	2/5/2016	9/5/2016
S001105	ນ.ແສງຈັນ ວັນນະສີ	ອົບຮົມພາສາອັງກິດດ້ວຍຕົນເອງ	ຍຸທອນ ໄທຍະສອນ	L070	2/5/2016	9/5/2016
S001105	ນ.ແສງຈັນ ວັນນະສີ	High-ED.s Mathematices	ສະໃໝ ຫຼ້າວງນ້ອຍ	M010	2/5/2016	9/5/2016
A011203	ທ. ອິນຕະພອນ ສຸວັນນະສີ	ລາວສຶກສາ 2 Ls202	ບຸນສົມ ສົກທາຍກາດ	C040	2/5/2016	9/5/2016
C002014	ທ. ແກ້ວອິນຄຳ ປານແສງ	speaking Chinese about China(Revised)/T.Lin	Kohli A.B	L070	2/5/2016	9/5/2016
C008001	ທ. ຄຳກົງລີ ປ່າມ້ວ	Landscapes of defence/ jonh R.Gold	Gold, John R.	H050	2/5/2016	9/5/2016
C009011	ທ.ປັກໃຈຄຳ ອິນທະວອນ	PHP MySQL	ສົ່ງຄຳ ບຸນສີໄຕ	Journal050	2/5/2016	9/5/2016
C009011	ທ.ປັກໃຈຄຳ ອິນທະວອນ	the pronunciation of english	Jones, daniel	L070	2/5/2016	9/5/2016
P004011	ອ.ຈ ຕາວັນ ຊຸນບຸໄລ	ແບບເຝິກຫັດ ແລະ ຜົກາໝິກທາງແກ້ບັນຫາໃຈເຄມ ຄູລິດ	ວິໃຈ ທິບມານີ	M030	2/5/2016	9/5/2016
B011041	ທ. ຄູນະພາ ສິປະເທິດ	ເຄມີທົ່ວໄປຫຼ້ມ 1(ສະບັບລວມຮັດ)	ລິດຖາ ມີສຸກ	H650	2/5/2016	9/5/2016
C008012	ນ. ພອນສະຫວັນ ແກ້ວມັນກີ	Intermediate spoken chinese part Two	Delian,Liu	L010	2/5/2016	9/5/2016
C008012	ນ. ພອນສະຫວັນ ແກ້ວມັນກີ	the state stare Ref refrom in Lation America	Education Lpra	R100	2/5/2016	9/5/2016
C008012	ນ. ພອນສະຫວັນ ແກ້ວມັນກີ	PHP MySQL	ວິໃຈ ທິບມານີ	Journal050	2/5/2016	9/5/2016
B041015	ນ.ຈິນທະຈອນ ຈິນທະວົງ	ຄູ່ມີໂປຼແກຼມ Matlab ສະບັບສົມບູນ	ມະນິດສິງ້ອວນສິນ	A070	2/5/2016	9/5/2016
B120035	ນ.ຈິນທະຈອນ ຈິນທະວົງ	ຄູ່ມືການອ່ານວິໄຈຕາມຄຸລິດ 1	ສຸກັນຍາ ສນິດວົງະຍຸທາ	M520	2/5/2016	9/5/2016
B120035	ທ. ກິດສະໜາ ພອນສະຫວັນ	Physical science	Tiller,Bill W	H740	2/5/2016	9/5/2016

Figure 3: user's borrowing record

Figure 3: Data set user's record from database include user name, name of book, authors, group of book, date issue and date due, for this training set we use only group of book to find out which groups of books users like to borrow. In the column group for example the full name of code E110 is English, L070 is Language, M010 is Mathematics, and H050 is History and so on.

After extracting data from the system, then we cut and remove some column that does not relate to this training set for example we cut user name and date issue off after complete screening data. Each attribute is a binary or has a value that can be entered in each column as follows.

- Denote t user borrows book.
- Denote? user null borrows book.

While preparing data set from the record, we proceeded to change data files to files suitable with program WEKA used to analyze the data. Therefore, we convert the original file format of Microsoft Excel 2007 to file .arff that can be open in the program WEKA.
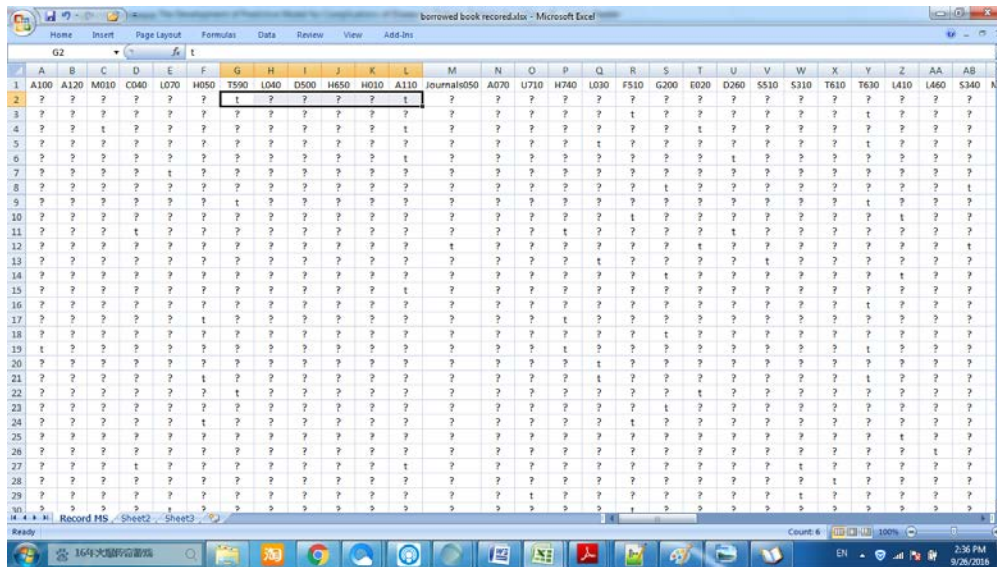
Figure 4: transform record in format of Microsoft excel

Figure 4 shows how to transform data for finding out association rules. This section describes the following: the user borrows book from group T590 and she/her will borrows from group A110 together.

4.1 Setting Parameters

In the WEKA software we will use is the default algorithm selected. However, in order to change the parameters for this run "support and confidence…" we click on the text box of the 'Choose' button then select Apriori look at figure 5:
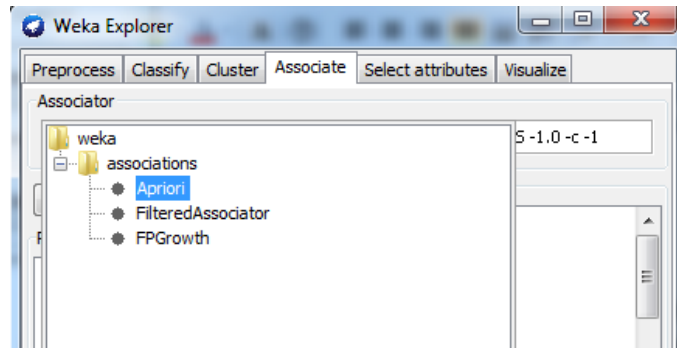

Figure 5: select Apriori algorithm

To find the relation that is the most reliable we can set the parameter associated with Apriori algorithm by program WEKA to find association rule by setting lower bound minimum support 0.03(3%) upper Bound minimum support to 0.1(100%) and setting num. rule is 100 rules look at figure 6.
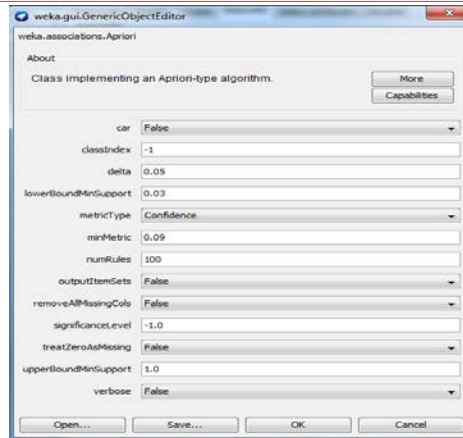
Figure 6: Apriori setting program in WEKA

The data set used for the Training Set was 6,615 user's record from created association rule. This process was found 8 rules showed in figure 7:
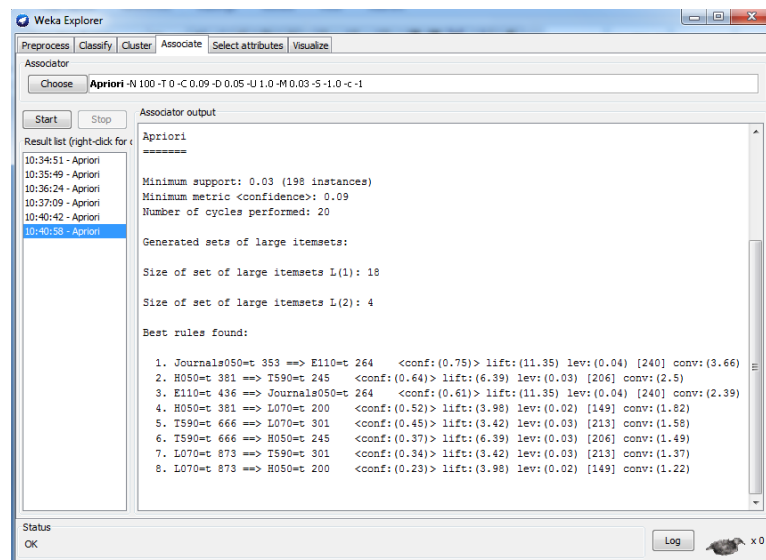

Figure 7: Result association rule found by WEKA 3.7

Display the rules associated with the Minimum Support 0.03 Number of Cyclic Performance 20 Minimum Metric 0.09 is the most appropriate look as figure: 7

1. Rule 1:Journals050 = t 353==> E110 = t 264 that mean if user borrowing book from group Journal050 he/she will borrow from group E110 together by confidence 0.75(75%).
2. Rule 2: H050 = t 381 ==> = t 245 that mean if user borrowing book from group 50 H050 he/she will borrow from group T590 together by confidence 0.64(64%).
3. Rule 3: E110 = t 436 ==> Journals050 = t 264 that mean if user borrowing book from group E110 he/she will borrow from group Journal050 together by confidence 0.61(61%).
4. Rule 4: H050 = t 381 ==> L070 = t 200 that mean if user borrowing book from group H050 he/she will borrow from group L070 together by confidence 0.52(52%).

5. Rule 5: T590 = t 666 ==> L070 = t 301 that mean if user borrowing book from group T590 he/she will borrow from group L070 together by confidence 0.45(45%).

6. Rule 6: T590 = t 666 ==> H050 = t 245 that mean if user borrowing book from group T590 he/she will borrow from group H050 together by confidence 0.37(37%).

7. Rule 7: L070 = t 873 ==>T590 = t 301 that mean if user borrowing book from group L070 he/she will borrow from group T590 together by confidence 0.34(34%).

8. Rule 8: L070 = t 873 ==> H050 = t 200 that mean if user borrowing book from group L070 he/she will borrow from group H050 together by confidence 0.23(23%).

5   Conclusion

Data Mining System is useful to study borrowing behavior of the student in library. In particular we use data mining techniques and provides Association rules for apply to library find out of frequent patterns, correlation from between item in the dataset and implement decision support in library was to selected by Apriori algorithm. From the experimental results by using program WEKA we can found relationships and rules is very appeal. The study shows the total of 8 rules with the confidence of 75%, 64%, 61%, 52%, 45%, 37%, 34%, and 23% respectively the results showed that user's behavior can use to manage the resource library. According to the results we can recommend to Management in library (change library layout) may move group of books nearby together to make comfortable for student. Model in this paper can be benefit to manage in resource library.

REFERENCE

[1]   KanyaratBussaban"Analysis of Users' Behavior on Book Loan Log Based On Association Rule Mining" International Journal of Computer, Electrical, Automation, Control and Information Engineering, Vol: 8, No: 1, 2014

[2]   U. M. Fayyad, G. Pitatesky-Shapiro, P. Smyth, and R. Uthurasamy, *"Advances in Knowledge Discovery and Data Mining",* AAAI/MIT Press, 1996.

[3]   https://www.techopedia.com/definition/30306/association-rule-mining

[4]   Dr. M. Dhanabhakyam , Dr. M. PunithavalliVaibhav Pandey, "A Survey on Data Mining Algorithm", Global Journal of Computer Science and Technology ,Volume 11 Issue 11 Version 1.0 July 2011.

[5]   R. Agrawal, T. Imielinski, and A.N. Swami, A. N., "*Mining association rules between sets of items in large databases*", In Proceedings of the1993 ACM SIGMOD International Conference on Management of Data,pp. 207-216,1993.

[6]   Tang, K., Chen, Y., & Hu, H. (2008).Context-based market basket analysis in a multiple-store environment. Decision Support Systems, 45, 150-163.

[7]   Z. Qiankun, "*Association Rule Mining: A Survey, Technical Report*", CAIS, Nanyang Technological University, Singapore, 2003

[8]   Rakesh Agrawal, Ramakrishnan Srikant, 1994. Fast Algorithms for Mining Association Rules. In: the 20th International Conference on Very Large Databases (VLDB). Santiago, Chile.