Expert Systems with Applications 28 (2005) 359–369

Expert Systems with Applications

www.elsevier.com/locate/eswa

Mining changes in customer buying behavior for collaborative recommendations

Yeong Bin Cho[a,*], Yoon Ho Cho[b,1], Soung Hie Kim[c]

[a]Department of e-Business, Far East University, 5 san Wangjang, Gamgok, Eumsung, Chungbuk, 369-851, South Korea
[b]School of e-Business, Kookmin University, 861-1 Jungnung-dong, Sungbuk-gu, Seoul 136-702, South Korea
[c]Graduate School of Management, Korea Advanced Institute of Science and Technology, 207-43 Cheongryangri-Dong, Dongdaemun, Seoul, 130-012, South Korea

Abstract

The preferences of customers change over time. However, existing collaborative filtering (CF) systems are static, since they only incorporate information regarding whether a customer buys a product during a certain period and do not make use of the purchase sequences of customers. Therefore, the quality of the recommendations of the typical CF could be improved through the use of information on such sequences.

In this study, we propose a new methodology for enhancing the quality of CF recommendation that uses customer purchase sequences. The proposed methodology is applied to a large department store in Korea and compared to existing CF techniques. Various experiments using real-world data demonstrate that the proposed methodology provides higher quality recommendations than do typical CF techniques, with better performance, especially with regard to heavy users.

Keywords: Recommender systems; Purchase sequence; Collaborative filtering; SOM; Association rules

1. Introduction

Recommender systems have been a recent focus of researchers and practitioners. Many companies hope that the use of recommender systems may be a means of surviving in a competitive environment. Recommender systems are particularly suited to retail business, as compared to other types of business, since retail markets are distinguished by several characteristics, such as repeated buying over a particular time horizon, large numbers of customers, and a wealth of information detailing past customer purchases (Schmittlein & Peterson, 1994).

In general, retail companies operate purchase databases in a longitudinal way, such that all transactions are stored in chronological order. A record-of-transaction database typically contains the transaction date for and the products bought in the course of, a given transaction. Usually, each record also contains a customer ID, particularly when the purchase was made using a credit card or a frequent-buyer card. Therefore, the purchasing sequence of a customer in the database who has made repeat purchases can easily be determined. This purchase sequence provides a description of the changes in a customer's preferences over time. However, in our domain of knowledge, there has been little study of the question of whether recommendations based on purchase sequences may be more accurate than existing recommender system predictions, based on non-sequential patterns. In this study, for the purpose of enhancing the quality of recommendations, we propose a new methodology that considers the way in which a customer's purchase sequence evolves over time.

1.1. Motivation

To date, a variety of recommender systems (Balabanović & Shoham, 1997; Basu, Hirsh, & Cohen, 1998; Hill, Stead, Rosenstein, & Furnas, 1995; Lawrence, Almasi, Kotlyar, Viveros, & Duri, 2001; Resnick, Iacovou, Suchak, Bergstrom, & Riedl, 1994; Sarwar, Karypis, Konstan, & Riedl,

* Corresponding author. Tel.: +82 2 958 3864; fax: +82 2 958 3604.
E-mail addresses: ybcho@kgsm.kaist.ac.kr (Y.B. Cho), www4u@kookmin.ac.kr (Y.H. Cho), seekim@kgsm.kaist.ac.kr (S.H. Kim).
[1] Tel.: +82 2 910 4950; fax: +82 2 910 4519.

0957-4174/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2004.10.015

2001; and Shardanand & Maes, 1995) has been developed. Collaborative filtering (CF) has thus far been the most successful recommendation technique and has been used in a number of different applications, such as in the recommendation of web pages, movies, articles and products (Hill et al., 1995; Resnick et al., 1994; and Shardanand & Maes, 1995). Collaborative filtering works by recommending products to a target customer through a process of identifying people who share similar preferences for products and looking for those products that target customers are most likely to purchase.

The recommendation processes of typical collaborative filtering in retail business consist of the following three steps (Sarwar, Karypis, Konstan, & Riedl, 2000, 2001).

(1) *Customer profile construction*

The purchase transaction records of a customer for a certain period are used to build a customer profile describing his or her likes and dislikes. The system represents the customer profile, A, such that a_{ij} is one if a customer i has purchased a product j, and zero, otherwise.

(2) *Neighborhood formation*

This is the most important part of the CF-based recommender systems. The system finds a set of customers, known as neighbors, who, in the past, have exhibited similar behavior (i.e. bought a similar set of products), through calculating the correlations among customers for the customer profile. A set of K customers is usually found (a neighborhood of size K), which is formed according to the degree of similarity between each of the neighbors and a target customer.

(3) *Recommendation generation*

Once a neighborhood is formed for a target customer, the system generates a set of the top N products that the target customer is most likely to purchase, by searching for products that the neighbors have purchased and that the target customer has not yet purchased.

As mentioned above, typical collaborative filtering is static, since it only makes use of information relating to whether the customer bought a product during a certain period, and does not use information on the purchase sequences of customers in the determination of the neighbors of a target customer. However, customers in retail business are not static, and their buying behavior changes over time. Thus, the quality of the recommendation of the typical CF could be further improved through the use of the available information on the purchase sequences of customers.

To illustrate the importance of this potential improvement in accuracy, let us consider the following example. Table 1 presents typical transaction records for a retail company and the customer profile is provided in Table 2. This example determines products that target customer ID011 is likely to buy, using transaction records for consumers CID001 through CID011.

Assume that the typical CF algorithm is used for solving the problem, that the neighborhood size (K) is three, and that the number of products recommended (N) is two. The typical collaborative filtering algorithm considers the correlation of preferences between the target customer and the other customers. All of the four customers, CID011, CID001, CID002 and CID003, commonly bought 'Perfumes' and 'Skincare Products.' The similarities between CID011 and the other three customers are equivalent; that is, the Pearson correlation coefficient is 0.67. Therefore, a recommender system based on the collaborative filtering algorithm will determine that CID001, CID002 and CID003 are the nearest neighbors and have the same preferences as the target customer. However, it is quite difficult to select two products that should be recommended to CID011, because CID001, CID002 and CID003 each purchased different additional products: 'Dresses,' 'Shoes,' and 'Knits,', respectively. In this case, two different products to be recommended to CID011 would have to be selected randomly. Accordingly, the recommendations would not necessarily be very appropriate for the preferences of the target customer.

Table 1
Transactions for illustrative example

Customer ID	Transaction time	Products category bought
001	July 11 2000	Perfumes
001	August 17 2000	Skincares
001	September 14 2000	Dresses
002	July 15 2000	Perfumes
002	August 13 2000	Shoes
002	September 25 2000	Skincares
003	July 19 2000	Skincares
003	August 22 2000	Perfumes
003	September 18 2000	Knits
:	:	:
:	:	:
010	September 27 2000	Dresses
011	July 22 2000	Perfumes
011	August 26 2000	Skincares

Table 2
A customer profile for typical CF

CID	Perfumes	Skincares	Knits	Dresses	Shoes
001	1	1	0	1	0
002	1	1	0	0	1
003	1	1	1	0	0
004	1	0	0	1	0
005	1	0	0	0	1
006	0	1	1	0	0
007	0	1	0	1	0
008	0	1	0	0	1
009	0	0	1	0	0
010	0	0	0	1	0
011	1	1	0	0	0

Table 3
A customer profile rearranged by time

CID	Perfumes			Skincares			Knits			Dresses			Shoes		
	July	Aug.	Sep.	July	Aug.	Sep.	July	Aug.	Sep.	July	Aug.	Sep.	July	Aug.	Sep.
001	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0
002	1	0	0	0	0	1	0	0	0	0	0	0	0	1	0
003	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0
004	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0
005	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0
006	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
007	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
008	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
009	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
010	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
011	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0

1.2. A new approach

Table 3 provides a customer profile rearranged with regard to each customer's transaction time. From the table, it can be seen that the purchase sequence of the target customer CID011 occurred in the order of Perfumes, followed by Skincare Products. Similarly, the purchase sequences of CID001, CID002 and CID003 were Perfumes → Skincares → Dresses, Perfumes → Shoes → Skincares, and Skincares → Perfumes → Knits, respectively. Assume now that the neighbors of the target customer are determined based on the purchase sequence of each customer. Customers who have purchase sequences similar to that of CID011 include CID 001 and CID002. The purchase sequences of CID011 and CID001 are exactly same, as both bought the same products during the same month; therefore, the nearest neighbor of the target customer CID011 is CID001. The next nearest neighbor is CID002. Therefore, the products recommendable to CID011 are 'Dresses,' and 'Shoes.' As mentioned above, when the past purchase sequences of each customer are available, this knowledge can be used to enhance the quality of the recommendations made.

However, if a customer profile rearranged by time is applied directly, this application results in a significant problem, namely, the sparsity problem. It is well known that a sparse data set, having few nonzero entries, decreases recommendation accuracy (Mobasher, Cooly, & Srivastava, 2000). The sparsity level, defined as 1- (nonzero entries/total entries), of Table 2 is 60% (=1−(22/55)), while that of Table 3 is 87% (=1−(22/165)). In general, the sparsity level of a typical data set in the field of recommendation is over 95%. Rearrangement of an input matrix by time results in an increased time dimension, as compared to that of a typical customer profile, and thereby makes the input data set sparser. Therefore, a new solution to the sparsity problem must be found.

In our research, we employ a clustering technique that groups the transactions of customers into homogeneous subgroups. The SOM (Self-Organizing Map) technique, which has been applied frequently of late, is used for clustering (Kohonen, 1990). With the aid of the SOM, all the transactions of customers may be allocated to a certain cluster and a cluster number imposed. The change in the cluster number resulting from each transaction determines a customer purchase sequence.

By observing changes in the cluster number of each customer over time, a buying sequence can be built for each customer. These buying sequences are potentially capable of predicting the future purchases of a target customer. However, since not all buying sequences have a statistical validity sufficient to guarantee the generalization of the prediction, the association rule mining technique may be used to extract the sequential patterns from the buying sequences (Agrawal, Imielinski, & Swami, 1993).

2. Proposed methodology

2.1. Overall procedure

Generally, most marketing campaigns are conducted based on transactions occurring during a specified time period (e.g. 3 months or 6 weeks). We assume that a time period of length l is used to detect the purchase sequence of a customer and that a product recommendation for a target customer is made at time T. In other words, our problem can be described as follows: When the purchase sequence and buying history of a target customer for the past $l-1$ periods prior to time T are given, which product is the target customer most likely to purchase at time T?

For solving the above problem, our recommendation procedure is divided into two components, called a 'model-building phase' and a 'recommendation phase.' Fig. 1 presents the overall procedure. A model-building phase is performed once to create a reliable model from the customer transaction database, while a recommendation phase is used to recommend products that target customers are highly likely to buy.

The model-building phase is divided into the following three steps. First, transaction clustering is conducted, so that all the transactions of customers are clustered. The SOM

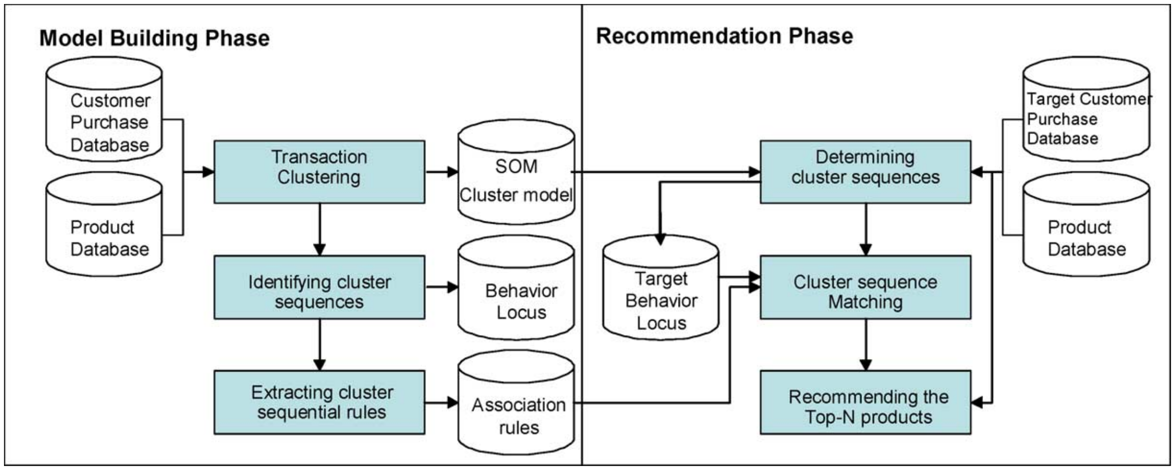


Fig. 1. Overall procedure.

model used for transaction clustering is stored in a SOM cluster model base. This SOM model is used to cluster target customers' transactions in the recommendation phase. The second step is to detect the evolving customer purchase sequences as time passes. These customer behaviors, which are derived from a change in the cluster number of each customer, are kept in the purchase sequence database. In the final step of the model-building phase, sequential purchase patterns over user-specified minimum support and confidence are extracted using the association rule. The sequential purchase patterns are then stored in the association rule database.

The recommendation phase begins with the transaction clustering of target customers using the SOM model built in the model-building phase. In a manner similar to that of the first two steps of the modeling phase, all the transactions of a target customer are converted to a purchase sequence. The second step of the recommendation phase consists of a matching, such that a target customer's purchase sequence is compared with the purchase sequence stored in the association rule base. Finally, after the most similar purchase sequences have been identified, our approach generates a set of products that the target customer is most likely to purchase by selecting the top N most commonly purchased products in the cluster.

2.2. Model-building phase

2.2.1. Transaction clustering

In this section, we present a transaction clustering for the purpose of constructing a recommendation model. We use a SOM technique to obtain transaction clusters, as mentioned above. However, SOM clustering technique often breaks down when handling very high-dimensional data. The numbers of dimension are the products and can number in the ten thousand in retail business. Our approach suggests that using product taxonomy can provide an effective dimensionality reduction method while improving clustering results. Product taxonomy represents the hierarchical relationships among products as the domain-specific knowledge of marketing managers or domain experts (Cho, Kim, & Kim, 2002; Cho & Kim, 2004; Lawrence et al., 2001). Fig. 2 presents an example of product taxonomy for those goods in a large department store that women are most likely to purchase.

We shall assume that a product class set P is classified into n different subclasses, and that each subclass consists of subclasses at a lower level, or eventual leaf products, as follows

$$P = \{P_1, P_2, \ldots, P_n\}. \tag{1}$$

Suppose that A is the set of the transactions of m customers during l periods before time T. More specifically, let A be composed as follows

$$A = \{A_{1,T-k}, A_{2,T-k}, \ldots, A_{m,T-k}\}, \quad k = 0, 1, \ldots, l-1, l \geq 2, \tag{2}$$

where $A_{j,T-k} \in A$ is a non-empty subset of P.

Each $A_{j,T-k}$ represents the product class or classes from which customer j purchased products at time $T-k$.

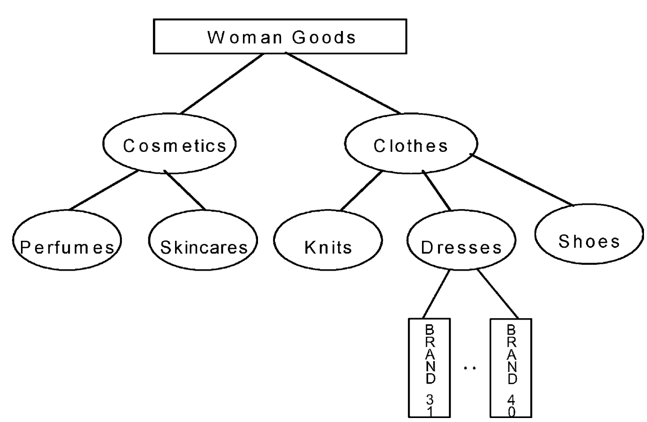


Fig. 2. An example of product taxonomy.

Each $A_{j,T-k}$ is transformed into an input matrix composed of a bit vector, and the matrix to be transformed is used in the transaction clustering. The time-ordered vectors for a particular customer represent the purchasing history of the customer; this input matrix can be thought of as the dynamic profile of the customer. We define a dynamic customer profile as follows.

Definition 1. Dynamic customer profile

Let $\bar{A}$ be a dynamic customer profile. Then, $\bar{A}$ is defined by the following matrix for n product classes and m customers over the course of l periods

$$\bar{A}_{j,T-k} = \langle P_1^{A_{j,T-k}}, P_2^{A_{j,T-k}}, \ldots P_n^{A_{j,T-k}} \rangle, \quad j = 1, 2, \ldots, m,$$
$$k = 0, 1, \ldots, l-1, l \geq 2, \tag{3}$$

where

$$P_i^{A_{j,T-k}} = \begin{cases} 1, & \text{if } P_i \in A_{j,T-k} \\ 0, & \text{otherwise} \end{cases}$$

Example 1. Throughout this paper, we use the example given in Section 1 to illustrate proposed terminologies. The set of product classes given in Table 3 is P = {Perfumes, Skincare Products, Knits, Dresses, Shoes}, and the transactions of customer CID001 are $A_{001,\text{July}}$ = {Perfumes}, $A_{001,\text{Aug.}}$ = {Skincares}, $A_{001,\text{Sep.}}$ = {Dresses}. The dynamic customer profile of customer CID001 from July to September may thus be represented as $\bar{A}_{001,\text{July}} = \{1,0,0,0,0\}$, $\bar{A}_{001,\text{Aug}} = \{0,1,0,0,0\}$, $\bar{A}_{001,\text{Sep}} = \{0,0,0,1,0\}$.

All the transactions of customers in the training customer purchase database are transformed into dynamic customer profiles based on their prior purchase behaviors. We then use the SOM clustering technique to assign each transaction to a group. This transaction clustering facilitates the discovery of the dynamic cluster sequence of a customer in a way that produces a change in the cluster of a customer over time. The SOM model is then stored in the SOM cluster model base and is used for predicting the target customer's dynamic behavior.

2.2.2. The identification of cluster sequences

The transaction clustering results in the following set of q clusters

$$\mathrm{C} = \{C_1, C_2, \ldots, C_q\}, \tag{4}$$

where each C_i is a subset of the $\bar{A}$ given in (3).

Each cluster represents only a group of transactions with similar patterns. A rearrangement of these clusters by customer and by time period is necessary for the identification of the dynamic behavior of each customer. It is possible to learn the cluster sequence of a customer by identifying the cluster to which each transaction of the customer belongs, during each time period. To formalize this concept, we use the following terminology:

Definition 2. Customer Behavior Locus

Let L_i be the behavior locus of customer i. Then, the behavior locus L_i is identical to the changes in the cluster number of customer i during l periods and is defined as follows

$$L_i = \langle C_{i,T-l+1}, \ldots, C_{i,T-1}, C_{i,T} \rangle, \quad i = 1, 2, \ldots, m, \tag{5}$$

where $C_{i,T-k} \in \mathrm{C}$, $k = 0,1,2,\ldots,l-1, l \geq 2$.

The process of searching for a behavior locus can be simply conducted through transaction clustering. The following example presents the sample behavior locus of a customer.

Example 2. The behavior locus of CID001, L_{001}, is ⟨10,3,9⟩, as shown in Table 4. L_{001} indicates that customer CID001 belonged to the tenth cluster in July and moved into the third cluster in August, thereafter reaching the ninth cluster in September (Fig. 3).

2.2.3. The extraction of sequential cluster rules

All customers have a behavior locus based on their prior transactions, as was illustrated in the previous section. However, a behavior locus with little statistical validity is not sufficient for use as the rule predicting the behavior of a target customer, because it does not constitute the general behavior pattern, only a dynamic behavior locus, of the customer. The association rule technique is well suited for determining the most frequent pattern with confidence, since it provides automatic filtering capabilities.

We intend to discover the cluster of a target customer at time T based on his/her past behavior. For doing this, we divide the input data into a conditional part and a consequential part. The conditional part of the association rule is composed of the left-hand-side $\langle C_{i,T-l+1},\ldots,C_{i,T-1} \rangle$ of formula (5), and $C_{i,T}$ is assigned to the consequential part. This approach, termed the goal-oriented association rule by Chiang, Wang, Lee, and Lin (2003), has been known to

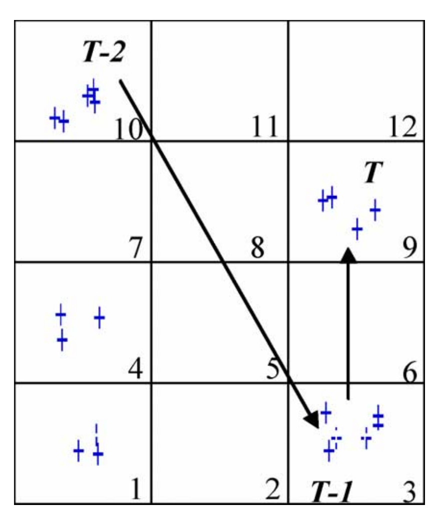


Fig. 3. A behavior locus of customer CID001.

increase rule-finding efficiency. For the typical association rules, we subsequently find those having support and confidence higher than the predetermined minimum value.

We represent the association rule R_j over the user-specified minimum support and confidence in the following form

$$R_j : r_{j,T-l+1}, \ldots, r_{j,T-1} \Rightarrow r_{j,T} \ (\text{Support}_j, \text{Confidence}_j), \quad (6)$$

where $r_{j,T-k} \in \mathrm{C}$ or ϕ, and $r_{i,T} \in \mathrm{C}$.

A rule R_j indicates that, if the locus of a customer is $r_{j,T-l+1},\ldots,r_{j,T-1}$, then the behavior cluster for that customer is $r_{j,T}$ at time T.

2.3. Recommendation phase

In this phase, given target customers, we seek those products that are best matched to the dynamic behaviors of these customers. The target customers' transactions are converted into behavior locus using the SOM model, as in the previous phase. Furthermore, the best-matching loci stored in the association rule base are extracted, and the search methodology for determining the products that the target customers are most likely to buy is isolated.

2.3.1. The determination and matching of the cluster sequences of target customers

Behavior locus prediction begins when a target customer's transactions are entered into the SOM model. It is necessary to know the degree to which the behavior locus of a target customer during $l-1$ periods before T is similar to the rules of the association rule base. The cluster locus, transformed via the SOM model, of a target customer is compared with the association rules derived from other customers' loci, and then the best-matching locus is determined. Execution of this process requires new measures for calculating the degree of correspondence between the association rules in the model base and the behavior locus of a target customer.

The degree of similarity between the two, or the extent to which the behavior locus of a target customer is identical to the conditional part of the association rule in the model base, in the same period, can be used as the correspondence measure. We define this similarity measure as follows.

Definition 3. Similarity measure

Let L_i^C be the behavior locus of target customer i during $l-1$ periods and R_j^C be the conditional part of association rule j stored in the model base. Then, $L_i^C = \langle C_{i,T-l+1}, \ldots, C_{i,T-1} \rangle$, and $R_j^C = \langle r_{j,T-l+1}, \ldots, r_{j,T-1} \rangle$. Let SM_i^j denote the similarity measure between L_i^C and R_j^C. SM_i^j is defined as follows

$$\mathrm{SM}_i^j = \sum_{k=1}^{l-1} S_{i,T-k}^j, \quad (7)$$

where

$$S_{i,T-k}^j = \begin{cases} 1 & \text{if } C_{i,T-k} = r_{j,T-k} \\ 0 & \text{otherwise} \end{cases}$$

The above definition indicates that, if the behavior locus of a target customer i is equal to the conditional part of association rule j in the same period, then $S_{i,T-k}^j$ is equal to one, but is otherwise equal to zero. Note that a higher SM_i^j value suggests a greater correspondence between the behavior locus of a target customer and the conditional part of association rule j.

However, even if the similarity measure is high, a choice of the association rule suited to the prediction of the cluster of a target customer at time T is nonetheless difficult, since such a rule is not general, given that the support and confidence of the association rule may be remarkably low. Thus, to assure a good fit between the behavior locus of a target customer and the conditional part of the association rule, we need an alternative measure of fitness. We introduce such a measure as follows:

Definition 4. Fitness measure

Let FM_i^j be a measure of the goodness-of-fit between the behavior locus i and the association rule j of the model base. Then, FM_i^j is defined as follows

$$\mathrm{FM}_i^j = \mathrm{SM}_i^j \times \text{Support}_j \times \text{Confidence}_j \quad (8)$$

Using the above definition, we can determine that the cluster of a target customer at time T is a consequential part $r_{j,T}$ of the association rule j with maximum FM_i^j.

Example 3. Table 4 presents the loci of customers, and Table 5 presents the association rules derived from customers' loci, up to a minimum support of 0.1 and a minimum confidence of 0.5. The similarities between the locus of customer CID011 and the derived rules are $\mathrm{SM}_{011}^3 = 2$ and $\mathrm{SM}_{011}^1 = 1$. Thus, customer CID011 belongs to the ninth cluster at time T, since the fitness between CID011 and the rules are $\mathrm{FM}_{011}^3 = 0.2$ and $\mathrm{FM}_{011}^1 = 0.2001$.

2.3.2. Recommendation of the top N items

The final step involves the derivation of the top N recommendations from the predicted cluster for a target

Table 4
Behavior loci of customers

CID	$T-2$	$T-1$	T
001	10	3	9
002	10	1	3
003	3	10	4
004	10	–	9
005	1	–	10
006	4	–	3
007	–	9	3
008	3	–	1
009	–	4	–
010	–	–	9
011	10	3	?

Table 5
Derived association rules

Rules	$T-2$	$T-1$	T	Support	Confidence
1	10	–	9	0.3	0.667
2	3	10	4	0.1	1.0
3	10	3	9	0.1	1.0
4	10	1	3	0.1	1.0
5	1	–	10	0.1	1.0
6	–	10	4	0.1	1.0
7	3	–	4	0.2	0.5
8	3	–	1	0.2	0.5
9	–	3	9	0.1	1.0
10	–	1	3	0.1	1.0
11	–	9	3	0.1	1.0
12	4	–	3	0.1	1.0

customer at time T. For each target customer, we produce a recommendation list of N products that the target customer should be willing to purchase. Recommendations for a specific target customer are derived from the purchase database of target customers and are drawn from the list of popular products in the cluster assigned to that customer at time T.

Let us denote by C^* the predicted cluster of a target customer at time T, as determined in the previous section. C^* may include the transactions in which products were previously purchased by the target customer, as well as any transactions in the product class level, rather than in the individual product level (e.g. leaf node in the product taxonomy). Thus, we need to select the transactions that are most suitable for recommending the top N products to the target customer. We only select transactions at time T in the creation of a recommendation list, since products that the target customer is likely to buy consist of what other customers in C^* bought at time T. It is also necessary to carry out a reverse decomposition that reduces a higher product class in the product taxonomy into the individual products in the leaf nodes. Finally, previously purchased products are excluded from the recommendation list in order that each customer's purchase patterns, or the coverage, are broadened.

We may now determine the top-N product recommendation list for a target customer as the most frequently purchased products from among the products in the cluster.

Definition 5. Recommendation list for a target customer

Let $\mathrm{MF}(r_1)$ denote the most frequently purchased product at time T in C^*. Similarly, $\mathrm{MF}(r_2)$ is ranked the next highest, and $\mathrm{MF}(r_N)$ is ranked the Nth highest. Then, the recommendation list for the target customer is given by $\mathrm{MF}(r_1)$, $\mathrm{MF}(r_2),\ldots,\mathrm{MF}(r_N)$, while $\mathrm{MF}(k)$ is computed as follows

$$\mathrm{MF}(k) = \sum_{j\in C*} P_{ik}^{A_{j,T}} \times N_{ik}^{T}, \qquad (9)$$

where

Table 6
A product list purchased by other target customers in selected cluster

	Purchased products (Brand)
CID 012	Brand 31(2), 33(3)
CID 013	Brand 31(2), 37(2)
CID 015	Brand 33(2), 38(3)

Number in parenthesis denotes purchase quantity.

$$P_{ik}^{A_{j,T-k}} = \begin{cases} 1, & \text{if } P_{ik} \in A_{j,T} \\ 0, & \text{otherwise} \end{cases}$$

N_{ik}^{T} is the quantity of product P_{ik} sold in period T, and P_{ik} is the kth leaf product in product class i.

Example 4. Suppose that Table 6 presents the products (brands) purchased by target customers who are assigned to the ninth cluster at time T (the number in parentheses indicates the number of products bought by each customer). Furthermore, assume that the top three products are recommended for CID011. Then, $\mathrm{MF}(r_1)=$Brand 33, $\mathrm{MF}(r_2)=$Brand 31, $\mathrm{MF}(r_3)=$Brand 38.

3. Applications and evaluations

3.1. Data sets

We used real-world data to examine the performance of the proposed approach. The data used in the experiment were transaction records for those goods sold by the H department store, the third largest department store in Korea, that were commonly purchased by women. In addition, we used transaction records obtained during the eight-month period from May to December 2000, in order to establish the behavioral characteristics of the customers over time. The input data from the H department store database consisted of 18,843 transactions and 557 products, and contained customer purchase data for 1833 customers. Customers selected as suitable to receive recommendations were restricted to loyal customers who had purchased frequently and recently, as it is difficult to identify the dynamic purchase behavior of customers who purchase goods only rarely. A month was chosen as the time unit for analysis, because customers only rarely purchased products from the same department repeatedly on a daily or weekly basis. Interviews with domain experts indicated that loyal customers could be identified as those customers who had made a purchase at least once each month during four consecutive months. Three hundred and ten customers fell into this category, and the number of products they purchased was equal to the total number of products mentioned above.

3.2. Evaluation method

The product taxonomy used by the H department store consisted of three levels: the top level contained ten product classes, the next level contained 25 product classes, and bottom level contained all 557 products.

The period between May and August of 2000 was set as the training period for model building, and the period from September to December was set as the test period for recommendations. The SOM model was applied to the training data, and a fourfold cross-validation was conducted in the course of building the model. Furthermore, for the cases of 17 of the differently numbered clusters of the SOM model, we experimented in order to learn the extent to which the number of clusters impacted the accuracy of the recommendations. In a real application, it is desirable to discover the number of clusters in the SOM model that will yield the highest level of accuracy and then recommend products that the target customer is likely to buy using this model. Experiments for various numbers of clusters were conducted for the reason that the determination of the optimal number of clusters is crucial to model-building, as recommendations and their accuracy vary according to the number of clusters used.

Moreover, we chose a minimum support level of 2% and a minimum confidence level of 50% for the selection of sequential rules, both of which were higher than those used in the study by Lawrence et al. (2001). We fixed the number of recommendations at ten. We selected 132 customers who had purchased at least once each month, consecutively, during the test period, as the set of target customers.

3.3. Evaluation measures

To evaluate the quality of the recommendation set, measures of recall and precision have been widely used in the field of recommender systems (Basu et al., 1998; Billsus & Pazzani, 1998; Lin et al., 2000; Lin, Alvarez, & Ruiz, 2002; Sarwar et al., 2000). Recall is defined as the ratio of the number of products that are recommended to the number of actual products purchased by the target customer at time T. In contrast, precision is defined as the ratio of the number of recommended products purchased to the number of products on the recommendation list. Recall measures how many of the products in the actual customer purchase list consist of recommended products, whereas precision measures how many of the recommended products belong to the actual customer purchase list. These measures are simple to compute and intuitively appealing, but they are in conflict, since increasing the size of the recommendation set leads to an increase in recall but, at the same time, to a decrease in precision (Sarwar et al., 2000). Hence, a widely used combination metric called the 'F1 metric' (Billsus & Pazzani, 1998; Rijsbergen, 1979; Sarwar et. al., 2000, 2001), which gives equal weight to both recall and precision, was also employed in the course of our evaluation. These are computed as follow

$$\text{Precision} = \frac{\text{Number of hit products}}{\text{Total number of recommended products}} \tag{10}$$

$$\text{Recall} = \frac{\text{Number of hit products}}{\text{Total number of products purchased at period } T} \tag{11}$$

$$F1 - \text{measure} = \frac{\text{Recall} \times \text{Precision}}{(\text{Recall} + \text{Precision})/2} \tag{12}$$

3.4. Results and discussion

3.4.1. Impact of the number of SOM clusters

We performed evaluations in which the number of clusters varied, to determine the effectiveness of

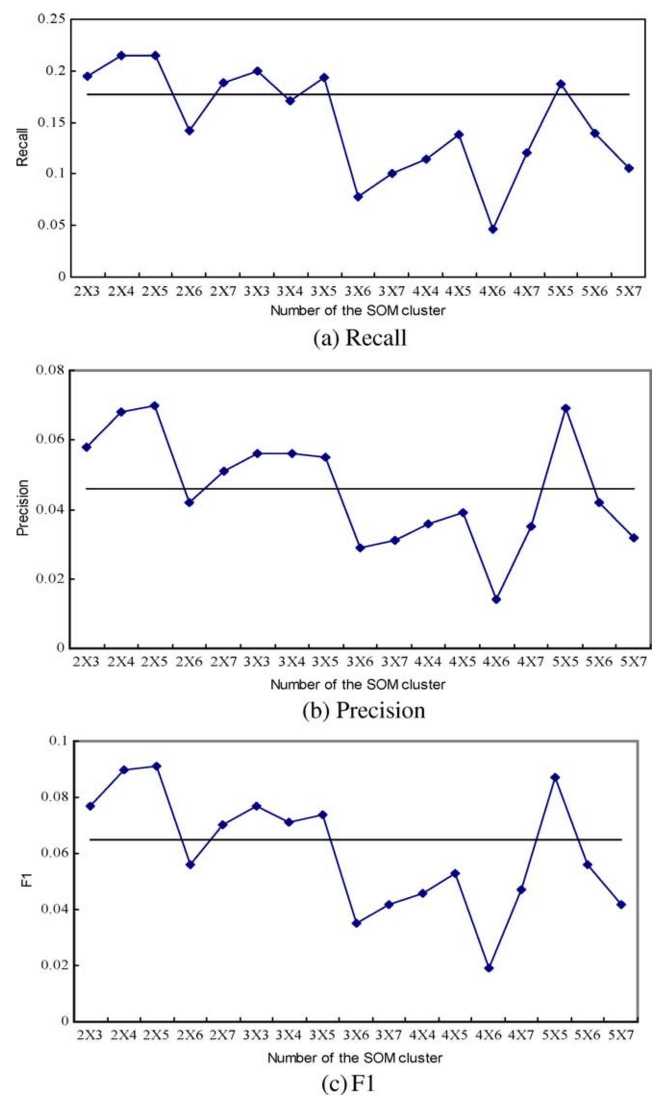


Fig. 4. Comparison of the proposed methodology and CF by the number of clusters, where the straight line indicates the accuracy of benchmark CF with the best accuracy.

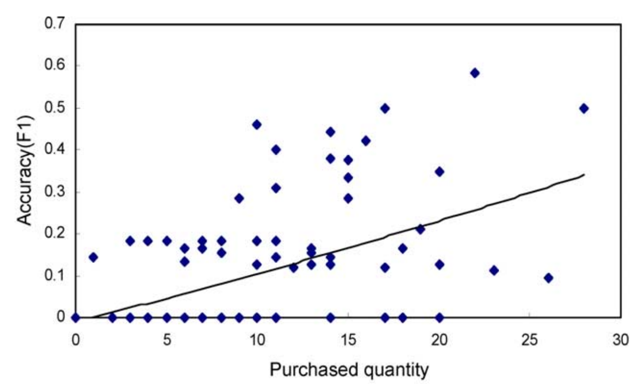
Table 7
Two tail paired t-test by the number of clusters

	F1			Recall			Precision		
	Mean	Standard deviation	SIG	Mean	Standard deviation	SIG	Mean	Standard deviation	SIG
SOM2X3	-1.14×10^{-2}	0.12518	+	-1.78×10^{-2}	0.4	+	-1.21×10^{-2}	9.62×10^{-2}	+
SOM2X4	-2.47×10^{-2}	0.13645	+(*)	-3.79×10^{-2}	0.41	+	-2.20×10^{-2}	0.105	+(*)
SOM2X5	-2.59×10^{-2}	0.12895	+(*)	-3.87×10^{-2}	0.41	+	-2.35×10^{-2}	9.99×10^{-2}	+(**)
SOM2X6	9.57×10^{-3}	0.13		3.51×10^{-2}	0.39		4.54×10^{-3}	9.76×10^{-2}	
SOM2X7	-4.23×10^{-3}	0.13	+	-1.14×10^{-2}	0.38	+	-4.55×10^{-3}	9.91×10^{-2}	+
SOM3X3	-1.22×10^{-2}	0.13	+	-2.31×10^{-2}	0.39	+	-9.85×10^{-3}	9.28×10^{-2}	+
SOM3X4	-6.03×10^{-3}	0.12996	+	6.11×10^{-3}	0.39		-9.85×10^{-3}	0.103	+
SOM3X5	-8.32×10^{-3}	0.13258	+	-1.64×10^{-2}	0.4	+	-9.09×10^{-3}	9.88×10^{-2}	+
SOM3X6	3.01×10^{-2}	0.11		9.88×10^{-2}	0.34		1.74×10^{-2}	8.48×10^{-2}	
SOM3X7	2.37×10^{-2}	0.13		7.68×10^{-2}	0.39		1.52×10^{-2}	9.49×10^{-2}	
SOM4X4	1.88×10^{-2}	0.13		6.25×10^{-2}	0.36		1.06×10^{-2}	0.1	
SOM4X5	1.25×10^{-2}	0.12485		3.86×10^{-2}	0.4		7.57×10^{-3}	8.84×10^{-2}	
SOM4X6	4.64×10^{-2}	0.13		0.13	0.34		3.18×10^{-2}	9.72×10^{-2}	
SOM4X7	1.79×10^{-2}	0.14		5.55×10^{-2}	0.43		1.14×10^{-2}	0.11	
SOM5X5	-2.21×10^{-2}	0.13	+(*)	-1.05×10^{-2}	0.35	+	-2.27×10^{-2}	0.10	+(*)
SOM5X6	9.48×10^{-3}	0.12793		3.71×10^{-2}	0.39		4.54×10^{-3}	9.36×10^{-2}	
SOM5X7	2.28×10^{-2}	0.12256		7.07×10^{-2}	0.36		1.44×10^{-2}	9.14×10^{-2}	

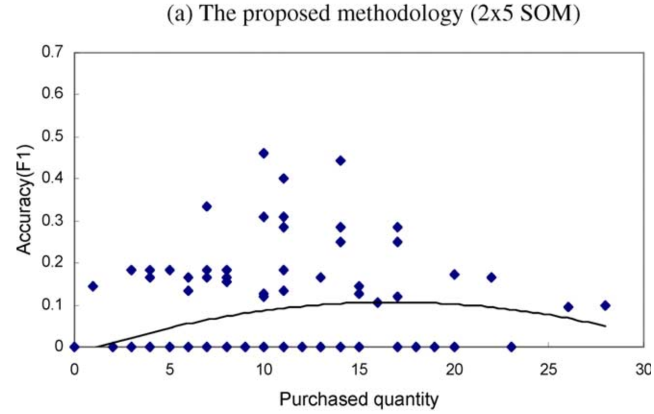
*, $P<0.05$, **, $P<0.01$.

the recommendations; we did this by computing the three metrics described above. In addition, we compared the recommendation accuracy of our proposed methodology with that of the most accurate CF benchmark algorithm. Fig. 4 presents our experimental results. All three measures were higher in the cases for the following numbers of clusters than for the benchmark CF algorithm: 6(2×3), 8(2×4), 9(3×3), 10(2×5), 14(2×7), 15(3×5), and 25(5×5). In contrast, they were lower in the cases with the following numbers of clusters: 12(2×6, 3×4), 16(4×4), 18(3×6), 20(4×5), 21(3×7), 24(4×6), 28(4×7), and 30(5×6). In the case of 12(3×4) clusters, recall was lower than under the CF, while the other two measures were higher in this case.

We were unable to identify a general tendency with regard to the number of clusters such as, for example, an increase in accuracy with an increase in the number of clusters, or vice-versa. Having examined the results, we inferred that the number of clusters used did affect the accuracy of the top N recommendations. Intuitively, this result is fairly reasonable, because all existing clustering methods have suffered from the problem of variation in results as a function of the number of clusters, and the determination of their optimal number continues to be a challenging problem (Nour & Madey, 1996). These results depended on the extent to which each SOM model took the behavior locus of the target customer into account.



(a) The proposed methodology (2x5 SOM)



(b) The benchmark CF technique

Fig. 5. Scatter diagram on recommendation accuracy.

Table 8
The Chow test between the proposed methodology and the benchmark CF

	RSS	d.o.f	F
Total	3.5387	261	–
The proposed method	1.9799	129	–
CF	1.3737	129	–
The Chow test			4.747**

**, $P<0.01$, $F(3,262.001)=3.85$.

To prevent over-fitting as a result of clustering, it was necessary to determine the number of SOM clusters that would exactly explain the locus of the customer.

Table 7 presents the results of two-tailed *t*-tests for each of the SOM clusters, compared with the most accurate CF benchmark. The values for F1 and precision, in the cases with 8(2×4), 10(2×5), and 25(5×5) clusters, were statistically significant. The proposed methodology exhibited better performance than the traditional CF technique, when the numbers of the SOM clusters were chosen well. The proposed methodology employing the optimal number of clusters, in this case 10(2×5), performed still better, achieving an average improvement of 40 and 52%, in F1 and precision, respectively.

3.4.2. Impact on the quantity purchased

Fig. 5 presents a scatter diagram of the quantity purchased (*X*-axis) and F1 (*Y*-axis) for the target customers, where (a) indicates case 10(2×5) for the proposed methodology, and where (b) is the benchmark CF. At a glance, the proposed methodology displayed better accuracy than the benchmark CF with regard to increases in the quantity purchased. In order to resolve the difference in performance between the two, we added a trend line to best fit the scatter pattern in each figure. After several trials, we arrived at the quadratic regression line shown in Fig. 5. This regression line indicates that the proposed methodology corresponds to a straight line that is proportional to the quantity purchased, while that of the CF decreases inversely in the same. To test whether a structural change occurred from one model to the other, we applied the Chow test popularly used in this case (Chow, 1960). The test indicated, with a statistical significance level of $p<0.01$, that the proposed methodology is structurally different from the CF, as is shown in Table 8. Thus, we conclude that the proposed methodology constitutes an improvement, on average, especially with regard to the behavior of heavy users, when compared to the existing CF techniques.

4. Conclusion

The preferences of customers change over time. In this study, we described a model-based approach for mining the changes in customer buying behavior over time and discussed solutions to several problems: data preprocessing, behavior locus extraction, and recommendation formulation based on extracted loci. Using the derived recommendation list, companies may be able to perform effective one-to-one marketing campaigns by providing individual target customers with personalized product recommendations.

The research presented in this paper makes a contribution to the related recommender systems literature. We took into consideration changes in customer preferences to improve the accuracy of the recommendations made. In particular, we determined that the proposed methodology is more suitable for heavy users.

Some possible extensions to this work are as follows. From the results of this study, we know which products target customers are likely to buy, but we have not yet explored the times at which these purchases are likely to occur. Further research analyzing customers' past purchasing patterns should likewise enable prediction of the most appropriate times for recommendations to be given. In addition, since the accuracy of all model-based approaches deteriorates as time passes, the model must be dynamically updated to reflect the users' evolving interests. The way in which the predictive capabilities of the model decrease as time passes should be investigated, with the goal of creating a repair plan. Furthermore, one interesting research extension would be the setting up of a real marketing campaign, in which customers would be targeted using our methodology, which could then be evaluated with regard to its performance.

Acknowledgement

This work was supported by research program 2005 of Kookmin University in Korea.

References

Agrawal, R., Imielinski, T., & Swami, A. (1993). *Mining association rules between sets of items in large databases Proceedings of the ACM SIGMOD conference on management of data* pp. 207–216.

Balabanović, M., & Shoham, Y. (1997). Content-based, collaborative recommendation. *Communacations of the ACM*, *40*(3), 66–72.

Basu, C., Hirsh, H., & Cohen, W. (1998). Recommendation as classification: using social and content-based information in recommendation. In: *Proceedings of the 1998 workshop on recommender systems* (pp. 11–15), AAAI Press.

Billsus, D., & Pazzani, M.J. (1998). Learning collaborative information filters. *Proceedings of the 15th international conference on machine learning* (pp. 46–54).

Cho, Y. H., & Kim, J. K. (2004). Application of web usage mining and product taxonomy to collaborative recommendations in e-commerce. *Expert Systems With Applications*, *26*(3), 234–246.

Cho, Y. H., Kim, J. K., & Kim, S. H. (2002). A personalized recommender system based on web usage mining and decision tree induction. *Expert Systems With Applications*, *23*(3), 329–342.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two regressions. *Econometrica*, *28*(3), 591–605.

Hill, W., Stead, L., Rosenstein, M., & Furnas, G. (1995). Recommending and evaluating choices in a communication of use. In: *Proceedings of CHI 95*.

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, *78*(9), 1464–1480.

Lawrence, R. D., Almasi, G. S., Kotlyar, V., Viveros, M. S., & Duri, S. S. (2001). Personalization of supermarket product recommendation. *Data mining and Knowledge Discovery*, *5*(1–2), 11–32.

Lin, W., Alvarez, S. A., & Ruiz, C. (2002). Efficient adaptive-support association rule mining for recommender systems. *Data Mining and Knowledge Discovery*, *6*(1), 83–105.

Mobasher, B., Cooly, R., & Srivastava, J. (2000). Automtic personalization based on web mining. *Communications of ACM*, *43*(8), 142–151.

Nour, M. A., & Madey, G. R. (1996). Heuristic and optimization approaches to extending the Kohonen self organizing algorithm. *European Journal of Operational Research*, *93*(2), 428–448.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). Grouplens: an open architecture for collaborative filtering of netnews. In: *Proceedings of the ACM 1994 conference on computer supported cooperative work* (pp. 175–186).

Rijsbergen, C. J. (1979). *Information retrieval* (2nd ed.). London: Butterworths.

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2000). Analysis of recommendation algorithms for E-commerce. In: *Proceedings of ACM E-commerce 2000 conference* (pp. 15–167).

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithm. In: *Proceedings of the tenth international world wide web conference* (pp. 285–295).

Schmittlein, D. C., & Peterson, R. A. (1994). Customer base analysis: an industrial purchase process application. *Marketing Science*, *13*(1), 41–67.

Shardanand, U., & Maes, P. (1995). Social Information Filtering: Algorithms for Automating 'Word of Mouth'. In: *Proceedings of CHI 95*.

Chiang, D., Wang, Y., Lee, S., & Lin, C. (2003). Goal-oriented sequential pattern for network banking churn analysis. *Expert Systems with Applications*, *25*(3), 293–302.