



# Using data mining technology to provide a recommendation service in the digital library

Data mining  
technology

711

Chia-Chen Chen

*Department of Information Management, Tunghai University,  
Taiching, Taiwan, and*

An-Pin Chen

*Institute of Information Management, National Chiao-Tung University,  
Hsinchu, Taiwan*

Received 15 September 2006  
Revised 30 October 2006  
Accepted 31 October 2006

## Abstract

**Purpose** – Since library storage has been increasing day by day, it is difficult for readers to find the books which interest them as well as representative booklists. How to utilize meaningful information effectively to improve the service quality of the digital library appears to be very important. The purpose of this paper is to provide a recommendation system architecture to promote digital library services in electronic libraries.

**Design/methodology/approach** – In the proposed architecture, a two-phase data mining process used by association rule and clustering methods is designed to generate a recommendation system. The process considers not only the relationship of a cluster of users but also the associations among the information accessed.

**Findings** – The process considered not only the relationship of a cluster of users but also the associations among the information accessed. With the advanced filter, the recommendation supported by the proposed system architecture would be closely served to meet users' needs.

**Originality/value** – This paper not only constructs a recommendation service for readers to search books from the web but takes the initiative in finding the most suitable books for readers as well. Furthermore, library managers are expected to purchase core and hot books from a limited budget to maintain and satisfy the requirements of readers along with promoting digital library services.

**Keywords** Digital libraries, Data collection, Electronic document delivery, Libraries, Cluster analysis, Programming and algorithm theory

**Paper type** Research paper

## 1. Introduction

The number of people using the internet has increased dramatically because of the widely accepted web environment. The internet has also rapidly accumulated a huge mass of data and has grown to be one of the most powerful means of information storage. In such a web environment, the concept of the digital library is fascinating as it includes information technology which could produce plenty of complex data for end-users. The emergence of digital libraries storing digitized data makes it possible to search more easily and conveniently. Traditionally, the library used to play a passive role in that it merely provided books for borrowing. It is a crucial subject, however, for a library manager to think about how to guide readers to find what they want in an aggressive way and promote the borrowing rate at the same time.



---

This paper specifies how digital libraries can benefit from immense digital resources to enhance the quality of various services, and an approach is presented to identify valuable and relevant online resources. In past research, most researchers have analyzed the content of digital documents. Then, they tried to discover the relationship between documents and documents, as well as between documents and users. However, there are more and more formats for digital publications such as audio, video, picture, etc. Under these circumstances, it is hard to analyze the keywords or content in themselves so as to refine users' recommendation information.

A new recommendation system architecture is established in this paper to enable customized services and management in the digital library. The association rules and clustering along with the data mining methods have been applied to discover the most adaptive readers of a book. First, loan records in the digital library are clustered according to some characteristics of readers. The proposed approach utilizes the automatic clustering feature of the Ant Colony Clustering Algorithm to form a user group with similar properties. Second, based on minimal support and confidence, the Apriori Algorithm is used to exhibit the ability of locating the associated rules between subjects to generate recommending rules. The association rules will judge which books borrowed by the readers in the same cluster are used as the basis of book recommendation. Finally, an automatic online recommendation system is proposed.

This paper not only constructs a real-time recommendation service for readers in searching books from the web, but also takes the initiative in finding the most suitable books for readers. Furthermore, library managers are expected to purchase core and hot books from a limited budget to maintain and satisfy the requirements of readers along with promoting digital library services. Digital libraries could provide better services via the seamless integration of diverse approaches towards collecting, organizing, storing, accessing, and applying knowledge.

## 2. Literature review

This section provides a general definition of data mining, which is the main component of the proposed methods.

### 2.1 Knowledge discovery in databases

The definition of knowledge discovery in databases (KDD), given by Fayyad *et al.* (1996), is defined as the "nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data". In their view, the term "knowledge discovery in database" is used to denote the entire process of turning chaotic data into valuable knowledge. They also illustrated that the whole KDD process covered several key steps: data cleaning, data reduction and transformation (data integration), data mining, pattern evaluation and then knowledge discovered. The overall process is outlined in Figure 1 (Han and Kamber, 2001). It is without doubt that data mining is considered as a central step in the process that involves extracting patterns from data (Chang and Chen, 2006). Additional steps are also essential to make certain that what we extract from data is useful knowledge. This paper also follows additional steps to demonstrate the proposed method in section 3.

## 2.2 Data mining

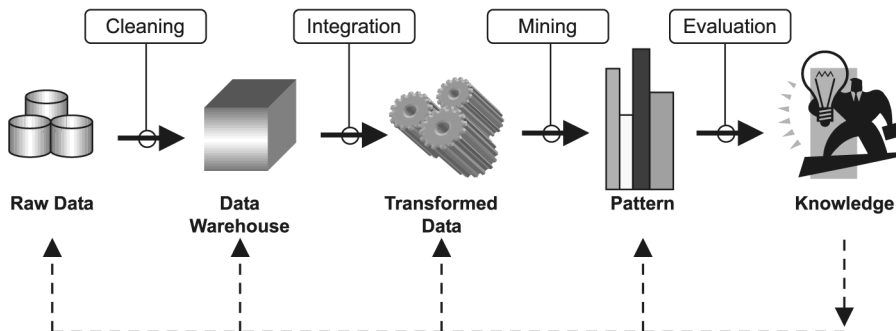
The illiberal definition of data mining is the application of specific algorithms to uncover useful information from a large degree of data, and its purpose is to explore interesting knowledge from a database, data warehouse, or some other large information storage unit (Han and Kamber, 2001). From a technical viewpoint, it combines a method of gathering and cataloging information then proceeds to generate rule-like knowledge from a large amount of data. The more common model functions in the current data mining algorithms include classification, regression, clustering, association rules, rule generation, summarization, dependency modeling, and sequence analysis (Mitra *et al.*, 2002).

Actually, data mining has been applied to various domains, such as customer service support, decision support, web intelligence, etc. (Fong *et al.*, 2002; Han and Chang, 2002; Hui and Jha, 2000). The most well-known example is “bear and diapers”, where the giant supermarket chain WalMart wanted to know that which items were sold together from their huge sales records. They analyzed billions of transaction records and finally found that bears and diapers punched together could stimulate a purchase.

In this paper, we applied association rules and clustering algorithms to extract similar interests of readers and recommend books to them. These are briefly explained below.

**2.2.1 Association rules.** Depending on the above-mentioned discussion of data mining applications, in recommender systems, one of the widely used examples of data mining is the discovery of association rules, especially market basket analysis. Huge amounts of customer purchase data are collected daily at the checkout counters of shopping malls and retailers are interested in purchasing the behavior of their customers. This technique, association rules, frequently found co-purchase items. Moreover, the uncovered relationships can be represented in the form of association rules. This provided retailers an opportunity for cross-selling their products to customers.

**2.2.2 Clustering.** Clustering techniques work by identifying groups of users who appear to have similar preferences and dividing groups who have very different preferences. Unlike classification, the class label of each group is unknown. This is the way to naturally segment data into undefined groups, called clustering. In contrast, classification is assigning data into defined groups (Edelstein, 2000). Briefly, a good clustering method produces high quality clusters with high intra-class but low



**Figure 1.**  
Overview of KDD process

inter-class similarity. However, how good a cluster is ultimately depends on the opinion of the user.

### 2.3 Association rules

The Apriori Algorithm was proposed by Agrawal and Srikant (1994), and is a famous algorithm in the mining association rule area. Those things that appear simultaneously in certain events or data are called associations. Association rules mining aims to discover interesting associations or correlation relationships from large data sets (Han and Kamber, 2001). In the research of Berry and Linoff (1997), they were applied to analyze market baskets and would indicate which items should be bought at the same time. *Support* and *confidence* are the two important parameters required to generate effective association rules. Support is the number of transactions with all the items in the rule, and confidence is the ratio of the number of transactions with all the items in the rule to the number of transactions with just the items in the condition (Berry and Linoff, 1997). Therefore, the support ( $A \Rightarrow B$ ) can be described as  $P(A \cup B)$ , and the confidence ( $A \Rightarrow B$ ) can be described as  $P(B|A) = A \cup B/A$ .

The major work in mining association rules is to find all the large item sets. There is a great amount of research on how to determine these large item sets, with the Apriori Algorithm being the representative algorithm. It was first proposed by Agrawal and Srikant to discover association rules (Agrawal *et al.*, 1993). However, these algorithms must scan databases many times to find the large item sets. Moreover, when they generate a candidate item set, the apriori-gen function wastes a lot of time checking whether its subsets are large or not (Han and Kamber, 2001). Agrawal and Srikant (1994) proposed the AprioriHybrid, which scales linearly with the number of transactions. It also has excellent scale-up properties with respect to the transaction size and the number of items in the database.

### 2.4 Ant Colony Optimization Algorithm

The natural metaphor on which ant algorithms are based is that of ant colonies. Real ants are capable of finding the shortest path from a food source to their nest and they communicate with others by exploiting *pheromone* information. While searching for food, ants deposit pheromones on the ground, and in all probability, follow pheromones previously deposited by other ants. As more and more ants pass by the same path, the pheromone on the path is increased, but the pheromones would decay with time on other paths. The ants' behavior is shown by Figure 2 (Dorigo and Gambardella, 1997a, b).

Dorigo *et al.* (1996) proposed an ant colony optimization algorithm (ACO), which has been applied successfully to several combinatorial optimization problems and produced many promising evolutions. Amongst these successes are the traveling salesman problem (TSP) (Dorigo and Gambardella, 1997a) and the quality of service problem (QoS) (Caro and Dorigo, 1998; Leguizamon and Michalewicz, 1999). There are three ideas from the natural ant colony that have been transferred to the artificial ant colony:

- (1) the preference for paths with a high pheromone level;
- (2) the higher rate of growth in the amount of pheromones on shorter paths; and
- (3) the information exchanged among ants (Dorigo and Gambardella, 1997a, b).

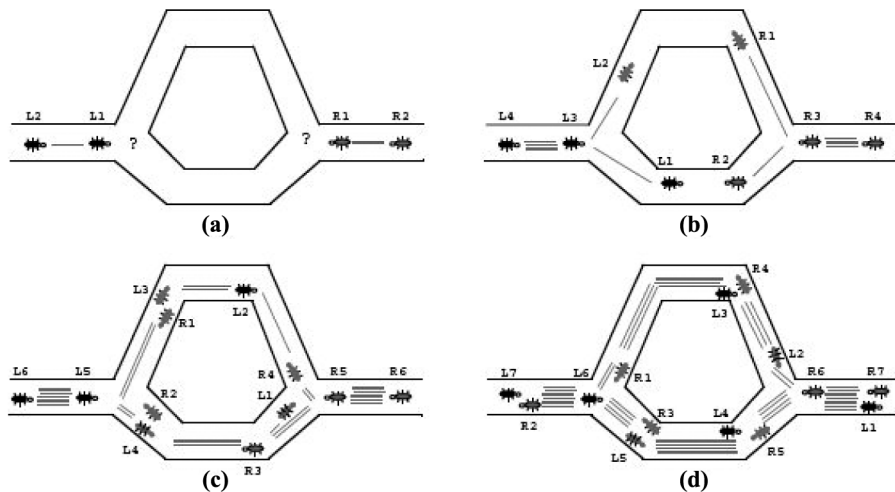


Figure 2.  
The behavior of real ants

The ACO algorithm can be summarized as follows:

- Step 0: Set parameters and initialize pheromone trails.
- Step 1: Each ant constructs its solution.
- Step 2: Calculate the scores of all solutions.
- Step 3: Update the pheromone trails.
- Step 4: If the best solution has not been changed after some predefined iterations, terminate the algorithm; otherwise, go to Step 2.

### 2.5 Digital library using data mining

In his research, Borgman considered that digital libraries are a set of electronic resources and associated technical capabilities for creating, searching and using information (Borgman, 1999). Due to the popularity of electronic commerce and personalized trends, the technique of data mining is also widely used to analyze consumers' behavior. This is to determine personal preference and to provide related product information in order to raise the level of consumption (Agrawal *et al.*, 1993).

Applying data mining techniques in a digital library service is also considered a trend as it can automatically filter out useful knowledge by user profiles and the function of statistical analysis. For example, filtering out popular topics from each borrowing history can help promote book circulation in the library. The digital library can also use functions of statistical analysis along with data mining to provide information on books, articles, topics and other long-term personal services for promoting circulation.

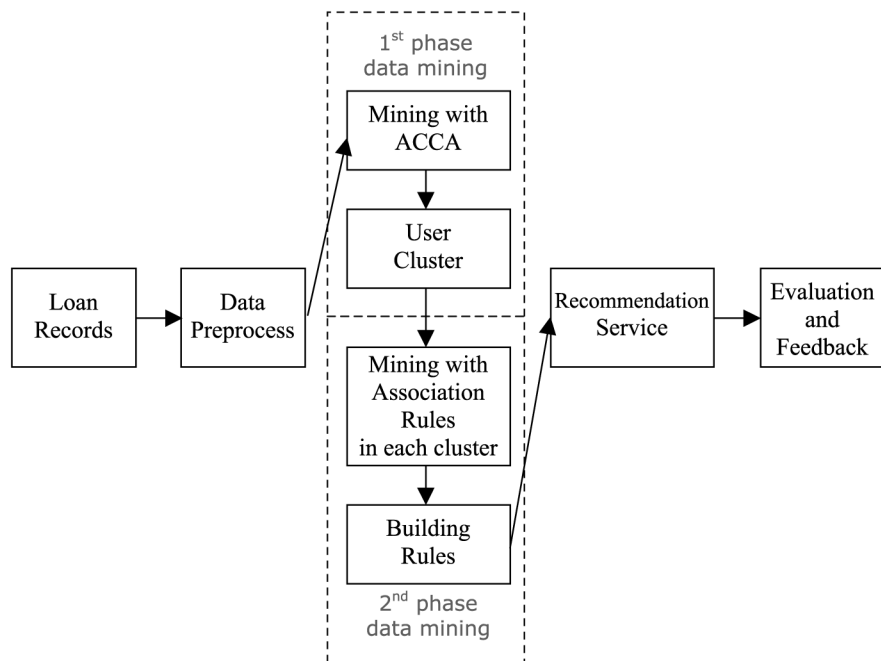
## 3. Problem definition and methodology

### 3.1 Problem definition

In the digital era, people can get information easily because of the development of information technology and the internet. They can discover interesting information

and digital content via surfing on the internet. When users access the digital library, they often input appropriate keywords and use the “Search” function to discover the information that they want. However, the results do not always satisfy users. In the past, there has been some research on searching by keywords. However, the keywords were provided by document authors (or publishers, librarians, and indexers) (Rocha and Bollen, 2001), and could not necessarily reflect the semantic expectations of users. Therefore, further research tried to support some recommendation for users to aid in keyword searching. In 1999, Luis led a project named the Active Recommendation Project (ARP) at the Los Alamos National Laboratory. It was developing research on recommendation systems for large databases and the worldwide web (WWW), which adapt to the expectations of users (Rocha, 1999). Heylighen and Bollen (2002) proposed recommendation system based on Hebbian algorithms.

In this paper, we proposed a two-phase data mining recommendation service through analyzing the access behavior of users. In the first phase, we used the Ant Colony Clustering Algorithm as the data mining method and separated users into several clusters depending on access records. Users who have similar interests and behavior are collected in the same cluster. In the second phase, we further analyzed the user records in the same cluster. We used association rules as the data mining method and discovered the associations among users’ interests and access behavior. Then, the rules for the recommendation service were built. The experimental process employed is shown as Figure 3.



**Figure 3.**  
Experimental process of  
two-phase data mining

### 3.2 Methodology

In this paper, we propose a recommendation service model which combines the Ant Colony Clustering Algorithm (Chen and Chen, 2006) with association rules to discover readers with the same interests. A detailed description of the data mining methods follows.

**3.2.1 Ant Colony Clustering Algorithm.** The proposed method uses the Ant Colony Clustering Algorithm, and its main process is shown in Figure 4.

The first step is to initialize the parameters. A set of artificial ants is positioned on the first job according to an initialization rule (e.g. randomly). Each ant constructs its own cluster. Once the ants have completed their clusters, each cluster's variance ( $CV_{intra}$ ) is calculated. The percentage of the farthest nodes are chosen to be regrouped into the cluster with the shortest distance to  $O_{center}(M)$ . If the new variance ( $CV'_{intra}$ ) is smaller than  $CV_{intra}$ , that means the nodes in the updated cluster are more similar than the nodes in the previous cluster. While applying new clusters, an ant will simultaneously update the amount of pheromone on its visited paths (by applying the local updating rule). After all of the ants have built solutions, the pheromone trails on the paths of the global best cluster are modified again (by applying the global updating rule) up to the current iterations. The process is terminated after predefined iterations.

The complete Ant Colony Clustering Algorithm is summarized as follows:

- Input:  $n$  nodes.
- Output: the number of predefined clusters.
- Step 0. Initialize the parameters, which include the number of ants  $m$ , parameters  $q_0$ ,  $\beta$ , the pheromone decay parameter  $\alpha$ ,  $\rho$ , the percentage of farthest nodes chosen to regroup  $\gamma$ .
- Step 1. Place  $m$  ants on the nodes randomly.
- Step 2. Group the collected nodes into clusters. An ant  $k$  at node  $r$  chooses the node  $s$  to move along the nodes that do not belong to its working memory  $M_k$ . The state transition rule is applied by the following probabilistic formula, which provides a direct way to balance between exploration of new edges and exploitation of *a priori* and accumulated knowledge about the problem:

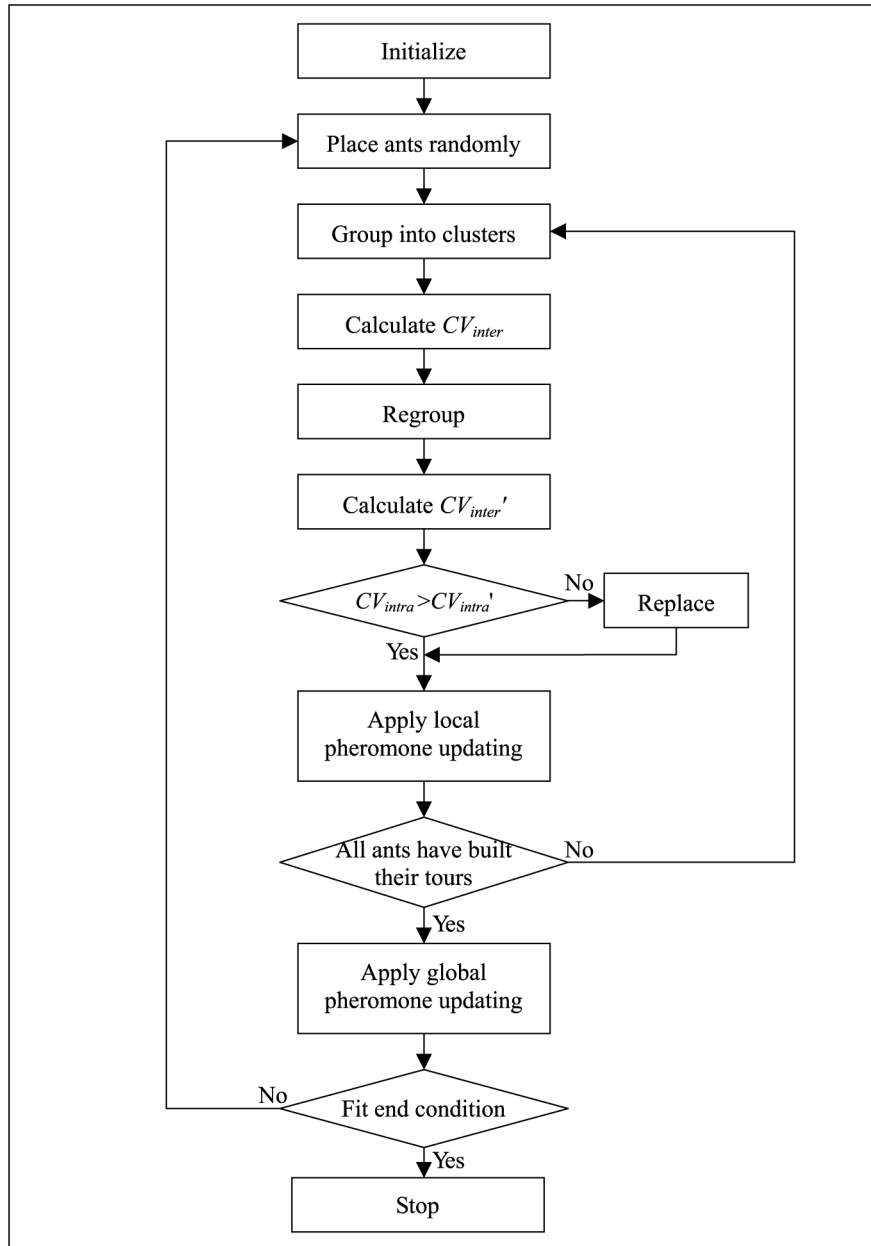
$$s = \begin{cases} \arg \max_{u \notin M_k} \{[\tau(r, u)] \cdot [\eta(r, u)]^\beta\} & \text{if } q \leq q_0 \quad (\text{exploitation}) \\ S & \text{otherwise} \quad (\text{exploration}) \end{cases}, \quad (1)$$

where  $S$  is a random variable selected according to the probability distribution given in equation (2), which favors edges that are shorter and have a higher level of pheromone trail:

$$p_k(r, s) = \begin{cases} \frac{[\tau(r, s)] \cdot [\eta(r, s)]^\beta}{\sum_{u \notin M_k} [\tau(r, u)] \cdot [\eta(r, u)]^\beta} & \text{if } u \notin M_k \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

If  $p_k(r, s) \geq \bar{p}_k$ , ant  $k$  collects the node  $s$ .

- Step 3. Calculate  $O_{center}(M)$  and  $CV_{intra}$  of each cluster. The nodes that are in  $\gamma$  are chosen to be regrouped to the closest group.



**Figure 4.**  
Flow chart of the Ant  
Colony Clustering  
Algorithm



- Step 4. Calculate  $CV'_{intra}$ . If  $CV'_{intra} > CV_{intra}$ , the replaced result is adopted and local pheromone is updated on all edges according to:

$$\tau(r, s) = (1 - \rho) \cdot \tau(r, s) + CV_{inter}^{-1}. \quad (3)$$

- Step 5. Global pheromone updating is intended to allocate a greater amount of pheromone. When all ants have built their tours, global pheromone is updated according to:

$$\tau(r, s) = (1 - \alpha) \cdot \tau(r, s) + CV^{-1}, \quad (4)$$

where  $CV$  is the sum of the smallest  $CV_{intra}$  in all  $CV_{intra}$ .

- Step 6. The process is iterated until the end condition is met.

**3.2.2 Association rules.** In this paper, we used the Apriori Algorithm to discover the association rules. Generally, there are two steps to mining the association rules.

Step 1: Find all the large item sets

- (1) The supports of large item sets should be larger than the minimal supports defined by users.

$$\text{Support}(AB) = p(AB).$$

- (2) If there are  $k$  items in a large item set, then we call it a large  $k$ -item set.

Step 2: Use the large item sets generated in the first step to generate all the effective association rules

- (1) Calculate the confidence:

$$\text{Confidence}(A \Rightarrow B) = (P|B) = \text{support\_count}(AB) / \text{support\_count}(A).$$

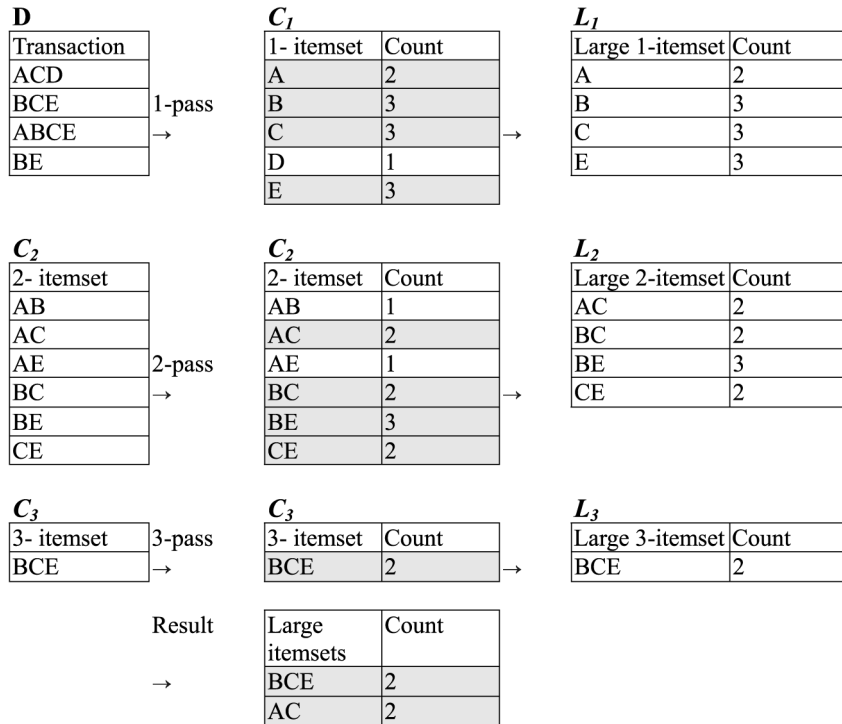
- (2) If the confidence of association rule is larger than the minimal confidence defined by users, then it is effective.

The algorithm terminates when no more candidate item sets can be constructed for the next round. An example of the Apriori Algorithm which sets minimum support at 50 percent and minimum confidence at 80 percent is shown in Figure 5.

#### 4. Implementation

Before proceeding to data mining, the source of data (loan records in library) needs to be preprocessed. The completeness of source data is one of the keys to successful data mining. This shows that data preprocessing should spend most of the time in the whole KDD process.

The major tasks in data preprocessing include data cleaning, data integration and data transformation. In order to ensure the degree of data purity, it is necessary to identify outliers and smooth out noisy data. The loan patterns are found out by using clustering and association rules. Since an association rule is an expression of the form  $X \Rightarrow Y$ , where  $X$  and  $Y$  are sets of items, the main goal is searching for association relationship between two sets of items. In this paper, we take the campus digital library as an example. There are thousands of loan records every year and the unavailable data should be eliminated in order to reduce computing run-time. At this point, if a reader borrowed only one book in half a year, the loan record would be considered as useless data in this system. Therefore, we clean the useless data at the beginning.



**Figure 5.**  
Example of the Apriori  
Algorithm

The first phase for the data mining method is to cluster data in each group. We assumed that all data is clustered into three groups. It is necessary to input the reader's department, sex, role on campus (undergraduate student, graduate student, staff, or faculty), and books borrowed. After iterating several rounds by the Ant Colony Clustering Algorithm, the three groups can be output.

The second phase for the data mining method is to find out the patterns of relationships in each cluster by association rules. Before processing the method, the data must be integrated. As mentioned, we can take the item set in a customer's basket for a transaction record. Likewise, loan records are regarded as transaction records in a university digital library database. A loan record is defined as a set of records that a reader borrows continuously in a period of time. For example, there are loan records which record a reader's borrowing list, shown in Table I. Each attribute contains a serial number, reader's id, object id, object type, loan date and return date. In order to strengthen the relationship between two items, the records are linked serially until the period of time is over three months between the loan date and the least return date. The digital library integrated data are presented in Figure 6.

Association rules whose support and confidence exceed user-supplied thresholds are output by the Apriori Algorithm (see Figure 7). Then, we can recommend books depending on association rules. The recommendation system architecture is shown in Figure 8.

SNo	Reader_ID	Object_ID	Type	Loan_Date	Return_Date
1	Stu_A_890001	X221237	Book	2004/2/5	2004/2/13
2	Stu_A_890001	VCD000276	VCD	2004/6/1	2004/6/10
3	Stu_A_890001	X253973	Book	2004/6/1	2004/6/10
4	Stu_A_890001	X215859	Book	2004/7/5	2004/7/12
5	Stu_A_890001	VCD000277	VCD	2004/7/12	2004/7/20
6	Stu_A_890001	DVD000535	DVD	2004/7/12	2004/7/20
7	Stu_A_890001	X240492	Book	2004/11/20	2004/11/28
8	Stu_A_890001	X255340	Book	2004/11/20	2004/11/28
9	Stu_A_890001	X243228	Book	2004/12/13	2004/12/22

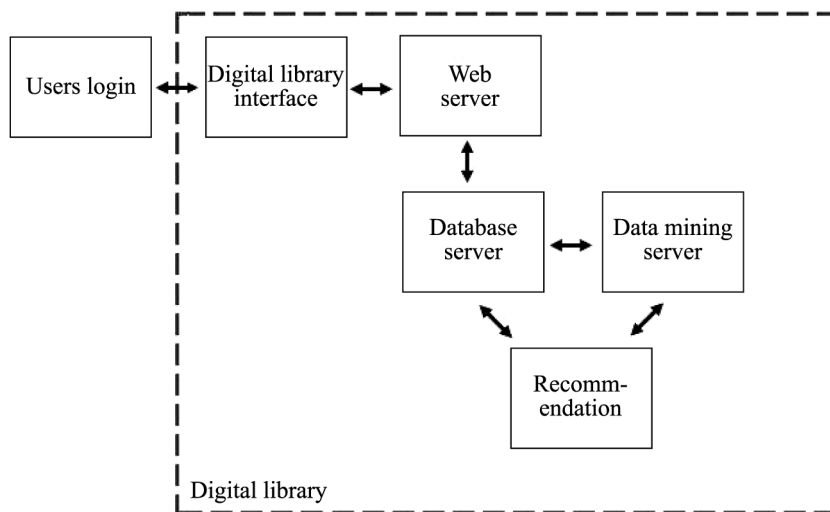
**Table I.**  
Original loan records

Stu\_A\_890001 {X221237}  
 Stu\_A\_890001 {VCD000276, X253973, X215859, VCD000277, DVD000535}  
 Stu\_A\_890001 {X240492, X255340, X243228}

**Figure 6.**  
Integrated loan records

1. X253973 ⇒ VCD000277
2. X240492 ⇒ X255340
3. DVD000188 ⇒ DVD000041
4. X243178 ⇒ F189208
5. ....

**Figure 7.**  
Apriori Algorithm output



**Figure 8.**  
Recommendation system architecture

### 5. Conclusion

As noted, in the digital era users can get information easily and conveniently via information technology tools. The powerful development of information technology makes the function of personal services more important than before. As per users' needs, it is worth providing valuable and proper information actively.

---

This paper has proposed a personalized recommendatory system architecture to enable personalized services and management in the campus digital library. The architecture applied data mining technology to support recommendation services for users based on users' interests. We used the Ant Colony Clustering Algorithm and association rules to design a two-phase data mining process to generate recommendations. The process considered not only the relationships in the users' cluster but also the associations among the information accessed. With the advanced filter, the recommendation supported by the proposed system architecture would closely meet users' needs. This paper has not only constructed a recommending mechanism for readers in searching books from the web but has also taken the initiative in finding the most adaptive readers for books. Library managers are expected to purchase core and hot books from a limited budget to maintain and satisfy the requirements of readers along with promoting digital library services. Furthermore, more and more emphasis will be put on personal service by users in the future. The proposed system architecture could be also applied to other digital platforms to support a personal recommendation service.

### References

- Agrawal, R. and Srikant, R. (1994), "Fast algorithms for mining association rules", *Proceedings of the 20th VLDB Conference, Santiago*, pp. 487-99.
- Agrawal, R., Imielinski, T. and Swami, A. (1993), "Mining association rules between sets of items in large databases", *Proceedings of the ACM SIGMOD Conference on Management of Data, Washington, DC*, pp. 207-16.
- Berry, M. and Linoff, G. (1997), *Data Mining Techniques for Marketing, Sales, and Customer Support*, Wiley, New York, NY.
- Borgman, C.L. (1999), "What are digital libraries? Competing visions", *Information Processing & Management*, Vol. 35, pp. 227-43.
- Caro, G.D. and Dorigo, M. (1998), "Antnet: distributed stigmergetic control for communications networks", *Journal of Artificial Intelligence Research*, Vol. 9, pp. 317-65.
- Chang, C.C. and Chen, R.S. (2006), "Using data mining technology to solve classification problems-a case study of campus digital library", *The Electronic Library*, Vol. 24 No. 3, pp. 307-21.
- Chen, A.P. and Chen, C.C. (2006), "A new efficient approach for data clustering in electronic library using ant colony clustering algorithm", *The Electronic Library*, Vol. 24 No. 4, pp. 548-59.
- Dorigo, M. and Gambardella, L.M. (1997a), "Ant colonies for the traveling salesman problem", *BioSystems*, Vol. 43, pp. 73-81.
- Dorigo, M. and Gambardella, L.M. (1997b), "Ant colony system: a cooperative learning approach to the traveling salesman problem", *IEEE Transactions on Evolutionary Computation*, Vol. 1 No. 1, pp. 53-66.
- Dorigo, M., Maniezzo, V. and Colnari, A. (1996), "The ant system: optimization by a colony of cooperating agents", *IEEE Transactions on Systems, Man, and Cybernetics – Part B*, Vol. 26 No. 1, pp. 29-42.
- Edelstein, H. (2000), *Building Profitable Customer Relationships with Data Mining*, white paper, SPSS, Inc., Chicago, IL.

- 
- Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996), *From Data Mining to Knowledge Discovery in Databases*, American Association for Artificial Intelligence, Menlo Park, CA, pp. 37-54.
- Fong, A.C.M., Hui, S.C. and Jha, G. (2002), "Data mining for decision support", *IEEE IT Professional*, Vol. 4 No. 2, pp. 9-17.
- Heylighen, F. and Bollen, J. (2002), "Hebbian algorithms for a digital library recommendation system", *Proceedings of International Conference on Parallel Processing Workshops, Vancouver*, pp. 439-44.
- Han, J. and Chang, K.C.-C. (2002), "Data mining for web intelligence", *IEEE Computer*, Vol. 35 No. 11, pp. 64-70.
- Han, J. and Kamber, M. (2001), *Data Mining: Concepts and Techniques*, Morgan Kaufmann, San Mateo, CA.
- Hui, S.C. and Jha, G. (2000), "Data mining for customer service support", *Information and Management*, Vol. 38 No. 1, pp. 1-13.
- Leguizamón, G. and Michalewicz, Z. (1999), "A new version of ant system for subset problems", *Proceedings of the Congress on Evolutionary Computation, Washington, DC*, pp. 1459-64.
- Mitra, S., Pal, S.K. and Mitra, P. (2002), "Data mining in soft computing framework: a survey", *IEEE Transactions on Neural Networks*, Vol. 13 No. 1, pp. 3-14.
- Rocha, L.M. (1999), "TalkMine and the Adaptive Recommendation Project, Proceedings of the Association for Computing Machinery (ACM) – Digital Libraries", *University Of California, Berkeley, CA*, pp. 242-3.
- Rocha, L.M. and Bollen, J. (2001), "Biologically motivated distributed designs for adaptive knowledge management", in Segel, L. and Cohen, I. (Eds), *Design Principles for the Immune System and other Distributed Autonomous Systems*, Santa Fe Institute Series in the Sciences of Complexity, Oxford University Press, Oxford, pp. 305-34.

### Further reading

- Agrawal, R., Mannila, H., Srikant, R., Toivonen, H. and Verkamo, A.I. (1996), "Fast discovery of association rules", in Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P. and Uthurusamy, R. (Eds), *Advances in Knowledge Discovery and Data Mining*, MIT Press, Cambridge, MA, pp. 307-28.
- Anand, S.S. and Buchner, A.G. (1998), *Decision Support Using Data Mining*, Prentice-Hall, Englewood Cliffs, NJ.
- Ausiello, G., Crescenzi, P., Gambosi, G., Kann, V., Marchetti-Spaccamela, A. and Protasi, M. (1999), *Complexity and Approximation: Combinatorial Optimization Problems and their Approximability Properties*, Springer, Berlin.
- Fayyad, U. (1996), "Data mining and knowledge discovery: making sense out of data", *IEEE Expert*, Vol. 11, pp. 20-5.
- Fayyad, U. and Uthurusamy, R. (1996), "Data mining and knowledge discovery in databases", *Communications of the ACM*, Vol. 39 No. 11, pp. 24-6.
- Sheikholeslami, G., Chatterjee, S. and Zhang, A. (1998), "WaveCluster: a multi-resolution clustering approach for very large spatial databases", *Proceedings of the 24th VLDB Conference, New York, NY*, pp. 428-39.

---

EL  
25,6

**About the authors**

Chia-Chen Chen is Assistant Professor in the Department of Information Management, Tunghai University, Taiwan. Chia-Chen Chen is the corresponding author and can be contacted at: [kid.iim92g@nctu.edu.tw](mailto:kid.iim92g@nctu.edu.tw)

An-Pin Chen received a PhD degree in Industrial and Systems Engineering from the University of Southern California in the research areas of artificial intelligence, financial investment analysis and policy decision. He is now a Professor in the Institute of Information Management at National Chiao-Tung University in Taiwan.

**724**

---