

Analysis of Users' Behavior on Book Loan Log Based On Association Rule Mining

Kanyarat Bussaban, Kunyanuth Kularbphettong

Abstract—This research aims to create a model for analysis of student behavior using Library resources based on data mining technique in case of Suan Sunandha Rajabhat University. The model was created under association rules, Apriori algorithm. The results were found 14 rules and the rules were tested with testing data set and it showed that the ability of classify data was 79.24percent and the MSE was 22.91. The results showed that the user's behavior model by using association rule technique can use to manage the library resources.

Keywords—Behavior, data mining technique, Apriori algorithm.

I. INTRODUCTION

NOWADAYS, Library plays an important role as the source of knowledge for Educational institutions. Currently library always supports on and off line services and library services provide much of resources, like book, journal, CD and online database, to support knowledge learning. Computer and communication technology has been currently used in library service to keep resource and user's information. Using technology to search for useful information from the saved data in order to improve library service is worthwhile. Providing precise resources to the users shows the library's efficiency in passing knowledge to the users. When we compare the book-loan behavior of each library, we see that one user can borrow many books at the same time. This can affect the accuracy in keeping track of books for future use; therefore, the library might not have enough resources for the users' need. Therefore, it is very important to analyze library users' behavior in order to improve and manage Library Management.

Data mining is a process to create knowledge from transactional database by using statistic procedure and machine learning and training set to get the exact information for future management. Data Mining can analyze relevant information results and produce different perspectives to understand more about the users' behavior so as to handle library resources. Moreover, it can figure out what factors encourage users to use the library, what categories are related to their interest, and how we predict the future needs.

The remainder of this paper is organized as follows. Section II presents the research methodologies used in this work.

Kanyarat Bussaban is with Mathematics Program, Suan Sunandha Rajabhat University, Bangkok, Thailand (phone: 662-150-1169; e-mail: kanyarat.bu@ssru.ac.th, ktanbee@yahoo.com).

Kunyanuth Kularbphettong is with Computer Science Program, Suan Sunandha Rajabhat University, Bangkok, Thailand (phone: 662-150-1169; e-mail: kunyanuth.ku@ssru.ac.th, kunyanuth.ku@gmail.com).

Section III presents the experimental results based on the purposed model based on data mining technique. This project demonstrates how to analyze user behavior from log-loan data and how to evaluate purposed methodology. Finally, in Section IV conclude the paper with future research.

II. THE METHODOLOGIES

In this section, we illustrate the specified methodologies used in this project.

Data Mining is the data analyzing process from different perspectives also summarizing the useful information results. The data mining process uses many principles as machine learning, statistics and visualization techniques to discover and present knowledge in an easily comprehensible form. There is another definition as "the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data" [1], [2].

Association rule mining is one of the most popular data mining approaches. It is used to discover interesting relationships between variables in databases. According to Agrawal et al. [3], an association rule explains a close correlation between items in a database in the form of $x \Rightarrow y$ where x and y are sets of Itemset I (x and $y \subset I$) and $x \cap y = \phi$. $I = I_1, I_2, \dots, I_m$ is a Itemset of m distinct attributes. The rule is indicated x implies y whereby x is called antecedent and y is called consequent. There are two importance thresholds for measurement association rule mining, minimum support and minimum confidence. The support of a rule $x \Rightarrow y$ is the probability of the Item set $\{x, y\}$ that means the relevance of the rule and the confidence of a rule $x \Rightarrow y$ is the conditional probability of y given x that indicate the accuracy of the rule.

$$\text{confidence}(x \Rightarrow y) = \frac{\text{support}(\{x,y\})}{\text{support}(x)} \quad (1)$$

and

$$\text{support}(x \Rightarrow y) = \frac{\text{support}(\{x,y\})}{\text{Total number of transacti in } D} \quad (2)$$

Set of transactions: $D = \{d_1, d_2, \dots, d_n\}$ each $d_i \subseteq I$

Therefore, confidence is an important measure of the association rules to indicate the strength rules. If the confidence of the association rule $x \Rightarrow y$ is 80%, it means that 80% of the transactions that contain x also contain y , based on users to indicate the specified minimum confidence [4].

Apriori Algorithm is an influential algorithm for association rule mining, purposed by proposed by Agrawal [5] as shown in Fig. 1. The Apriori Algorithm is used level-wise search for

frequent item sets, the sets of items that have minimum support.

Let C_k is Candidate itemsets of size k and L_k is itemsets of size k .

- 1) $L_1 = \{\text{large 1-itemsets}\}$;
- 2) **for** ($k = 2; L_{k-1} \neq \emptyset; k++$) **do begin**
- 3) $C_k = \text{apriori-gen}(L_{k-1})$; // New candidates
- 4) **forall** transactions $t \in \mathcal{D}$ **do begin**
- 5) $C_t = \text{subset}(C_k, t)$; // Candidates contained in t
- 6) **forall** candidates $c \in C_t$ **do**
- 7) $c.\text{count}++$;
- 8) **end**
- 9) $L_k = \{c \in C_k \mid c.\text{count} \geq \text{minsup}\}$
- 10) **end**
- 11) $\text{Answer} = \bigcup_k L_k$;

Fig. 1 Apriori Algorithm [5]

III. EXPERIMENTAL SETUP

In our research, we collected the user's behavior log from Suan Sunandha Rajabhat University Library, in the semesters of 2011 and 2012. The number of user's behavior loaning data was 3,826. The data is composed of personal records, borrowing log file from Library loaning system as Fig. 2.

| SuanSunandha Rajabhat University | | Loan History by Borrower | | | | | | |
|----------------------------------|--|--------------------------|----------------|----------|------------|-------------|-------------|-------------|
| Book # | Author | Copy # | Classification | Barcode | Cost | Date loaned | Date due | Date |
| 001.422.4373n | | 0 | | 11275074 | \$170.00 | 17 Aug 2011 | 31 Aug 2011 | 20 Aug 2011 |
| 001.422.4838# | | 0 | | 11267217 | \$270.00 | 17 Aug 2011 | 31 Aug 2011 | 20 Aug 2011 |
| 302.35.4345w | | 0 | | 11263065 | \$441.00 | 11 Aug 2011 | 25 Aug 2011 | 19 Aug 2011 |
| 658.3.A357D | Organizational behavior and management : an integrated skills approach / Aldag, Ramon J. | 0 | | 21030864 | \$3,525.30 | 11 Aug 2011 | 25 Aug 2011 | 17 Aug 2011 |
| 658.4+979w | Kozuhara, Loren W., Jr. auth. | 0 | | 11297711 | \$207.00 | 11 Aug 2011 | 25 Aug 2011 | 17 Aug 2011 |
| 658.3.H425D | Organizational behavior : Organizational behavior - a.3 / อนันต์ นามะ | 0 | | 21037811 | \$924.30 | 11 Aug 2011 | 25 Aug 2011 | 17 Aug 2011 |
| 658.4.H319D | Organizational behavior : foundations, realities & challenges / Nelson, Dabne L. | 0 | | 21029001 | \$2,395.00 | 11 Aug 2011 | 25 Aug 2011 | 17 Aug 2011 |

Fig. 2 Users' loaning log

In Fig. 3, the data is pre-processed, and transformed to be appropriated format in order to apply data mining techniques to discover association rules.

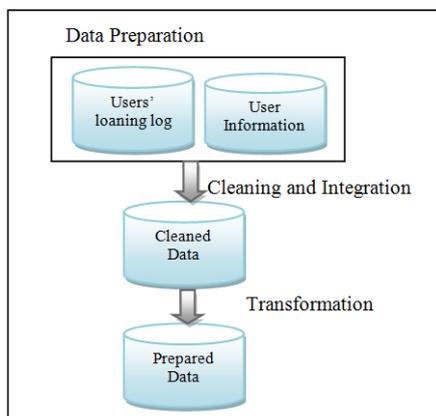


Fig. 3 The data preparation process

After preparation data, we apply the association rule mining algorithm to discover valuable patterns. Data was analyzed by WEKA. WEKA, the Waikato Environment for Knowledge Analysis, is a collection of machine learning algorithm to analyze data set for data mining tasks [6]. Association rule mining is very useful for discovering users' motivation behavior because it presents significant relationships between the activities of users with their activities. Apriori algorithm has been used for this research and evaluated with a minimum support of 0.2 and a minimum confidence of 0.9. Fig. 4 shows examples from the results of Apriori algorithm. To create the model, we divided data into two parts: learning set and testing set. Then the efficiency of the model was compared the predicted result with the real users' data, and calculated to find the right percentage of the prediction and the MSE. And Fig. 4 was shown the model of analyze the users' behavior of this project.

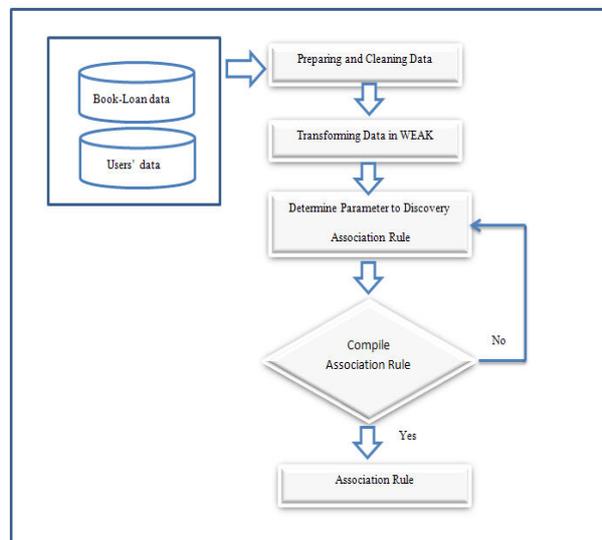


Fig. 4 The model of analyze the users' behavior

IV. RESULTS

In this book-loan behavior study, it was divided the result by the research purpose into 2 parts: developing the model to analyze book-loan behavior by using data mining and testing the model by using testing data set.

A. Developing the Model to Analyze Book-Loan Behavior by Using Data Mining

While developing the analyzing model using library users' data, this process was discovered 14 rules as shown in Table I.

TABLE I
RESULTS OF THIS PROCESS

| Rule | Meaning |
|------|--|
| 1 | "67% of library users who borrowed books from literature, history and geography categories would also borrow books from sociology category." |
| 2 | "67% of library users who borrowed books from miscellaneous and linguistic categories would also borrow books from sociology category." |
| 3 | "63% of library users who borrowed books in linguistic and literature categories would also borrow books from sociology category." |
| 4 | "58% of library users who borrowed books from sociology and literature categories would also borrow books from linguistic category." |
| 5 | "57% of library users who borrowed books in religion category would also borrow books from sociology category". |
| 6 | "56% of library users who borrowed books from miscellaneous and science categories would also borrow books from applied science & technology category" |
| 7 | "55% of library users who borrowed books from sociology and science categories would also borrow books from applied science & technology category" |
| 8 | "54% of library users who borrowed books from applied science & technology and geography & history shelves would also borrowed books from sociology shelves" |
| 9 | "54% of library users who borrowed books in miscellaneous and linguistic categories would also borrow books in applied science & technology category." |
| 10 | "54% of library users who borrowed books in literature category would also borrow books from sociology category." |
| 11 | "53% of library users who borrowed books from miscellaneous and philosophy categories would also borrow books from applied science & technology category." |
| 12 | "53% of library users who borrowed books in linguistic and history & geography would also borrow books in sociology category." |
| 13 | "51% of library users who borrowed books in philosophy and linguistic category would also borrow books in applied science & technology category." |
| 14 | "50% of library users who borrowed books from miscellaneous and linguistic categories would also borrow book from sociology category." |

B. Testing the Model by Using Testing Data Set

When tested the association rule of book-loan behavior analysis made by 924 testing data sets, we discovered that the association rule of book-loan behavior was the most accurate rule in terms of data sorting (79.24%) and the MSE was 22.91%.

V. CONCLUSION AND FUTURE WORK

In this paper, we presents the preliminary result showing a promising progress in this prototypes model for the ongoing improvement project and also this model can be beneficial to manage resource library. However, in term of the future experiments, we are looking forward to research about other data mining techniques to enhance this project and also apply the tool to handle resource in the library.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial subsidy provided by Suan Sunandha Rajabhat University.

REFERENCES

- [1] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurasamy, "Advances in Knowledge Discovery and Data Mining", AAAI/MIT Press, 1996.
- [2] W. Frawley, G. Piatetsky-Shapiro, and C. Matheus, "Knowledge Discovery in Databases: An Overview". AI Magazine, Fall 1992, pp. 213-228.
- [3] R. Agrawal, T. Imielinski, and A.N. Swami, A. N., "Mining association rules between sets of items in large databases", In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pp. 207-216,1993.
- [4] Z. Qiankun, "Association Rule Mining: A Survey, Technical Report", CAIS, Nanyang Technological University, Singapore , 2003
- [5] R. Agrawal, and R. Srikant, "Fast algorithms for mining association rules", In Proc. 20th Int. Conf. Very Large Data Bases, VLDB, pp. 487-499, 1994.
- [6] <http://www.cs.waikato.ac.nz/ml/weka/>

Kanyarat Bussaban received M.S. degree in Statistics. Her current research interests are in the area of Decision making.

Kunyanuth Kularbphetong received the B.S. degree in Computer Business, M.S. degree in Computer Science, and Ph.D. degree in Information Technology. Her current research interests are in Multi-agent System, Web Services, Semantic Web Services, Ontology and Data mining techniques.