

# Book Recommendation System for Digital Library Based on User Profiles by Using Association Rule

Pijitra Jomsri

Department of Information technology  
Suan Sunandha Rajabhat University  
Bangkok, Thailand  
Pijitra.jo@ssru.ac.th

**Abstract**— Due to wide application of management system, information data grows rapidly. On one hand, people have a large number of information resources. On the other hand, the time cost and difficulty of people finding the proper information increases. To tackle the problems, book recommendation is one of the solutions for university libraries which possess huge volumes of books and reading-intensive users. This paper proposes a library book recommendation system based on user profile loaning and apply association rule to create model. The result shows that new association rule algorithm suitable to apply for recommender book in library.

**Index Terms**—Recommendation System, data mining, association rule.

## I. INTRODUCTION

It seems that the use of Internet to search information is likely to increase. Especially, there is an increase on the use of Internet to search for books in the library or the use of on-line library to locate books. There has been continual development in library system on information retrieval to increase effectiveness of information searching to meet users' satisfaction at the highest level. One of the techniques used to foster information retrieval and service is Recommender System. This technique is applied to recommend information which yields more relevant results to users' needs resulting in enhancing users' satisfaction when searching for books in the library.

As previously mentioned, the researcher is interested in developing book recommendation system to serve individual needs. This research applied association rule technique to identify relationship between books that the users are interested in and the availability of books in the system in accordance with book categories. This could facilitate the users when searching for books, and give better searching results. The book recommendation system not only enhanced effectiveness of library system but also helped reduce cost of keeping. Additionally, it facilitated users when searching through a wide range of books on shelves and could develop users' reading habit.

The paper is structured as follows: related works are summarized in Section II. The framework for book recommendation is presented in Section III. Association rule analyses are explained in Section IV. Consequently

Experimental Setting are describes in section V and Experimental result is explained in Section VI. Finally, Section VII describes conclusions and future work.

## II. RELATED WORK

There has been extensive research attempting to solve the problem on sparse rating data found in collaborative filtering (CF). For example, Pazzani[1] investigated collaborative filtering through the use of content-based filtering and demographic filtering. Demographic information of the users such as gender, age and education was used to recommend restaurants to them. The research applied this information to rating points given by the users to add values. This method could reduce the problem on sparse rating data.

Mooney [2] used content-based book recommending technique for text categorization. In the paper, the researcher explored the development of a widely-used recommendation system by using collaborative filtering which used others users' interests to recommend individual users. However, it was found that the recommendation done by using content-based filtering yielded better results because it was directly based on the product. Yan-ge Ma [7] proposed a recommendation algorithm which takes the interconnected characteristic of books by content similarity into account, and considers the attributes information of one book.

In addition, the effectiveness of association rule mining for uncovering patterns and driving personalized marketing decisions has been known for a sometime [3]. There are several specific research projects use association rule on recommendation system such as Mobasher *et al.* [4] presented a system for web personalization based on association rules mining. Their system identifies association rules from page views co-occurrences based on users navigational patterns. Their approached outperforms a *k*NN-based recommendation system both in terms of precision and coverage. Smyth *et al.* [5] presented two different case studies of using association rules for RS. In the first case they use the *a priori* algorithm to extract item association rules from user profiles in order to derive a better item-item similarity measure. In the second case, they apply association rule mining to a *conversational* recommender. The goal here was to find co-occurrent

critiques – i.e. user indicating a preference over a particular feature of the recommended item. Lin *et al.* [6] presented a new association mining algorithm that adjusts the minimum support of the rules during mining in order to obtain an appropriate number of significant rules; therefore, addressing some of the shortcomings of previous algorithms. The measured accuracy outperforms previously reported values for correlation-based recommendation and was similar to the more elaborate approaches such as the combination of SVD and ANN.

Up to now, most of the algorithms recommend items based on the rank or score given by users, whereas in libraries' book-loan logs, no such data was available. Yan *et al.* [8] discussed the users' behavior from the library book-loan log, but the study mainly concentrates on the knowledge dependency of different majors. Some research has aim of the study to discussed book recommendation based on bookloan data, but their method is association rules, and the paper only proposed a service mode without experiment on the algorithm [10]. Zhu, Z [9] proposed an algorithm of probabilistic memory-based collaborative filtering (PMCF), which combines memory-based and model-based techniques. They use a mixture model built on the basis of a set of stored user profiles and use the posterior distribution of user ratings to make predictions. The approach proposed in this paper do not relay on user profiles, whereas the history of users' book-borrow information is needed to build the frequency distribution of users' reading interest.

This paper uses different views to recommend book with focus on the diversity and reliability. This paper extends the method of association rule that is data mining technique by using implicit and explicit data.

### III. FRAMEWORK FOR BOOK RECOMMENDATION

A framework for book recommender system is follows in Fig 1. The framework mechanism was designed in three steps:

#### A. Historical data of user loaning book

Historical data of each user loaning is collect after process of user loan book. The collected documents consist of book ID, book name, book category, author name, user ID, loan date, return date, and etc. In addition, the detail of each category is described as follow:

- 000 Miscellaneous
- 100 Philosophy and psychology
- 200 Religions
- 300 Social sciences
- 400 Languages
- 500 Science
- 600 Technologies
- 700 Arts & recreation
- 800 Literatures
- 900 History & geography

#### B. Association rule

This step is preparing and cleaning data for creating association rule model. The relationship while users loaned

book was analyzed. This technique was described in section IV.

#### C. Recommender Book

Recommender book is process to suggest books for each user by use the result from association rule process.

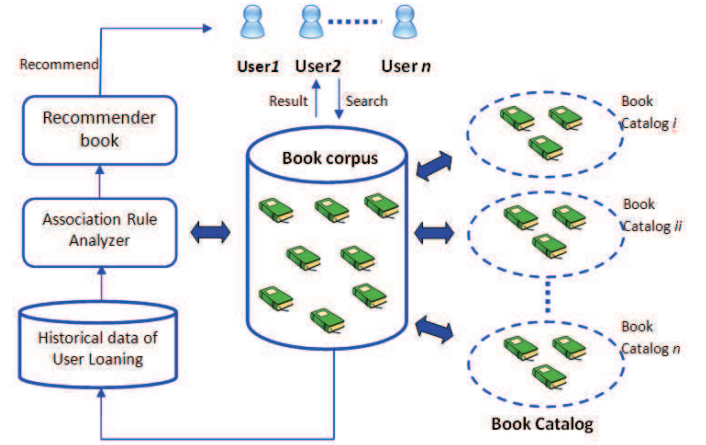


Fig. 1. A Framework for Book Recommendation System

### IV. ASSOCIATION RULE MINING

Association rule Mining is a popular data mining method and well researched method for discovering interesting relationship between variables in large databases. Many used the research lead data mining and association rule for analyzing and increasing efficiency in searching result [11], [12].

This paper used association rule technique from user profile for exploring pattern to improve book recommender system. Researcher explored association of a set user, book category and book title. We expect that relation of the history of user loaning will be significant for creating recommendation system. The detail of Association rule Mining model as follow:

The rules tab in the form of  $U, C, L \rightarrow T$  is applied for extracting rules can call  $UCLModel$  and  $U, L \rightarrow T$  is applied for extracting rules can call  $ULModel$ . Where  $U$ ,  $C$  and  $T$  are disjoint item sets of user ( $U$ ), book category ( $C$ ), book loan ( $L$ ) and book title ( $T$ ). For each rule of the form  $U, C, L \rightarrow T$ , researcher defines the *supp* and *conf* as the *support* and *confidence* as follows.

$$conf(U, C, L \rightarrow T) = \frac{count(U, C, L \text{ and } T)}{count(U, C, L)} \quad (1)$$

$$sup(U, C, L \rightarrow T) = \frac{count(U, C, L \text{ and } T)}{count(All)} \quad (2)$$

Table I shows examples of rules for book relation. Confidence and support value are used for rule selections.

Because plenty of rules are generated, some simple concerns in rule selections include:

- 1) Select the rule with maximum confidence.
- 2) Select the rule with maximum support if confidence value is equal.
- 3) Select the rule that happens first when confidence and support values are equal.

From table I, shows the rule explains:

- Support of  $U, C, L \rightarrow T$  is the probability that a book has in user ( $U$ ), book category ( $C$ ), book loan ( $L$ ) and book title ( $T$ ).
- Confidence of  $U, C, L \rightarrow T$  is probability that a book appear in book title ( $T$ ) given that the user ( $U$ ), book loan ( $L$ ), book category ( $C$ ).

TABLE I. RELATION MODELS OF BOOK THAT USER LOAN WITH CONFIDENCE AND SUPPORT VALUES

Rule			Conf (%)	Sup (%)
User 1, 000, Visual basic 2008	$\Rightarrow$	Quick Start VB. NET 2008	69.77	1.37

## V. EXPERIMENTAL SETTING

The experimental setting is divided into two sections. Section A) describes the data set, section B) discusses describes evaluation metrics.

### A. The data set

The collected documents consist of 65,521 transactions during January 2012 to February 2014. Each record in the historical loaning corpus contains: Book ID, book name, category number, return date, loan date, barcode, type of user and user id.

### B. Evaluation Matrix

The informal was conducted with thirty users that were recruited as experiment participants. In the step of measuring the system accuracy, we need to use information retrieval classification metrics, which evaluate the capability of the system to suggest a short list of interesting items to the user. The precision and recall are the standard measurement for the probability that the system makes a correct or incorrect decision about the user interest. With  $b_x$  being the book from recommend for user  $u$  and  $D(u, b)$  is the set of recommended books, recall and precision are defined as Equation (3) and (4):

$$recall = (D(u, b)) = \frac{1}{|U|} \sum_{u \in U} \frac{|bp(u, b) \cap D(u, b)|}{|bp(u, b)|} \quad (3)$$

$$precision = (D(u, b)) = \frac{1}{|U|} \sum_{u \in U} \frac{|bp(u, b) \cap D(u, b)|}{|D(u, b)|} \quad (4)$$

Where

$|bp(u, b)|$  is the number of is relevant documents,

$|D(u, b)|$  is the number of retrieved documents.

$|bp(u, p) \cap D(u, p)|$  is the number of relevant documents from the number of retrieved documents.

Recall measures the percentage of interesting items suggested to the users, with respect to the total number of interesting items. Whereas, precision measures the percentage of interesting items suggested to the users, with respect to the total number of suggested items. The values precision and recall are shown in section VI. The thirty subjects were considered as experts in the field participated in the experiment. Therefore, their relevancy ratings are assumed to be perfect. In the study setting, each subject is assigned to investigate the book obtained from the  $b_x$ . The 10 books for relevancy are displayed. Finally, the subjects were asked to rate the relevancy of the book recommendation on a two-point scale: score 0 is not relevant at all and score 1 is relevant.

## VI. EXPERIMENTAL RESULTS

The results of the paper were described in two sections. The first section is a result of association rule, and the second one is a result of evaluation by using precision and recall.

### A. Result of association rule

Form table II, the rule which has confidence value showed that more than 60% were choose. The strength rules were hold such as “User 1, 000, Knowledge management” has Confidence 100.00%. The rule means that user1 which loans in Miscellaneous category, book name: Knowledge management will loan in Knowledge management: theoretical concepts. Therefore, the relationship of these rules may help to recommend book to user. However, “User, 600, Digital system designs and practice” has Confidence 21.67%. So User which loans in Technologies category, book name: Digital system designs and practice will not loan in Communication skill always. Therefore, the relationship of this rule may not help to recommend book.

TABLE II. THE ACCURACY OF CONFIDENCE TRAINING VALUE AND CONFIDENCE TESTING VALUE

Rule			Conf (%)	Sup (%)	Rule hold
User 1, 000, Visual basic 2008	$\Rightarrow$	Quick Start VB. NET 2008	69.77	1.37	Yes
User 1, 000, Knowledge management	$\Rightarrow$	Knowledge management : theoretical concepts	100	10.5	Yes
User2, 600, Fundamental nutrition	$\Rightarrow$	Nutrition for healthy living	67.33	1.01	Yes
User2, 300, Food business management	$\Rightarrow$	Food preparation and theory	60.00	0.86	Yes
User3, 400, English in scientific context	$\Rightarrow$	Reading skill development	59.09	0.82	No

Rule			Conf (%)	Sup (%)	Rule hold
User3, 400, TOEFL reading comprehension	⇒	TOEIC reading review-tests	66.23	0.79	Yes
User4, 700, Music : the art of listening	⇒	Music of the masters	70.71	0.78	Yes
User4, 700, Folk music and traditional performing arts of Thailand	⇒	Jazz styles : history & analysis	82.65	0.76	Yes
User5, 100, Psychology in everyday life	⇒	General psychology	55.76	0.75	No
User5, 100, Educational technology research	⇒	Psychology of reading	67.74	0.73	Yes
User6, 800, Phra Aphai Mani II	⇒	Phra Aphai Mani III	52.38	0.72	No
User6, 800, Thai usage1	⇒	Thai usage2	100	0.69	Yes
User7, 000, JAVA OOP	⇒	Intro JAVA	63.00	0.67	Yes
User7, 000, PHP	⇒	C++	71.21	0.68	Yes
User8, 600, Digital system designs and practice	⇒	Communication skill	21.67	1.23	No
User8, 600, Principles of electric circuits	⇒	Electronic devices and circuit theory	78.57	2.82	Yes
User9, 300, Need to know : social science research methods	⇒	Doing action research in your own organization	37.98	1.75	No
User9, 000, Java	⇒	Into JAVA	48.86	2.82	No
User9, 000, HTML & XHTML	⇒	Flash 8	60.26	1.34	Yes
User10, 600, Nursing process	⇒	Nursing care of the child	89.21	1.89	No

### B. Result of evaluation Matrix

Since the subject relevancy ratings for book were recommender follow in equation (4). The experiment result is depicted in table III. The values in each cell the average of 30 search terms precision values. In this experiment compare between *UCLModel* with *ULModel*. *UCLModel* is technique of this paper suggestion and *ULModel* is traditional technique by using user profile. The result show that *UCLModel* better than *ULModel*. In high precision value indicates high accuracy and the result show that Precision In top-5 shows the performance better than Precision In top-1.

TABLE III. PERFORMANCE OF PROPOSED FRAMEWORK

Method	Precision In top-1 (%)	Precision In top-5 (%)
<i>UCLModel</i>	91.34%	92.00%
<i>ULModel</i>	90.00%	91.22%

## VII. CONCLUSION AND FUTURE WORKS

This research aims to present book recommendation system for digital library. The researcher analyzed user profiles which were the history of borrowing and book categories and related data Furthermore, The findings of a model of book recommendation system developed to facilitate individual users for book searching indicated that the users satisfied with the book recommendation system. In further studies, the researcher will include other factors to analyze relations, apply other data mining techniques and compare with the developed model

## ACKNOWLEDGMENT

The authors would like to thank Suan Sunandha Rajabhat University for scholarship support. The study is not possible without the data from Academic Resource Center.

## REFERENCES

- [1] M. J. Pazzani, "A Framework for Collaborative, Content-Based and Demographic Filtering," *Artificial Intelligence Review* 13, no. 5, p.393-408, 1999.
- [2] R. J. Mooney, and L. Roy, "Content-based book recommending using learning for text categorization," *Proc. of the Fifth ACM Conf. on Digital Libraries*, p. 195–204, 2000.
- [3] R. Agrawal, and R. Srikant, "Fast algorithms for mining association rules in large databases," In *Proceedings of the 20th International Conference on Very Large Data Bases*, 1994.
- [4] B.Mobasher, H. Dai, T. Luo, and M. Nakagawa, "Effective personalization based on association rule discovery from web usage data," In *Workshop On Web Information And Data Management, WIDM '01*, 2001.
- [5] B. Smyth, K. McCarthy, J. Reilly, D. O'Sullivan, L. McGinty, and D. Wilson, "Case studies in association rule mining for recommender systems," In *Proc. of International Conference on Artificial Intelligence (ICAI '05)*, 2005.
- [6] W. Lin, and S. Alvarez, "Efficient adaptive-support association rule mining for recommender systems," *Data Mining and Knowledge Discovery Journal*, 6(1), 2004.
- [7] Yan-ge Ma, "Research of Library Book Recommendation System Based on Cloud Computing," *Proceedings of the 9th International Symposium on Linear Drives for Industry Applications*, Volume 1, 2014.
- [8] F. Yan, M. Zhang, J. Tang, T. Sun, Z. Deng, L. Xiao, "Users' Book-Loan Behaviors Analysis and Knowledge Dependency Mining," *WAIM 2010. LNCS*, vol. 6184, p. 206–217.
- [9] Z. Zhu, J. Wang, "Book Recommendation Service by Improved Association Rule Mining Algorithm," *IEEE Press*, 6th International Conference on Machine Learning and Cybernetics, p.3864–3869, 2007.
- [10] K. Yu, A. Schwaighofer, V. Tresp, X. Xu, , H. Kriegel, "Probabilistic Memory-based Collaborative Filtering," *IEEE Transactions on Knowledge and Data Engineering* 16(1), P.56–69, 2004.
- [11] X. Chen, and Y. Wu, "Personalized Knowledge Discovery: Mining Novel Association Rules from Text," *Siam Conference on data mining*, 2006.

- [12] C. Schmitz, A. Hotho, R. Jäschke, and G. Stumme, “Mining Association rule in Folksonomies,” *Journal of Information Science (JIS)*, 2008.