Using Data Mining to Provide Recommendation Service

Ruey-Shun Chen [1,*], Yung-Shun Tsai [2], K.C. Yeh [2], D.H. Yu [2], and Yip Bak-Sau [3]
[1] Department of Information Management, China University of Technology,
530 Sec. 3 Chung Shan Road, Hukou, Hsinchu, Taiwan
[2] Institute of Information Management, National Chiao Tung University,
1001 Ta Hsueh Road, Hsinchu, Taiwan
[3] Institute of Biotechnology, Department of Life Science, National Tsing-Hua University
[3] Department of Neurology, Hsin-Chu General Hospital,
Department of Health, The Executive Yuan, Taiwan
[*] E-mail: rschen@cc.nctu.edu.tw

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.

For libraries, the catalyst that has moved them into a new dimension is information technology (IT). The application of IT in libraries includes integrated library systems, database development, online search and retrieval services, new modes of document delivery, and digital libraries [22], [25]. In the field of a digital library, a number of researchers have utilized a recommendation system to improve the core activities of library services. The most commonly used recommendation system methods are collaborative filtering, content-based filtering, and hybrid filtering [20]. Bollen et al. [23] proposed a minimalist methodology to develop usage-based recommender systems for multimedia digital libraries. Gao et al. [20] unified partition-based collaborative filtering and meta-information filtering and applied them to a digital resource management system. Lee et al. [21] presented two agent-based recommender systems for supporting

Internet commerce. Symeonidis et al. [24] proposed a model-based approach based on latent semantic indexing to improve the performance of collaborative filtering algorithms.

However, few studies have reported on the application of data mining and Bayesian Network on recommendation system for a digital library. The purpose of this study was to integrate data mining with Bayesian Network to introduce personalized recommendation service into library services. Furthermore, a library would demonstrate the proposed system. The findings of this study revealed that the proposed system should improve the library system and readers could make use of the resource of the library more effectively.

The rest of this paper is organized as follows: Section 2 reviews related literature. Section 3 describes related demand analysis. The framework of the proposed system is also outlined in Section 3. Section 4 describes the actual implementation and performance evaluation of the proposed system. Finally, Section 5 presents conclusions.

2 Literature Review

2.1 Recommendation System

Recommender system is a personalized service tool based on information filtering [5], [10]. The system learns user's preference through analyzing user behavior and interest and further provides product or information meeting individual's need [2], [33]. Schafer et al. [8] suggested that the effective utilization of recommender system into e-commerce website could bring in the following three benefits:

(1). Converting Browsers into Buyers: User often encounters too much product information when browsing over the Internet, and most information does not meet user's interest. Recommender system will provide information based on user's interest and preference. It could further stimulate consumption and prompt the user's purchasing desire and become a purchaser.

(2). Increasing Cross-sell: Recommender system could recommend product that other customers might be interested in based on the preference of the user and generates the benefit of purchasing at the same time. The cross sales opportunity could be improved if the product recommended could meet customer's need.

(3). Building Loyalty: In the commercial competition environment, securing customer loyalty is the key for business to profit. Recommender system could build up good relationship between business and customers, analyze consumer needs, and further provide personalized information and lower customer churn rate.

There are three recommending methods in the recommender system. One is Content-based Recommendation, which recommends the product preferred by the user. Second is Collaborative Filtering Recommendation, which recommends product with features similar to the product preferred by customer. Below is the description of the three recommendation methods. However, as both Content-based Recommendation and Collaborative Filtering Recommendation have their own respective restriction and drawbacks, many scholars subsequently recommended Hybrid Approach Recommendation to compliment their drawbacks and are described as below.

2.1.1 Content-based Recommendation

Content-based Recommendation [12] 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. 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. The results are then recommended to users. It could even further assign different weights [15], [29], [31] based on the degree of association between user's preferences and targeted contents in order to better fit users' requirements.

In addition, though the content-based recommendation method could recommend items meeting user's preference items base on user's past preference, there are several restrictions listed below [17]:

(1). Content-based recommendation method can only compare the preference and item features and check whether texts meet the criteria, but could not automatically analyze and judge the attributes of media including sound, picture, and video.

(2). Content-based recommendation cannot find out potential information that user has never attempted in the past but could be interested. It can only recommend information user has seen before or similar things. But in actual fact, in the process of looking for data, user often accidentally discovered information never thought before but very meaningful. This kind of information is the extra discovering service

that content-based recommendation still cannot achieve.

(3). Content-based recommendation cannot filter the quality of the recommended items. Assume two articles both meet the recommendation but with significant difference in quality. The content-based recommendation method has no way of telling their differences.

(4). Content-based recommendation will generate Synonymy problem. There are situation where many similar products are having different names in real life. Most content-based recommendation methods could not find out the relationship between different names but describing the same product. For example, user might wish to find out product information of "light pen". But light pen sometimes is called "laser pen". Now user obviously would not be able to obtain information related to "laser pen".

To improve the pitfalls of above described content oriented recommendation method, some scholars have recommended the collaborative filtering recommendation method in order to compensate the inadequacy of content-based recommendation method. Below is a detailed introduction of collaborative filtering recommendation method.

2.1.2 Collaborative Filtering Recommendation

Collaborative filtering recommendation is the recommendation mechanism currently used widely [34]. The earliest research applying collaborative filtering recommendation in the recommendation system is the TAPESTRY system [31] 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. The operating method of this technique is to find out a group of users with common interests in the beginning and analyze the common interests or preferences shared by the group members. Then, calculate the similarity between each user based on historical records and find out the neighbors with closer preferences. The opinions or recommendations of neighbors are then used to generate recommendation information for user who never experienced before but could be interested as reference. Therefore, what this method emphasizes is the cooperation between people. Then, Breese et al. [27] divided the collaborative filtering recommendation into two major types based on the algorithms used:

(1). Memory-based Collaborative Filtering: During recommendation the user's historical record must be use to make similarity comparison in order to find out the neighboring group with preference closer to user before making recommendation based on the group preferences. The most common method of this type is "Nearest Neighbors Method".

(2). Model-based Collaborative Filtering: This type is mainly to construct the user preference model through statistical method or machine learning method using user's historical data, and further generates recommendation using this preference model. For example, the Latent Semantic Indexing technique is using Single Value Decomposition technique to find out the neighboring group of certain user. The common methods belonged to this type includes "Latent Semantice Indexing", "Association Rule", "Bayesian Network", "Regression Analysis", etc. The Bayesian network method used in this research belongs to this type.

Collaborated filtering recommendation could resolve some of the problems with content-based recommendation. Using this method does not have to analyze the item content, as the system will use the preferences of the group members having same interests with the user as the target of reference for the recommendation [4]. In other words, it is based on the opinion of other customers to recommend product to target customer. Therefore, items that user might not prefer might also appear in the recommended product.

Though the collaborated filtering recommendation mechanism is being widely used at the present moment, there are still many restrictions exist as stated below [13], [30]:

(1). Sparsit: Collaborative filtering recommendation method usually will use large amount of product or items as evaluation and recommendation targets. Take Amazon.com as an example[16], as the items purchased by users couldn't possibly exceed one percent of all products, if system were to look users with similar preferences based on users purchasing similar products [11], i.e. using binary values of {0, 1} to express whether to purchase the product, it would be hard for the system to locate consumer with similar preference to assist in recommendation or filtering book when there are very few quantity of books are purchased by the same group of users.

(2). Scalability: When the amount of data in the database is exceedingly huge, the system could become overloaded. For example, using the nearest neighbor to calculate the similarity between users, the complexity of algorithm is proportional to the square of user's historical

trading data or the amount of browsing data. To a recommendation system with users and products exceeding a million, the problem of expansion often can be a serious problem in consuming the calculation resources.

(3). Cold Start: This problem means there is not enough user and product information and therefore it is hard for the recommendation system to calculate the similarity of users. For example, new customers often do not have enough trading information. Hence the collaborative filtering recommendation might not be able to effectively locate the same group or similar group as the content of recommendation.

2.1.3 Hybrid Approach Recommendation

Both content-based recommendation and collaborated recommendation have their restrictions and pitfalls. Therefore, many scholars subsequently proposed the Hybrid Approach Recommendation [9], [26] 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 [6]. This process is the so-called data mining. Data mining is the process of finding trends and patterns in data. The objective of this process is to sort through large quantities of data and discover new information [32]. 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 [16].

2.3 Bayesian Network

Bayesian network Technique can be viewed as a kind of network learning [14]. There are many variations to the problem of network learning, including whether the network structure is known or unknown, and whether the variables in the network is certain or uncertain, to which there are different emphasis in learning. Therefore, network learning was divided into four situations [14]:

(1). Known structure and certain variables: This is called parametric learning. It is in the completion of conditional probability tables for these variables.

(2). Unknown structure but certain variables: This is called structural learning. It is to find out structure from many possible network structures that meets the data description. Subsequent parametric learning is also required.

(3). Known structure but uncertain variables: This will have the problem of hidden variables, and similar to the neural network learning.

(4). Unknown structure and uncertain variables: Here the learning problem becomes very difficult, and there is still no suitable solution as of now.

Knowledge maps are Bayesian network. This is one of the sorting algorithms often used in data mining. Bayesian network is a kind of graphical knowledge expression method that expresses the uncertainty and causal relationship in the knowledge through network graphics [1]. Therefore, Bayesian network can take care of both quantitative and qualitative analyses.

There are many analysis methods for data mining. The four features of Bayesian network as below are distinctly different from the rest [7].

(1). Bayesian network can handle the situation where data is incomplete. To the incomplete data within the domain knowledge, Bayesian network can use some kind of algorithms to compensate the inaccurate prediction due to incomplete data.

(2). Bayesian network can learn the causality between variables. The causality allows the ease of understanding of domain knowledge and distinguishing the influence to the network when interference appears that is helpful in preparing decision and preventing accident.

(3). Bayesian network combines domain knowledge and historical data through Bayesian statistics.

(4). Bayesian network could avoid data overfit problem.

The type of Beyesian network is graphical model that could describe the probability relationship among variables. Combining graphical model with statistical method, it is very suitable for the application in data analysis problems. There are three advantages of using Beyesian network [7], [28]:

(1). Beyesian network could easily handle incomplete data, and Beyesian network provides a knowledge expression method expressing independence relationship.

(2). Beyesian network allows the learning of causality relationship.

(3). Through Beyesian statistical method, Beyesian network could combine the domain knowledge and data.

After the value of each node on the Beyesian network is known, the work a computer must be used to calculate the probability of each node because of the complexity is called inference. There are many related algorithms available currently, and in general are divided into two types, including exact inference and approximate inference methods.

The Beyesian network learning methods can be divided into structural and parametric learning. Structural learning refers to whether the modification of association between variables (i.e. between nodes) and variables are being considered in the network under the situation of both structure and probability are uncertain; whereas the parametric learning refers to the learning of parametric probability in the Beyesian network under the situation of both structure and probability are already certain [3]. The process of the whole Beyesian network learning can be divided into five major steps [7], [28]:

(1). First select the problem domain and the related variables.
(2). Select a quality evaluation equation and search algorithm.
(3). Put the possible Beyesian networks into Quality Measure one by one and evaluate.
(4). Pick out Beyesian network with higher quality.
(5). Find out the best Beyesian network using search algorithem.

3 System Design

3.1 Environment and Requirement Analysis

Before designing a complete system, it is necessary to perform environment and requirement analysis. The results of analysis are described below.

(1). System Environment Analysis:

To the book recommendation system to be designed, its major features are the number of types of books and the number of levels is all significantly large, and the relationship between books is also quite complicated. To be specific, there are a few features for the electronic library:

- It has many book types and books itself.
- There exists semantic problem on books.

The level of book types described in Figure 1 is quite complicated.

(2). System Requirement Analysis:

Two major points can be used for discussion in the analysis of system needs, i.e. reader's needs and library's needs.

- Reader Aspect: This system must be able to assist readers in the selection of books valuable, beneficial, and suitable to the reader through the information related to reader's individual book preference.
- Library Administrator Aspect: This system must be able to assist or replace the library administrator, properly make use of the huge library data, and historical borrowing data files, etc. in order to maximize the effectiveness of library resources, and at the same time reduce readers the barriers encountered in the search of interesting, needed, or potentially needed books.

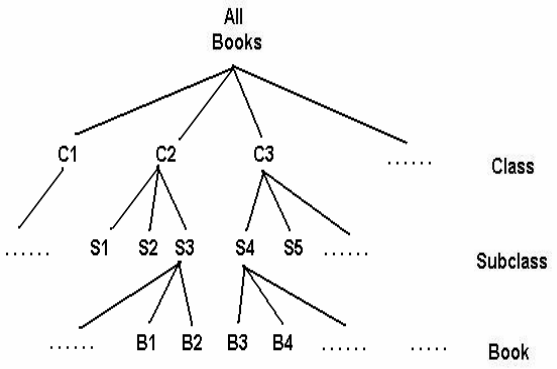


Fig. 1 Level structure of book types

3.2 System Function

Operationally, the book recommendation system could provide the following functions:

(1). Personalized Services:

Online readers could use the historical borrowing record file of known readers based on the same type or similar readers to calculate using the method of this system and make most suitable recommendation to future readers.

Offline readers could use the method of active notification method. For example, after readers borrow books, the system will regularly match the newly introduced library books focusing on readers' reading preference, borrowing habit, type, and range to make sure if the new books are suitable for readers to read. If the answer is yes, the system will automatically email or send short message to notify readers. In this way, readers could easily understand the content of the whole library and find out in time whether there are books suitable for themselves. This could further improve the utilization ratio of the library books and reach the objective of providing convenience to the public.

(2). Library Function Services:

There are four major functions of the library: (1) Keeping and maintaining books. (2) Organize and present. (3) Borrow/return and search. (4) Analyze, consolidate, and broadcast. The design of the system could help librarians in improving execution efficiency, or the so-called library automation. In other words, using new computers and network technologies as tools to assist librarians to perform collection, organizing, borrow/return, and analysis of all kinds of data in order to complete all functions a library should have.

3.3 Reader's Model of Borrowing

This allows the system, while users are in off-line state, to utilize the historical trading information and generate the Beyesian network model users belong to in order to find out "Under the situation certain user has chosen book A he intends to read, it will also select other types of books for his reading." And the variable group of the problem is the status of whether each book in the "historical borrowing database" has been borrowed. After confirming the problem domain and related variables, the Beyesian network can be built up. The task of constructing Beyesian network can be divided into two parts as shown below: (1) the construction of Beyesian network structure, and (2) generation of probability table. Below is the detailed description of the two major parts.

3.3.1 Structure of Bayesian Network

The general steps of constructing Beyesian network can be based on the method proposed by Heckerman [7] which includes the following four steps:

(1). Determine the purpose of constructing Beyesian network.
(2). Make sure all observation data possibly related to the problem are available.
(3). Determine variables valuable to network model under these observation data.
(4). Organize these variables to build up Beyesian network.

The Beyesian network itself has the feature of allowing system constructor subjectively setting data under the situation of insufficient data during system construction stage. This enables the system constructor intuitively assigns data based on the past experience during the Beyesian network construction stage. Beyesyian network consists of several nodes and directional links. Each link and its parent node and subnode at the two ends are treated as one body and turned into a rule.

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 Beyesian network, the reader borrowing history record data shown in Table 1 is converted into matrix of borrowing actions and borrowed books as shown in Table 2 for subsequent ease of calculation of related probabilities.

We use Apriori to analyze the association between books from the matrix in Table 2. 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.

Table 1 The table of reader's book borrow history recorder

Trans. ID	Reader ID	Book ID	Amount	Date
000031	0005	A	1	060722
000031	0005	C	1	060722
000032	0006	B	1	060724
000033	0007	A	1	060725
000033	0007	B	1	060725
000033	0007	E	1	060725
000034	0008	D	1	060801

Table 2 The table of the amount of reader's book borrowed

Trans. ID	Reader ID	A	B	C	D	E
000031	0005	1	0	1	0	0
000032	0006	0	1	0	0	0
000033	0007	1	1	0	0	1
000034	0008	0	0	0	1	0

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 Beyesian network construction rule stated above, the associations of reader historical borrowing data are examined one by one to generate the Beyesian network structure as Figure 2. We utilize MSBN to complete the construction of Beyesian network. As MSBN provides user a set of complete program library, the VB or C/C++ programming languages can be used easily to manipulate Beyesian network and at the same time to convert the constructed Beyesian network into XML format for subsequent program development and operation.

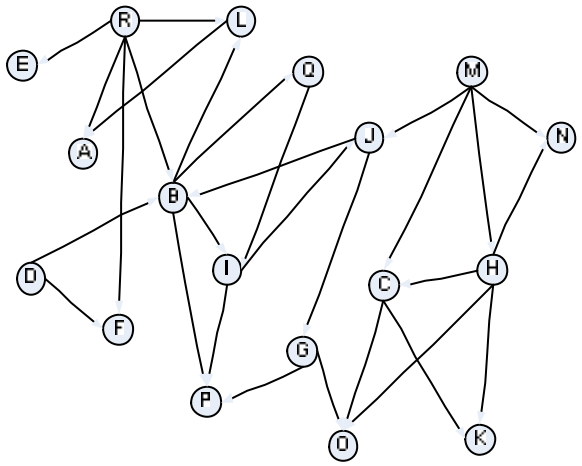


Fig. 2 The structure of Bayesian Network

3.3.2 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.

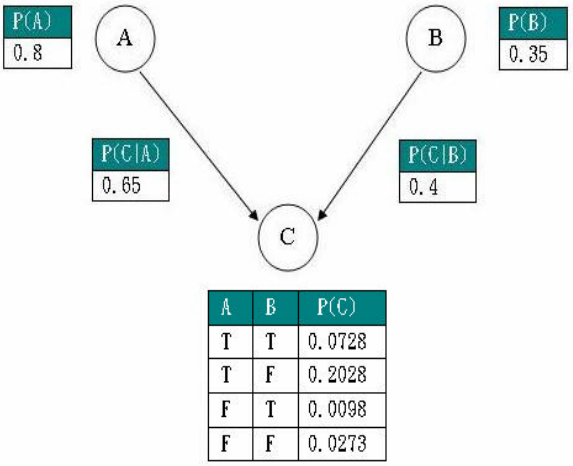
The calculation of Bayesian network probability is to use Equation (1), starting from the root node of each network and learn the condition probability table of each node. Using Figure 3 as an example, assume a reader wished to find out the probability table at node C. We must first know the probability 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.

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

$P(X)$: 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	0.0728
T	F	0.2028
F	T	0.0098
F	F	0.0273

Fig. 3 The structure and probability of Bayesian network

3.4 Personalized Recommendation System

In most examples, recommendation problem can be sorted based on the following three criteria:

(1). Whether the customer to be recommended is a whole or a specific customer group.

(2). Whether the purpose is to predict how much a specific customer likes a specific product, or to make sure whether a series of product will attract customers already existent is not of concern.

(3). Whether the recommendation is completed within a specific time frame or repetitively in progress.

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.

In the issue of solving book recommendation, the system helps a specific user to look for books suitable for their preference. Assume the flow to solve the recommendation problem is divided into six steps. Figure 4 illustrates the procedure of the recommendation system.

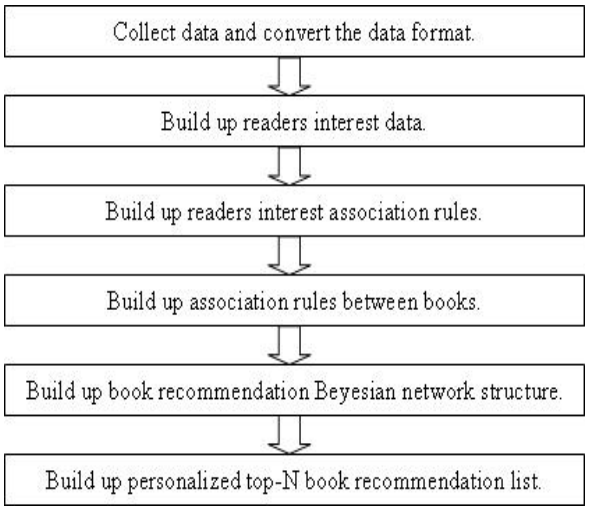


Fig. 4 The procedure of the recommendation system

After completion of Bayesian network recommendation structure, different reader will have different Bayesian network probability table. Assuming Figure 5 is the Bayesian network of reader A and Figure 6 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.

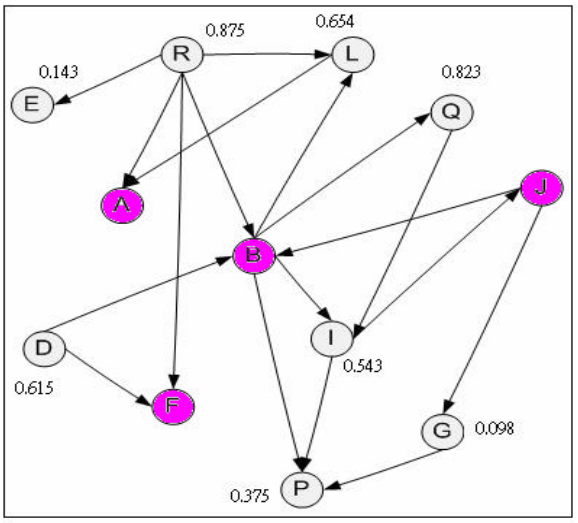


Fig. 5 Book recommendation Bayesian network of reader A

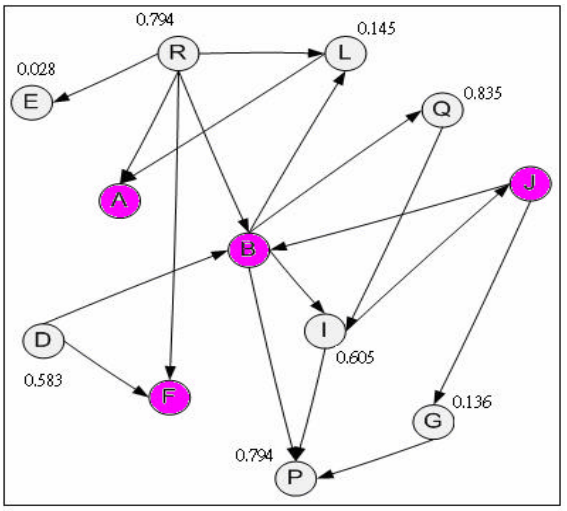


Fig. 6 Book recommendation Bayesian network of reader B

4 Example Analysis

4.1 System Environment

The design of personalized library book recommender system has been completed and the personalized service system related information table is also generated. The description regarding system development tool and environment platform as well as the personalized information service interest file are as shown below.

(1). System development tool and development environment.

- CPU: AMD Turion(tm) 64 CPU 1.8GHz
- RAM: 1GB
- HD: 60GB

(2). Software equipment and development tool.
- Windows XP Professional with SP2
- Microsoft SQL Server 2005
- Microsoft Internet Information Server 6.0 (IIS 6.0)

(3). Related information table.
- Reader borrowing trade record information table
- Reader basic information table
- Reader interest information table
- Library book record information table

4.2 Data Preprocessing

As the data collected from database might have noises or incompatible format problem, there is a need to perform suitable preprocessing. The data is rearranged into the dataset suitable for analysis before inserting into chosen algorithm for analysis.

The trading set required for collaborated filtering recommendation method is based on a type as a trading record. The types of Chinese Book Classification Method are used as classification standard. The type where each book is located is based on the classification number on the borrowing index. Each item in the trading represent a reader ever borrowed the type of book and the number of times the reader borrowed the type of books.

4.3 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. Here only part of the readers is listed for reference as shown in Table 3. 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 as shown in Table 4.

Subsequently the readers' book interest data table as shown in Table 5 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 the algorithm. Table 5 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 6 and Table 7.

Table 3 Data table of readers' number of times of borrowing type of books

Class / Reader ID	0	1	2	3	4	5	6	7	8	9	sum
55	0	0	0	0	0	21	0	0	0	0	21
98	0	0	0	3	0	0	0	0	0	3	6
132	5	0	0	6	0	5	0	0	0	0	16
158	6	3	0	0	0	0	0	0	7	0	16
190	0	0	0	0	0	9	0	0	0	2	11
239	0	0	0	8	2	0	0	10	0	0	20
248	2	0	0	0	0	0	0	0	6	0	8
327	2	0	0	0	0	0	2	0	0	0	4
334	0	0	31	0	0	2	0	0	0	3	36
376	0	5	0	0	6	0	14	0	0	0	25
421	0	0	3	0	0	1	0	0	1	0	5
432	0	7	3	3	0	0	7	0	0	0	20

Table 4 Values of Readers' interest in type of books

Class / ReaderID	0	1	2	3	4	5	6	7	8	9
55						1				
98				0.5						0.5
132	31.25			37.5		31.25				
158	0.375	0.1875							0.4375	
190						0.8182				0.1818
239				0.4	0.1			0.5		
248	0.25								0.75	
327	0.5						0.5			
334			0.8611			0.556				0.0833
376		0.2			0.24		0.56			
421			0.6			0.2			0.2	
432		0.35	0.15	0.15			0.35			

Table 5 Table of readers' interest in type of book.

Reader_ID	Interest_Class
55	5
98	3, 9
132	0, 3, 5
158	0, 8
190	5
239	3, 7
248	0, 8
327	0, 6
334	2, 5
376	6
421	2
432	1, 6

Table 6 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 7 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, this model must convert all book documents in the book database into Bayesian network structure. 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 called complete Bayesian network structure. Figure 7 illustrated the Book Beyesian network structure diagram. Through the comparison of association between nodes of the complete Bayesian network structure, we could discover the Bayesian network node style after learning. Based on the causality association the recommended threshold value of the node could be predicted to see whether it has reached the standard of recommendation. Finally, the book data represented by the node is recommended to the reader. The steps of constructing Beyesian network primitive model is as shown below in order:

- Newly added nodes: Building up new module. Convert to the node in the corresponding module set based on all nodes in the book database.
- Build up association: Build up links with causality association for all nodes in pairs in the corresponding module set.
- Build up initial Beyesian network structure: Build up the links with causality association based on the pairs for all nodes until all nodes of the tree shape structure are interlinked to complete the initial Beyesian network structure.
- Build up the probability value for each node.
- Determine whether the probability value exceeds preset threshold value based on the probability values exhibit in the estimation table : The node could be the book information the library user needs if the value exceeds the threshold value and should be recommended to the user.
- To clearly determine the state of the node, we also could convert the tabulated probability values into visual estimation bar chart. This could be quicker and more accurate in the processing efficiency.
- Evaluate the whole Beyesian network structure: Consolidate steps (5) and (6) by combining the associated probability value and bar chart with simplified Beyesian network structure to complete the evaluation of whole beyesian network structure. Use more obvious visual method to provide back end processing with more comparison processing function.

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.

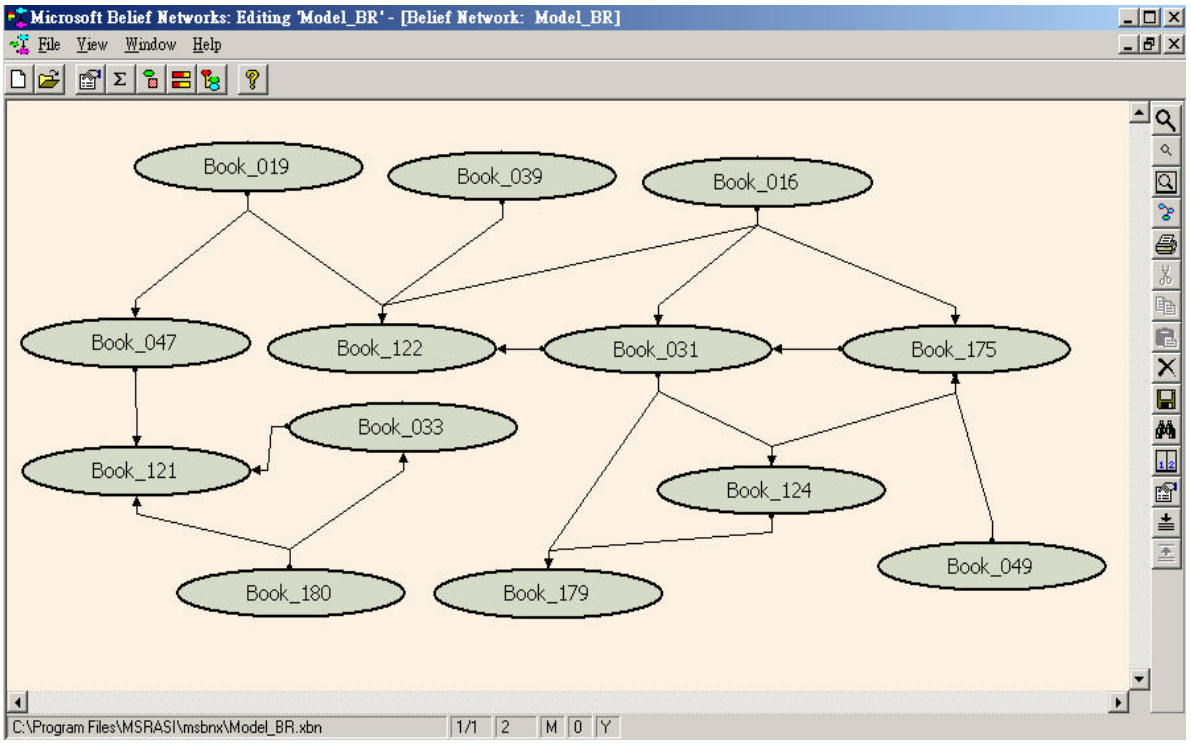


Fig. 7 Book Beyesian network structure diagram

(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 8. Then, the probability value of each book displayed in the evaluated bar chart as shown in Figure 9 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.

4.4 Analysis of Result

4.4.1 Evaluation Index

This study utilizes the precision and coverage measures which are often used as index to evaluate the proposed recommendation system. The precision is the percentage of satisfactory items for readers among recommended items, and its calculation method is as shown in the equations (2) below. The coverage is the percentage of correct recommendation items among satisfactory items for readers, and its calculation method is as shown in the equations (3) below. However, the numbers of recommendation items should affect the precision and coverage measures. When the numbers of recommendation items increase, the precision will decrease and the coverage will increase. Therefore, it is not enough to evaluate the performance of recommendation system if this study only respectively utilizes the precision and coverage measures.

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 (4) below. The F1 coefficient includes the precision and coverage measures, so it can objectively be utilized to evaluate the benefit of recommender system for readers.

$$\text{Precision}=\frac{|\text{Relevent} \cap \text{Recommendation}|}{|\text{Recommendation Items}|} \quad (2)$$

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

$$\text{F1}=\frac{2 * \text{Precision} * \text{Coverage}}{\text{Precision}+\text{Coverage}} \quad (4)$$

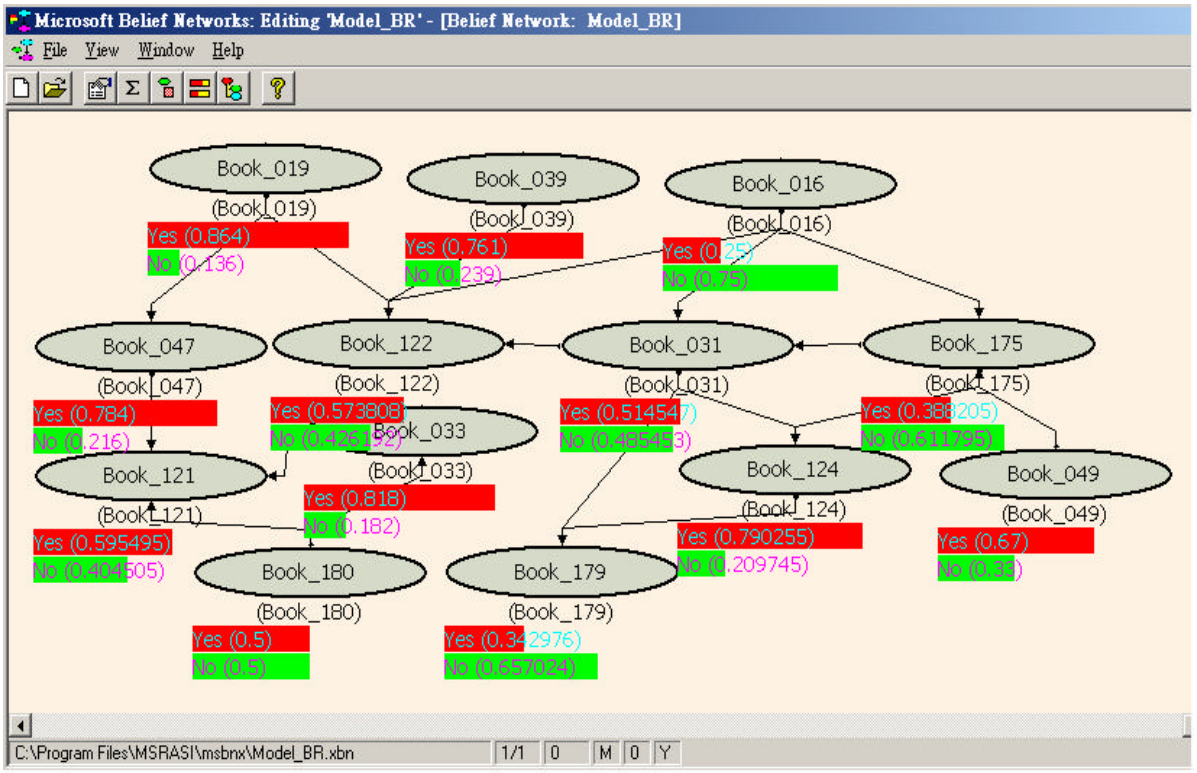


Fig. 8 Bayesian network structure of book recommendation

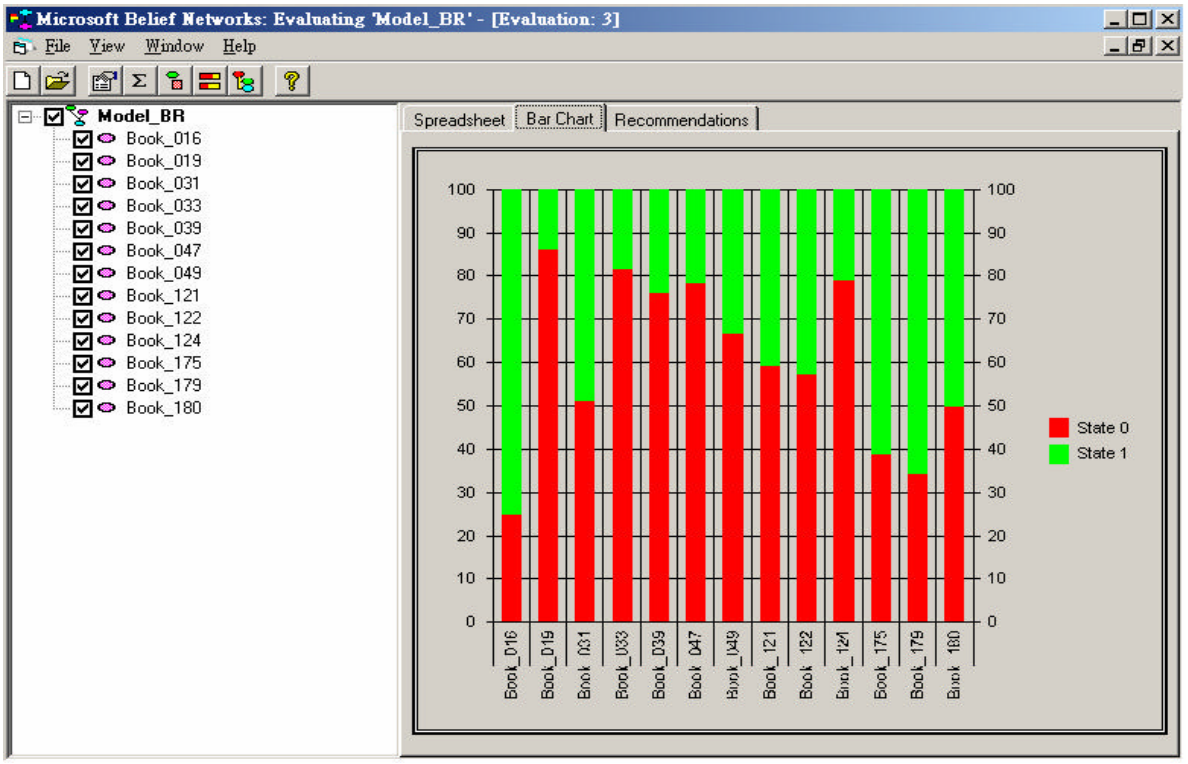


Fig. 9 Evaluation bar chart

Table 8 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.

4.4.2 Performance Analysis

This study focuses on each reader and evaluates his satisfaction of recommendation list. Table 9 is the sample data of each reader's satisfaction. 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

Table 8 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

Table 9 Sample data of each reader's satisfaction

Readers	Recommendation List	Satisfactory List	Satisfaction
1	5	4	0.8
2	5	4	0.8
3	5	3	0.6
4	5	4	0.8
5	5	5	1.0
6	5	2	0.4
7	5	4	0.8
8	5	3	0.6
9	5	4	0.8
10	5	3	0.6

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

4.5 Directions of Future Research

Recommender system is a personalized service tool based on information filtering [5], [10]. The most commonly used recommendation system methods are collaborative filtering, content-based filtering, and hybrid filtering [20]. However, these recommendation methods have their own respective restriction and drawbacks. For example, content-based recommendation method could not automatically analyze and judge the attributes of media including sound, picture, and video [17]. This method could not also find out potential information that user has never attempted in the past but could be interested [17]. In addition, collaborative filtering would be hard to locate users with similar preference to assist in recommendation or filtering product when there are very few quantity of products are purchased by the same group of users [13], [30]. Therefore, the directions of future research will focus on proposing new approaches that substantially improve the performance of collaborative filtering, content-based filtering, and hybrid filtering.

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] J. Ma and K. Sivakumar, Privacy-Preserving Bayesian Network Parameter Learning, *WSEAS Transactions on Information Science & Applications*, Vol.3, No.1, 2006, pp.1-6.

[2] M. Adda and C. Djeraba, Recommendation Strategy Based on Relation Rule Mining, *WSEAS Transactions on Information Science & Applications*, Vol.3, No.1, 2006, pp.7-14.

[3] A. Heni, M.N. Omri, and A.M. Alimi, Fuzzy Knowledge Representation Based on Possibilistic and Necessary Bayesian Networks, *WSEAS Transactions on Information Science & Applications*, Vol.3, No.2, 2006, pp.224-231.

[4] C. Vassiliou, D. Stamoulis, D. Martakos, and S. Athanassopoulos, A Recommender System Framework Combining Neural Networks & Collaborative Filtering, *WSEAS Transactions on Information Science & Applications*, Vol.3, No.7, 2006, pp.1202-1207.

[5] S. Segrera and M.N. Moreno, Application of Multiclassifiers in Web Mining for a Recommender System, *WSEAS Transactions on Information Science & Applications*, Vol.3, No.12, 2006, pp.2471-2476.

[6] R.C. Wu and R.S. Chen, The Application of Data Mining Technology for Intelligent Product Quality Analysis Improvement System, *WSEAS Transactions on Information Science & Applications*, Vol.4, No.4, 2007, pp.693-699.

[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] U. Fayyad and R. Uthurusamy, Data mining and knowledge discovery in databases, *Communications of the ACM*, Vol.39, No.11, 1996, pp.24-26.

[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.

[20] F. Gao, C. Xing, X. Du, and S. Wang, Personalized Service System Based on Hybrid Filtering for Digital Library, *Tsinghua Science and Technology*, Vol.12, No.1, 2007, pp.1-8.

[21] W.P. Lee, C.H. Liu, and C.C. Lu, Intelligent Agent-based Systems for Personalized Recommendations in Internet Commerce, *Expert Systems with Applications*, Vol.22, No.4, 2002, pp.275-284.

[22] C.C. Nauna, and S.M. Braxton, Developing Recommendations for Consortial Cataloging of Electronic Resources: Lessons Learned, *Library Collections, Acquisitions, & Technical Services*, Vol.29, No.3, 2005, pp.307-325.

[23] J. Bollen, M.L. Nelson, G. Geisler, and R. Araujo, Usage Derived Recommendations for a Video Digital Library, *Journal of Network and Computer Applications*, Vol.30, No.3, 2007, pp.1059-1083.

[24] P. Symeonidis, A. Nanopoulos, A.N. Papadopoulos, and Y. Manolopoulos, Collaborative Recommender Systems: Combining Effectiveness and Efficiency, *Expert Systems with Applications*, Vol.34, No.4, 2008, pp.2995-3013.

[25] C.L. Al-Qallaf and H.M. Al-Azmi, Information Technology in Public Libraries in Kuwait: A First Study, *Intl. Inform. & Libr. Rev.*, Vol.34, No.4, 2002, pp.289-308.

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

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] R.S. Chen and R.C. Wu, Design an Intelligent Quality Analysis Control System Based on Data Mining Technology for Semiconductor Packaging Industry, *WSEAS Transactions on Computers*, Vol.5, No.10, 2006, pp.2479-2487.

[33] Y. Yu, Y. Yu, H. Lin, and C. Chen, Mining on Web Logs for Recommendation, *WSEAS Transactions on Computers*, Vol.5, No.9, 2006, pp.1818-1822.

[34] H.R. Chen, Evaluating Recommendation Service Mechanisms for Asynchronous Digital Learning Systems, *WSEAS Transactions on Computers*, Vol.6, No.3, 2007, pp.612-617.