# Using Data Mining to Provide Recommendation Service

Ruey-Shun Chen[a], Yung-Shun Tsai[b], K.C. Yeh[b], Chia-Ming Sun[b,c], D.H. Yu[b], and Yip Bak-Sau[d]
[a] Department of Information Management, China University of Technology,
530 Sec. 3 Chung Shan Road, Hukou, Hsinchu, Taiwan
[b] Institute of Information Management, National Chiao Tung University,
1001 Ta Hsueh Road, Hsinchu, Taiwan
[c] Department of Accounting, National Yunlin University of Science & Technology,
No. 123, University Road, Section 3, Douliu, Yunlin County, 64002, Taiwan
[d] Institute of Biotechnology, Department of Life Science, National Tsing-Hua University
[d] Department of Neurology, Hsin-Chu General Hospital,
Department of Health, The Executive Yuan, Taiwan

*Abstract:* This research introduces personalized recommendation service into library services. Using the borrowing record of the library as basis, the association rules of data mining technique are used to look for book association by focusing on reader's borrowing mode, personal interest and trait in order to simplify the complexity of recommendation structure. The Bayesian network concept is used to build up a personalized book recommender system in order to generate different book recommendations, ranking from high to low, to help reader to locate book information most suitable to his requirement. Meanwhile we use user satisfaction questionnaire to understand the accuracy of recommended books and further to feedback information in order to help the post learning of Bayesian network parameter. This is for the perfection of the overall structure of recommender system so that readers could make use of the resource of the library more effectively and the value of the library system could be further improved.

*Key-Words:* Recommendation System, Bayesian Network, Data Mining

## 1 Introduction

The evolution of society is moving ever faster. Development of Internet has turned the society into a generation of automation. Libraries are also moving toward this trend of automation and are actively offering services to readers. The book recommendation with digital library as background is very similar to the so-called product recommendation in the e-commerce domain. In the past several years, we discovered the appearing of many recommender systems that could provide users with a kind of personalized recommendation service to their product varieties and services that allow users to select easily the product and service their prefer or are interested in, including articles, books, albums, or movies [8].

The establishment of digital library is to upgrade the traditional library service quality by using information automation and network technology. However, in today's environment where are too many data and too easy to secure, the flooding of information often render users unable to look for information they need as they desire. Therefore, how to allow users to secure the data they need correctly and quickly and developing personalized smart user interface is the important issue of the development of contemporary digital library [18]. Therefore, the personalized recommendation is introduced into library services to build up a personalized book recommender system through using existing historical record files in the database by focusing on readers' borrowing modes, personal interests and traits so as to assist readers to locate the book information that meet their requirement.

## 2 Literature Review

### 2.1 Recommendation System

Recommender system is a personalized service tool based on information filtering [10]. The system learns user's preference through analyzing user behavior and interest and further provides product or information meeting individual's need. Schafer et al. [16] suggested that the effective utilization of recommender system into e-commerce website could bring many benefits such as converting browsers into buyers, increasing cross-sell, and building loyalty.

There are three recommending methods in the recommender system. First, content-based recommendation mainly came from the concept of

information accessing and is a kind of recommendation method based on comparing users' preferences and associating contents between items in order to provide recommendations to users [12], [17]. This content-based method is also called feature-based recommendations [19] that judge and find out items users possibly interested in by analyzing the attributes and characteristics based on user profile. It could even further assign different weights [4], [6], [15] based on the degree of association between user's preferences and targeted contents in order to better fit users' requirements.

Second, collaborative filtering recommendation is the recommendation mechanism currently used widely. The earliest research applying collaborative filtering recommendation in the recommendation system is the TAPESTRY system [6] suggested by Goldberg in 1992. This system is an e-mail system developed by Xerox Palo Alto Research Center, using collaborative filtering mechanism to solve the e-mail flooding problem. Breese et al. [2] divided the collaborative filtering recommendation into two major types based on the algorithms used. They are memory-based collaborative filtering and model-based collaborative filtering. Though the collaborated filtering recommendation mechanism is being widely used at the present moment, there are still many restrictions such as sparsity, scalability, and cold start [5], [13].

Finally, both content-based recommendation and collaborated recommendation have their restrictions and pitfalls. Therefore, many scholars subsequently proposed the Hybrid Approach Recommendation [1], [9] that took the advantages and make up the shortcomings in order to compensate the each other's pitfalls.

### 2.2 Data Mining

Through the assistance of information technology, business could discover the knowledge and rules hidden in the complicated information through uninterrupted exploration from large amount of information. This process is the so-called data mining. To sum up, the function of data mining is divided into five types which are affinity grouping, data classification, data clustering analysis, estimation, and prediction.

### 2.3 Bayesian Network

Bayesian network technique can be viewed as a kind of network learning [14]. Bayesian network is a kind of graphical knowledge expression method that expresses the uncertainty and causal relationship in the knowledge through network graphics. Therefore, Bayesian network can take care of both quantitative and qualitative analyses. Bayesian network can handle the situation where data is incomplete, learn the causality between variables, combine domain knowledge and historical data through Bayesian statistics, and avoid data overfit problem [7].

## 3 System Design

### 3.1 Reader's Model of Borrowing

#### 3.1.1 Structure of Bayesian Network

The method of constructing Bayesian network is divided into two parts: One is for the experts to construct based on experience or domain knowledge, and the other is to construct using the existing data through learning. Before constructing Bayesian network, the reader borrowing history record data is converted into matrix of borrowing actions and borrowed books for subsequent ease of calculation of related probabilities.

We use Apriori to analyze the association between books from the matrix. After finding out the association, construct the complete Bayesian network structure based on the principle of condition attribute as parent and decision attribute as child, and the length of the rule will seriously affect the complexity during construction of network structure. Here the so-called "length of rule" refers to the number of condition attributes at the front of the rule. Assuming the rule is A^B? C, then the condition of the rule is A and B, and its length of rule is 2.

The length of rule is related to the causality during construction of network. Condition attribute can be converted to the cause in the network, or the parent node. Decision attribute can be converted to the result in the network, or the child node. Hence if the length of the rule is too long, the causality will become complicated and the possibility of associations after conversion into Bayesian network will be more, as in a certain rule, the attribute of parent can also become the parent and child causality in other rule. Therefore, the length of rule is limited to 2 for the purpose of reducing the complexity of network structure. Based on the Bayesian network construction rule stated above, the associations of reader historical borrowing data are examined one by one to generate the Bayesian network structure.

#### 3.1.2 Probability

After completion of Bayesian network, the next task is to build up the probability table of this Bayesian network structure, which is also the probability of occurrence of each node in the network.

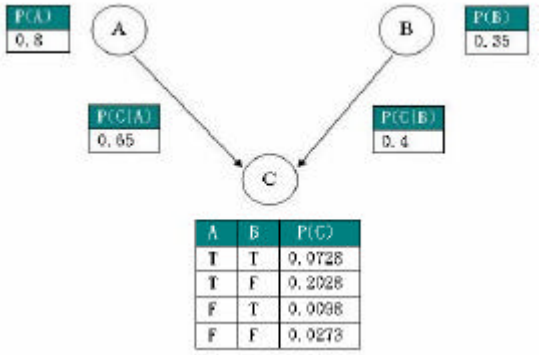
The calculation of Bayesian network probability is to use Equation (1), starting from the root node of each network and learn the condition probability table of each node. Using Figure 1 as an example, assume a reader wished to find out the probability table at node C. We must first know the probability of occurrence of parent node A and B at node C and the values of conditional probabilities P(C|A) and P(C|B) before we could build up the probability table at node C. Therefore, under the condition selecting book A and book B, the probability reader will also select book C is 0.0728. Hence the probability table of each node of the whole network can be obtained by applying this concept.

$$P(X) = \prod_{x} P(x \mid P_a(X)) \qquad (1)$$

$P(X)$ : node probability we wish to calculate.

$P_a(X)$ : parent node probability for the calculation of probability node.

$P(X) = \prod_{x} P(x \mid P_a(X))$ : indicates the probability of occurrence under the condition of occurrence at partent node.




| A | B | P(C) |
|---|---|---|
| T | T | 0.9728 |
| T | F | 0.2028 |
| F | T | 0.0098 |
| F | F | 0.0273 |

Fig. 1 The structure and probability of Bayesian network

### 3.2 Personalized Recomme ndation System

The task of this personalized recommendation system in this research is, while reader is borrowing book online, the system is deducing the Top-N list of other related books suitable to the reader from personalized Bayesian network and wishes to notify reader there might be other books in the library suitable to his preference or need and allow him to borrow all books he prefers right away and also have the opportunity to utilize the books in the library.

After completion of Bayesian network recommendation structure, different reader will have different Bayesian network probability table. Assuming Figure 2 is the Bayesian network of reader A and Figure 3 is the Bayesian network table of reader B, the values presented in the tables are the probabilities of occurrence of each node (or indicating the probability of being borrowed by the reader) under the condition of evidence nodes (dark color nodes in the figures).

Assume the threshold value is 0.6, or indicating the conditional probability satisfying being recommended is 0.6. Under this condition, the set of books system will recommend to reader A is { D, L, Q, R }. For the same reason, the set of books system will recommend to reader B is { I, P, Q, R }. In other words, though readers A and B borrowed the same book (i.e. input the same evidence node), the recommender system also might recommend different book items.

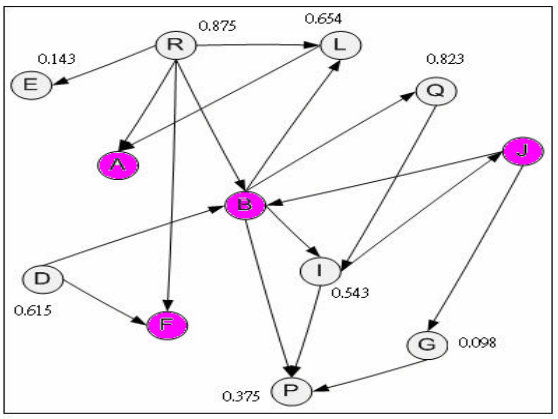



Fig. 2 Book recommendation Bayesian network of reader A

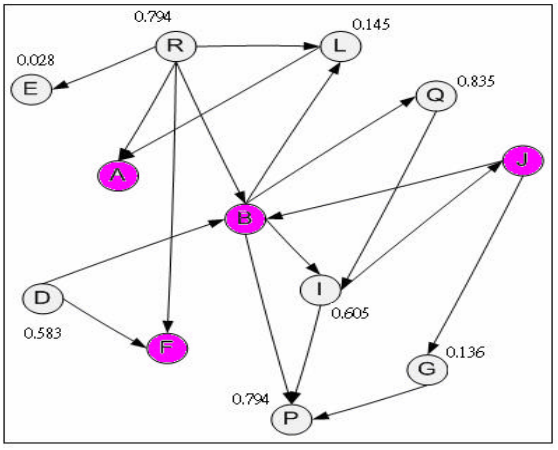



Fig. 3 Book recommendation Bayesian network of reader B

## 4 Example Analysis

### 4.1 The Construct Procedure of Recommendation System

A real example is used to explain the construction process of the book recommender system. There are

six steps, and they are described in sequence as shown below.

(1). Step1: Collect data and convert the data format.

The raw book borrowing information format is converted into the reader required and the borrowed book quantity information.

(2). Step2: Build up reader interest data.

Find out reader's interest to each type. Assume reader card number can be obtained from the data of readers' number of times books are borrowed during certain time section corresponding to the information table of type of books borrowed. The value in the table represents the number of times of the type of book each reader borrowed. Calculating based on the previously described reader card number corresponding to the books borrowed information table, the values of interest each reader has for each type of books can be obtained and are expressed in percentage.

Subsequently the readers' book interest data table can be obtained through further calculation based on the type ratio data readers are interested by setting the condition of threshold value of readers' interest as 0.2.

(3). Step3: Build up reader interest association rules.

Grouping readers based on interest by using algorithm. Data table is used to identify the group reader belongs to could generate association rule of reader's interest.

(4). Step4: Build up association rules between books.

Data table generated based by converting the data based on the historical book borrowing trade record. The association rule between books can be obtained through the analysis of ARMADA association module of Matlab software as shown in Table 1 and 2.

Table 1 Book association rule (length of rule is 1)

| Rule | Support(%) | Confident(%) |
|---|---|---|
| 047→121 | 61.25 | 93.45 |
| 019→047 | 51.64 | 79.89 |
| 019→122 | 44.78 | 95.86 |
| 180→033 | 57.89 | 96.25 |
| 180→121 | 45.98 | 91.64 |
| 039→122 | 48.16 | 78.12 |
| 049→175 | 49.54 | 88.74 |
| 016→175 | 59.12 | 92.56 |
| 016→031 | 54.88 | 84.62 |
| 016→122 | 53.26 | 67.88 |
| 175→124 | 44.16 | 74.65 |
| 175→031 | 41.76 | 91.54 |
| 124→179 | 42.46 | 65.83 |
| 031→122 | 60.36 | 79.25 |
| 031→124 | 50.24 | 76.83 |
| 031→179 | 45.28 | 86.54 |
| 033→121 | 59.54 | 92.35 |

Table 2 Book association rule (length of rule is 2)

| Rule | Support(%) | Confident(%) |
|---|---|---|
| 031∧124→175 | 58.65 | 93.75 |
| 016∧031→122 | 59.25 | 95.25 |
| 180∧033→121 | 56.72 | 98.16 |
| 175∧031→179 | 51.28 | 90.25 |

(5). Step5: Build up book recommendation Bayesian network structure.

First, the Bayesian network structure must be constructed by grouping similar book document attributes and their causality associations. After completion of Bayesian network structure, each node is assigned a probability value based on accumulated experiences. As each node represents a piece of book data and each link represents the association order on both ends of the node, through the handling steps and process we could complete the connecting probability structure from root node to leaf node. Finally, the book data represented by the node is recommended to the reader.

Based on the association rule between books generated in step 4, we construct Bayesian network structure model one by one using MSBN and save the final result into XML format. As MSBN could easily generate errors in the Chinese character processing, this example make use of book code to construct network nodes.

(6). Step6: Build up personalized top-N book recommendation list.

When readers logon to the system, the borrowing information is retrieved through the book borrowing historical record to recalculate its book classification interest data, interest association, and book association, etc. Finally, the reader's Bayesian network structure is built up as shown in Figure 4. Then, the probability value of each book displayed in the evaluated bar chart is used for judgement. Stat 0 indicates it meets the probability of interest. If the recommendation threshold value is 0.6, the top-n (setting n=5) book recommendation list will be generated with books of {019, 033,047, 124, 039} in sequence.

## 4.2 Analysis of Result

Statistics and evaluation of each reader's satisfaction of recommendation list shows that 115 readers are recommended 575 books and readers felt 439 books in the list are satisfactory. Therefore, satisfaction level is 76%, indicating readers have 76% satisfaction level about the recommended list. If this is combined with the library system, the reader's

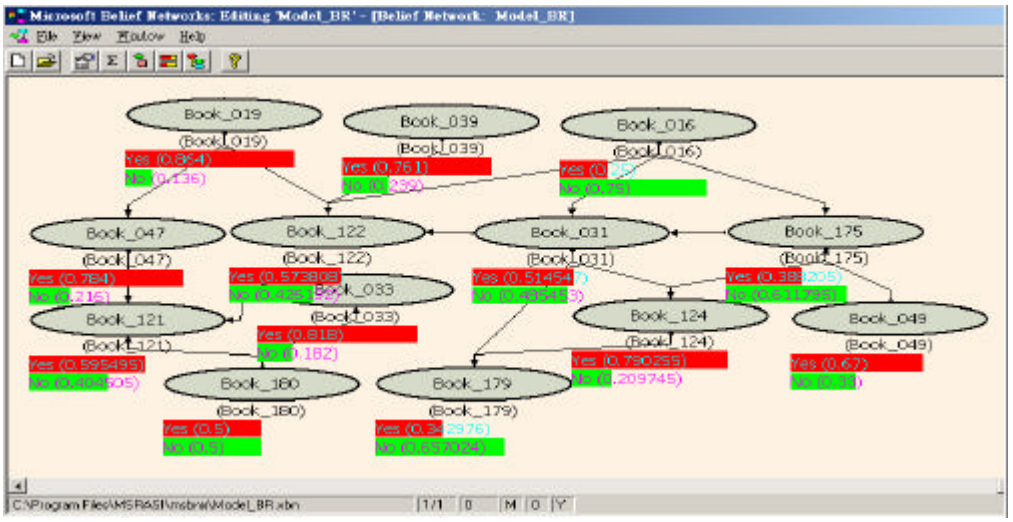


Fig. 4 Bayesian network structure of book recommendation

using rate of the system could be improved and the utilization ratio of the library resource could be increased.

The F1 coefficient proposed by Mobasher et al. is used as index to evaluate the benefit of recommender system and the calculation method is as shown in the equations (2)~(4) below.

$$F1=\frac{2*Precision*Coverage}{Precision+Coverage} \quad (2)$$

$$Precision=\frac{|Relevent \cap Recommendation|}{|Recommendation\ Items|} \quad (3)$$

$$Coverage=\frac{|Relevent \cap Recommendation|}{|Relevent\ Items|} \quad (4)$$

Table 3 is the calculation data. Looking from F1 index, F1 coefficient is largest when this recommender system is at threshold value of 0.6. This indicates that the benefit of recommender system under this condition is the largest, or 0.6 is the best recommendation threshold value.

Table 3 Evaluation index table of each recommended threshold value

| Threshold value | Accuracy rate | Coverage rate | F1 coefficient |
|---|---|---|---|
| 0.1 | 0.345 | 0.969 | 0.508 |
| 0.2 | 0.412 | 0.891 | 0.563 |
| 0.3 | 0.494 | 0.841 | 0.622 |
| 0.4 | 0.603 | 0.779 | 0.680 |
| 0.5 | 0.648 | 0.764 | 0.701 |
| **0.6** | **0.689** | **0.748** | **0.717** |
| 0.7 | 0.789 | 0.603 | 0.684 |
| 0.8 | 0.821 | 0.509 | 0.628 |
| 0.9 | 0.912 | 0.315 | 0.468 |

## 5 Conclusion

Use of Bayesian network technology in the model-based type of collaborated filtering recommendation method to construct a book recommender system on one hand could reduce the complexity of constructing Bayesian network structure through utilizing association rule and on the other also could generate book list through the inference of Bayesian network. In the Bayesian network part, the Bayesian post inference to compensation the recommendation misses of association rules.

Below are several points contributed by this research:

(1). Application of Bayesian network in personalized book recommendation: Book recommendation made by Bayesian network technique does not have the so-called cause and effect nodes. Each book itself could be the cause affecting other book being borrowed and could also be the result affected by other books. Therefore, the application of Bayesian network technique is quite different from the past.

(2). Using association method to simplify the complexity during construction of Bayesian network: Many Bayesian network construction algorithms need to be based on the "node sequence already known" condition. The purpose is to reduce the complexity of constructing Bayesian network structure through this, and also to simplify the complexity of constructing Bayesian network by assuming association rules between books on the book recommendation problem.

(3). Possible application in deciphering the book recommendation: As Bayesian network is the

knowledge expression method developed by combining Bayesian probability theory and graphical mode, therefore applying the Bayesian network technique on the book recommendation could assist in the explanation of book recommendation result.

(4). Assist readers in the efficiency of borrowing book: The final purpose is to assist readers using library resource in effectively receiving books suitable to their interests. Using the recommender system, the system could provide readers with recommendation book list through the recorded data of reader historical borrowing behavior or preference of similar readers and assist their efficiency and convenience in borrowing books.

*References:*


[1] M. Balabanovic and Y. Shoham, Fab: Content-based Collaborative Recommendation, *Communications of ACM*, Vol.40, No.3, 1997, pp. 66-72.

[2] J.S. Breese, D. Heckerman, and C. Kadie, Empirical Analysis of Predictive Algorithms for Colaborative Filtering, *Proceedings of the 14th Annual Conference on Uncertainty in Artificial Intelligence*, 1998, pp. 43-52, 1998.

[3] J. Cheng, D. Bell, and W. Liu, An Algorithm for Bayesian Belief Networks Construction from Data, *Proceedings of AI & START' 97*, 1997, pp. 83-90.

[4] K.W. Cheung, J.T. Kwok, M.H. Law, and K.C. Tsui, Mining Customer Product Ratings for Personalized Marketing, *Decision Support Systems*, Vol.35, No.2, 2003, pp. 231-243.

[5] H.Y. Cho and J.K. Kim, Application of Web Usage Mining and Product Taxonomy to Collaborative Recommendations in E-commerce, *Expert Systems with Applications*, Vol.26, No.2, 2004, pp. 233-246.

[6] D. Goldberg, D. Nichols, B.M. Oki, and D. Terry, Using Collaborative Filtering to Weave an Information TAPESTRY, *Communications of the ACM*, Vol.35, No.12, 1992, pp. 61-70.

[7] D. Heckerman, *A Tutorial on Learning with Bayesian Networks*, Technical Reporter MSR-TR-95-06, Microsoft Research, 1995.

[8] J.B. Schafer, J.A. Konstan, and J. Riedl, E-commerce Recommendation Applications, *Data Mining and Knowledge Discovery*, Vol.5, No.1, 2001, pp. 115-153.

[9] B.D. Kim and S.O. Kim, A New Recommender System to Combine Content-based and Collaborative Filtering Systems, *The Journal of Database Marketing*, Vol.8, No.3, 2001, pp. 244-252.

[10] Y. Li, L. Lu, and L. Xuefeng, A Hybrid Collaborative Filtering Method for Multiple-interests and Multiple-content Recommendation in E-commerce, *Expert Systems with Applications*, Vol.28, No.1, 2005, pp. 67-77.

[11] T.H. Roh, K.J. Oh, and I. Han, The Collaborative Filtering Recommendation Based on SOM Cluster-indexing CBR, *Expert Systems with Applications*, Vol.25, No.3, 2003, pp. 413-423.

[12] S.J. Russell and N. Peter, *Artificial Intelligence: A Modern Approach*, Prentice-Hall, 1995.

[13] M. Gery, M.H. Haddad, Evaluation of Web Usage Mining Approaches for User's Next Request Prediction, *Fifth ACM International Workshop on Web Information and Data Management*, WIDM 2003, New Orleans, Louisiana, USA, November 7-8, ACM Press, 2003, pp. 74 -81.

[14] A. Mild and T. Reutterer, An Improved Collaborative Filtering Approach for Predicting Cross-category Purchased Based on Binary Market Basket Data, *Journal of Retailing and Consumer Services*, Vol.10, No.3, 2003, pp. 123-133.

[15] R.J. Mooney, L. Roy, Content-based Book Recommending Using Learning for Text Categorization, *Fifth ACM Conference on Digital Libraries*, June 2-7, 2000, San Antonio, USA, pp. 195-204.

[16] J. Schafer, K. Ben, A. Joseph, and J. Riedl, E-commerce Recommendation Applications, *Data Mining and Knowledge Discovery*, Vol.5, No.1, 2001, pp.115-153.

[17] U. Shardanand and P. Maes, Social Information Filtering: Algorithms for Automating Work of Mouth, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1995, pp.210-217.

[18] S. Cohen, J. Fereira, A. Home, B. Kibbee, H. Mistlebauer, and A. Smith, My Library: Personalized Electronic Services in the Cornell University Library, *D-Lib Magazine*, April 2000.

[19] S.S. Weng, and M.J. Liu, Feature-based Recommendation for One-to-one Marketing, *Expert Systems with Applications*, Vol.26, No.4, 2004, pp.493-508.