Article

Detecting the change of customer behavior based on decision tree analysis

Jae Kyeong Kim,[1] Hee Seok Song,[2] Tae Seong Kim[1] and Hyea Kyeong Kim[1]

(1) School of Business Administration, KyungHee University, 1 Hoeki-Dong, Dongdaemoon-Gu, Seoul 130-701, Korea
E-mail: jaek@khu.ac.kr
(2) Department of Management Information Systems, Hannam University, 133 Ojung-Dong, Daeduk-Gu, Daejeon 306-791, Korea
E-mail: hssong@hannam.ac.kr

Abstract: *Understanding and adapting to changes in customer behavior is an important aspect for survival in a continuously changing environment. This paper develops a methodology based on decision tree analysis to detect the change in classified customer segments automatically between two data sets collected over time. We first define three types of changes as the emerging pattern, the unexpected change and the added/perished rule. Then, similarity and difference measures are developed for rule matching to detect all types of change. Finally, the degree of change is developed to evaluate the amount of change. Our suggested methodology based on decision tree analysis in the change detection problem can be used in more structured situations in which the manager has a specific research question and it also detects the change of classification criteria in a dynamically changing environment. A Korean Internet shopping mall case is evaluated to represent the performance of our suggested methodology, and practical business implications for this methodology are also provided. We believe that the change detection problem and the suggested methodology will become increasingly important as more data mining applications are implemented.*

Keywords: data mining, decision tree, change analysis, Internet shopping mall

1. Introduction

Understanding and adapting to changes in customer needs is an important aspect for survival in a continuously changing environment. In this paper, we propose a methodology for mining the change in customer behavior before and after a certain point in the contexts of decision tree classification. More specifically, this paper aims to give answers about the following questions. What are the differences in customer segmentation before and after a new service adoption? Which customer group's purchasing amount is gradually increasing? Which customer groups are churned over the years? These are typical questions which relate to the change in customer classification. Recent results of interviews performed by a consulting company show the importance of knowing the change in customer needs.

For this purpose, we discover the change in customer behavior by comparing two classification rule sets which are generated from different periods. Our suggested methodology is based on data mining techniques, especially decision tree analysis. Data mining is the process of exploration and analysis of large quantities of data in order to discover meaningful patterns and rules. In this context, data mining shows great potential for tracking changes in customer behavior automatically. Therefore, we use data mining techniques to detect changes in customer behavior. But most data mining techniques such as association rule mining, decision trees and neural networks cannot be applied alone to answer the above research questions, because they cannot handle dynamic situations well.

Liu *et al.* (2000) suggest a change mining methodology based on decision trees to find interesting relationships among a large set of data items. However, their research cannot give an answer about how much change has occurred. It is important to detect what kinds of changes have occurred, but it is also important to differentiate which changes are more serious. Song *et al.* (2001) developed a change detection procedure and a measure for evaluating the amount of change based on association rule mining. The measure for evaluating the amount of change due to Song *et al.* (2001) is adapted with modification in this research. Change detection based on association rule mining can be useful to identify changes of customer behavior in unstructured and ill-defined situations because of the unsupervised learning feature of association rule mining. However, decision tree analysis in change detection problems can be used in more structured situations in which the manager has a specific research question and it also detects the change in classification criteria in a dynamically changing environment.

Changes detected by decision tree analysis are usefully applied to plan various niche-marketing campaigns. For example, in a shop, if a manager can find out that the criteria of a certain customers' group for choosing a product has changed from price to design, then he/she will modify the existing merchandising strategy for such a group of customers. The methodology suggested in this paper detects changes automatically from customer profiles and sales data at different periods of time. The most common approach to discovering changes between two data sets is to generate decision trees from each data set and directly compare the rules from the decision trees by rule matching. But this is not a simple process for the following reasons. First, some rules cannot be easily compared due to different rule structures. Second, even with matched rules, it is difficult to know what kind of change and how much change has occurred. To simplify these difficulties, we first define three types of changes as the emerging pattern, the unexpected change and the added/perished rule. Then we develop similarity and difference measures for rule matching to detect all types of change from different time snapshot data. Finally, the degree of change is evaluated to detect significantly changed rules.

2. Background

2.1. Decision trees

Classification using decision trees can be used to extract models describing important data classes or to predict future data trends. The example of a classification model using decision trees is the bank loan model which characterizes customers as either safe or risky. Classification and prediction have numerous other applications including credit approval, medical diagnosis, performance prediction and selective marketing. Data classification is a two-step process as explained in Figure 1.

In the first step, a model is built describing a predetermined set of data classes or concepts. The model is constructed by analyzing database tuples described by attributes. Each tuple is assumed to belong to a predefined class, as determined by one of the attributes called the class label attribute. Typically, the learned or trained model is represented in the form of classification rules, decision trees or mathematical formulae. For example, given a database of customer credit information, classification rules can be learned to identify customers as having either excellent or fair credit ratings (see Figure 1(a)). The rules can be used to categorize future data samples, as well as provide a better understanding of the database contents. In the second step in Figure 1(b), the model is used to classify future data tuples or objects for which the class label is not known. For example, the classification rules learned in Figure 1(a) can be used to predict the credit rating of new or future (i.e. previously unseen) customers. A decision tree is a flow-chart-like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and leaf nodes represent classes or class distributions. The decision trees can easily be converted to classification rules. When the decision trees are built, many of the branches may reflect noise or outliers in the training data. Tree pruning attempts to identify and remove such branches, with the goal of improving classification accuracy on unseen data.

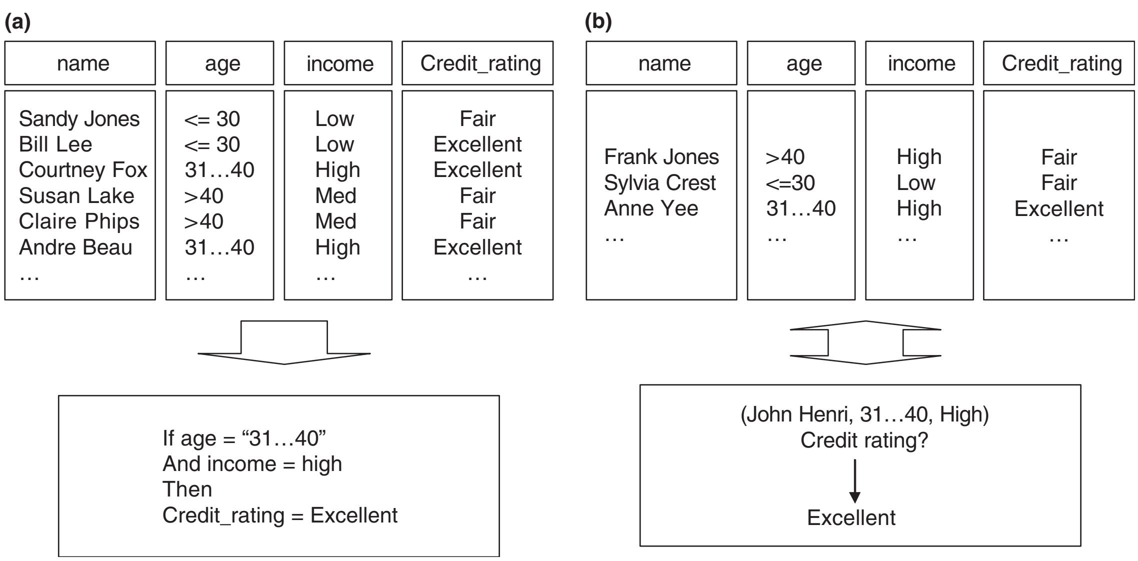

Figure 1: *Data classification process.*

2.2. Data mining in a changing environment

There is existing work on learning (Helmbold & Long, 1994; Widmer, 1996; Freund & Mansour, 1997) and mining (Bay & Pazzani, 1999; Ganti *et al*., 1999; Liu *et al*., 2000; Han & Kamber, 2001) in a changing environment. All the following related studies focus on dynamic aspects or comparison between two different data sets or rules. They are clustered as six categories in this paper.

The first field of study that examines mining in a changing environment is rule maintenance (Cheung *et al*., 1996; Feldman *et al*., 1997; Thomas *et al*., 1997). The purpose of these studies is to improve accuracy in a changing environment. For example, in the study of Cheung *et al*. (1996), incremental updating techniques are proposed for the efficient maintenance of discovered association rules when new transaction data are added to a transaction database. But these techniques do not provide any changes for the user. They just maintain existing knowledge.

The second research trend is to discover emerging patterns (Agrawal & Psaila, 1995; Dong & Li, 1999; Li *et al*., 2000), which are defined as item sets whose supports increase significantly from one data set to another. Emerging patterns can capture emerging trends in time-stamped databases or useful contrasts between data classes. But they do not consider the structural changes in the rules. For example, in a market basket, these techniques can discover significant rule changes which increase the growth rate or decrease the rate of consumption over time but cannot detect any unexpected changes such as a change from coffee $\Rightarrow$ tea to coffee $\Rightarrow$ milk.

Another related area of research is subjective interestingness in data mining (Liu & Hsu, 1996; Silberschatz & Tuzhilin, 1996; Liu *et al*., 1997; Suzuki, 1997; Padmanabhan & Tuzhilin, 1999). These papers give a number of techniques for finding unexpected rules with respect to the user's existing knowledge. This technique cannot be used for detecting changes as its analysis only compares each newly generated rule with each existing one to find degrees of difference. It does not find which aspects have changed, what kinds of changes have taken place and how much change has occurred.

The fourth research stream is mining from time-series data. There is increasing interest in discovering regularity in time-series data (Das *et al*., 1997, 1998; Han *et al*., 1999). Das *et al*. (1998) consider the problem of finding rules relating patterns in a time series to other patterns in that series, or patterns in one series to patterns in another series. Han *et al*. (1999) present several algorithms for efficient mining of partial periodic patterns, by exploring some interesting properties related to partial periodicity. Das *et al*. (1997) also present an intuitive model for measuring the similarity between two time series. But these studies are rather different from our research which focuses on the detection of irregularity rather than regularity in data.

The fifth research field is mining class comparisons to discriminate between different classes (Bay & Pazzani, 1999; Ganti *et al*., 1999; Han & Kamber, 2001). Ganti *et al*. (1999) present a general framework for measuring changes in two models. Essentially, the difference between two models is quantified as the amount of work required to transform one model into the other. Their framework covers a wide variety of models, including frequent item sets, decision tree classifiers and clusters, and captures standard measures of deviation such as misclassification rate and the χ^2 metric as special cases. But it cannot be directly applied to detect customer behavior changes because it does not provide which aspects have changed and what kind of changes have occurred. Bay and Pazzani (1999) and Han and Kamber (2001) also provide techniques for understanding the differences between several contrasting groups. But these techniques can only detect change about the same structured rule.

Finally, Liu *et al*. (2000) present a technique for change mining by overlapping two decision trees generated from different time snapshots. In a decision tree, each path from the root node to a leaf node represents a hyper rectangle region. A decision tree essentially partitions the data space into different class regions. Changes in the decision tree model thus mean changes in the partition and changes in the error rate. Our object in change mining is to discover the changes in the new data with respect to the old data and the old decision tree, and present the user with the exact changes that have occurred. However, their research cannot identify how much change has occurred. If there are a large number of changes, their methodology cannot rank what the most significant changes are. Song *et al*. (2001) developed a change detection system based on association rule mining. However, the application of association rule mining and decision trees is different. Change detection based on association rule mining can be used to identify changes of customer behavior in unstructured and ill-defined situations because of the unsupervised learning feature of association rule mining. However, using decision trees in the change detection problem can be used in more structured situations in which the manager has specific research questions and it also detects the change of classification criteria in a dynamically changing environment.

3. Problem

In this section, we examine all possible types of change based on past research and business requirements (Liu & Hsu, 1996; Liu *et al*., 1997; Suzuki, 1997; Dong & Li, 1999; Lanquillon, 1999; Padmanabhan & Tuzhilin, 1999; Song *et al*., 2001). After that, each type of change and change detection problem is defined. Let us define the following notation.

D^t, D^{t+k}	data sets at time t, $t+k$
R^t, R^{t+k}	discovered decision tree rule sets at time t, $t+k$
r_i^t, r_j^{t+k}	each rule from the corresponding rule set R^t, R^{t+k}, where $i=1,2,\ldots,\lvert R^t\rvert$, $j=1,2,\ldots,\lvert R^{t+k}\rvert$
$\sup^t(r_i)$	support of r_i in time t data set

Dong and Li (1999) introduced the emerging patterns concept that captures significant changes and differences between data sets. The emerging patterns are defined as item sets whose supports increase significantly from one data set to another. More specifically, the emerging patterns are item sets whose growth rates are larger than a given threshold value. When applied to time-stamped databases, the emerging patterns can capture emerging trends in business or demographic data. We bring from the study of Dong and Li (1999) the term emerging pattern with the following modified definition for our research.

Definition 1: Emerging patterns For rule r_j^{t+k}, if the following two conditions are met, then we call it the rule of emerging patterns with respect to r_i^t.

(1) Conditional and consequent parts are the same between r_i^t, r_j^{t+k}.
(2) Supports of the two rules are significantly different.

Example 1

r_i^t: If Income = High, Age = High, then Sales amount = High (support = 0.1)
r_j^{t+k}: If Income = High, Age = High, then Sales amount = High (support = 0.13)

In this case, r_j^{t+k} is the emerging pattern with respect to r_i^t if we specify the minimum growth rate to be 0.2. This is because the two rules have the same rule structure and their growth rate is 0.3.

The other type of change is unexpectedness, which is found from many studies about discovering interesting patterns (Liu & Hsu, 1996; Silberschatz & Tuzhilin, 1996; Liu *et al*., 1997; Suzuki, 1997; Padmanabhan & Tuzhilin, 1999). The unexpected changes can be found from newly discovered rules, which are different from users' existing beliefs. Liu and Hsu (1996) defined the unexpected changes as rule similarity and difference aspects. The approach is based on a syntactic comparison between a rule and a belief. In their study, a rule and a belief are 'different' if the consequents are 'far apart' but the conditions are 'similar', where 'similarity' and 'difference' are defined syntactically based exclusively on the structure of the rules. We redefine unexpected changes adapting from the study by Liu and Hsu (1996).

Definition 2: Unexpected changes (or unexpected consequent changes) r_j^{t+k} is unexpected change with respect to r_i^t if the conditional parts of r_i^t, r_j^{t+k} are similar but the consequent parts of the two rules are quite different.

Example 2

r_i^t: If Income = High, Age = High, then Sales amount = High
r_j^{t+k}: If Income = High, Age = High, then Sales amount = Low

In this case, r_j^{t+k} is an unexpected consequent change with respect to r_i^t since the conditional parts of r_i^t, r_j^{t+k} are similar but the consequent parts of the two rules are quite different.

Other types of change are the added rules and the perished rules (Lanquillon, 1999). An added rule is a newly arisen rule which could not be found in the past and a perished rule is a disappeared rule, which can only be found in the past but not the present. The added and perished rules are defined as follows.

Definition 3: Added rules/Perished rules r_j^{t+k} is an added rule if all the conditions and consequents are quite different from any of r_i^t in R^t, and r_i^t is a perished rule if all the conditions and consequents are quite different from any of r_j^{t+k} in R^{t+k}.

We used the terms 'similar' and 'quite different' in the above definitions. These terms are used to compare two rules in syntactic aspects and to judge their degree of similarity and difference. But the terms 'similar' and 'quite different' are subjective and different for each individual. Therefore we define the *rule matching threshold* (RMT) which can be differently determined by individual users. Finally, we define the *degree of change* as the measure of how much change has occurred. Evaluation of the degree of change will be explained in the next section. Now, the change detection problem is defined as follows using the above definitions of each change type.

Definition 4: Change detection problem The change detection problem consists of finding all emerging patterns, unexpected changes and added/perished rules between data sets which are collected from different periods and ranking the changed rules in each type by the degree of change.

4. Methodology

4.1. Overall procedure

The methodology for detecting the change in customer behavior consists of the following three phases, as illustrated in Figure 2.

In phase I, two rule sets are generated from each data set by using decision tree analysis. For this purpose, we present two basic approaches to change mining in the decision tree model: a new decision tree, and the same attribute and same cut. Details are explained in the following section.

In phase II, the changed rule set is generated using the rule matching method which compares two rules selected from each rule set. We adapted the rule matching method developed by Liu and Hsu (1996) and modified it to detect all types of changed rules including emerging patterns, unexpected changes, and added and perished rules. For efficient rule matching, similarity and difference measures are developed.

In phase III, various changed rules detected in phase II are ranked according to the predefined degree of change that is a measure of how much change has occurred. In the case of emerging patterns, the growth rate or rate of decrease of the *support* value of the changed rule is computed and then the rules are sorted by the absolute value of the rate. The growth rate or rate of decrease plays the role of the degree of change in an emerging pattern. Second, in the case of unexpected change, we adapted the unexpectedness concept from the study of Padmanabhan and Tuzhilin (1999). The degree of exception for certain existing rules plays the role of degree of change in the case of unexpected consequent change. Third, in the case of added or perished rules, we use the support value and maximum similarity value of new or disappeared rules as the degree of change. Maximum similarity value is defined in a later section, and more detailed explanations for the measures and procedures are provided in the following.

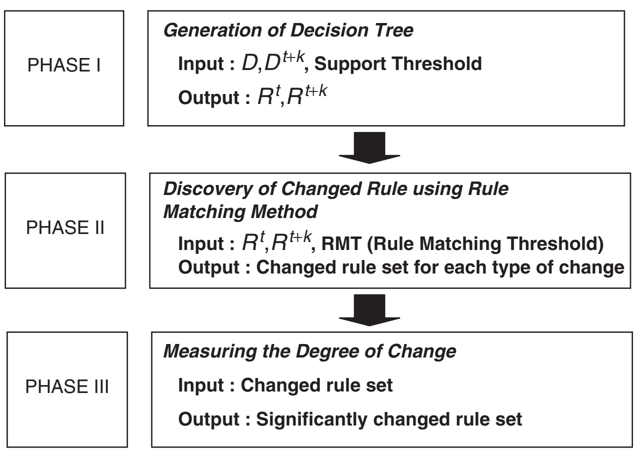

Figure 2: *Overall procedure to detect change.*

4.2. Discovery of rule change in the decision tree

In this paper, two approaches are used to mine changes in the decision tree model: a new decision tree, and the same attribute and same cut (Liu *et al.*, 2000). The first approach modifies the original tree structure, which makes the comparison with the original structure difficult. In the second approach it is easy to compare the two decision trees, but the support values of the same rule of each tree might be different. Note that the basic decision tree algorithm we use in our study is C4.5. We have modified it in various places for change mining purposes.

4.2.1. Same attribute and same cut A decision tree is developed with the data set of time t. At time $t+k$, not only is the same attribute as in the tree of time t used but also the same cut point in the tree of time t at each step of the partitioning. If the algorithm has not reached the leaf node of a particular branch of the tree, the tree of time $t+k$ needs to go beyond the depth of the corresponding branch in the tree of time t. Rule sets of time t and time $t+k$ are generated based on the decision trees of time t and time $t+k$ respectively. In the case of the same attribute and same cut in time $t+k$, a similar procedure is performed but a decision tree of time $t+k$ is developed first.

4.2.2. New decision tree Decision trees are generated respectively using the data set of time t and that of time $t+k$. Two rule sets are generated from the two decision trees of time t and time $t+k$. The added rules or perished rules are usually found on the basis of this method.

4.3. Discovery of a changed rule

In this phase, three types of changed rules are detected by rule matching based on syntactic comparison. The inputs of phase II are discovered rule sets at time t and $t+k$ and the RMT which is specified by the user. For the explanation of our rule matching method, some notations are briefly defined.

δ_{ij} difference measure: degree of difference between r_i^t and r_j^{t+k} ($-1 \le \delta_{ij} \le 1$)

s_{ij} similarity measure: degree of similarity between r_i^t and r_j^{t+k} ($0 \le s_{ij} \le 1$)

ℓ_{ij} degree of attribute match of the conditional parts

$$\ell_{ij} = \frac{|A_{ij}|}{\max(|X_i^t|, |X_j^{t+k}|)}$$

c_{ij} degree of attribute match of the consequent parts

$$c_{ij} = \begin{cases} 1 & \text{if same consequent attribute} \\ 0 & \text{otherwise} \end{cases}$$

$|A_{ij}|$ number of attributes common to the conditional parts of both r_i^t and r_j^{t+k}

$|X_i^t|$ number of attributes in the conditional parts of r_i^t

$|X_j^{t+k}|$ number of attributes in the conditional parts of r_j^{t+k}

x_{ijk} degree of value match of the kth matching attribute in A_{ij}

$$x_{ijk} = \begin{cases} 1 & \text{if same value} \\ 0 & \text{otherwise} \end{cases}$$

y_{ij} degree of value match of the consequent attribute

$$y_{ij} = \begin{cases} 1 & \text{if same value} \\ 0 & \text{otherwise} \end{cases}$$

Now we provide a *similarity measure* as follows, adapted from the study of Liu and Hsu (1996).

$$s_{ij} = \begin{cases} \dfrac{\ell_{ij} \sum_{k \in A_{ij}} x_{ijk} c_{ij} y_{ij}}{|A_{ij}|} & \text{if } |A_{ij}| \neq 0 \\ 0 & \text{if } |A_{ij}| = 0 \end{cases}$$

In $s_{ij}, \ell_{ij} \sum_{k \in A_{ij}} x_{ijk}/|A_{ij}|$ represents a similarity of the conditional part and $c_{ij}y_{ij}$ represents a similarity of the consequent part between r_i^t and r_j^{t+k}. If the conditional and consequent parts between r_i^t and r_j^{t+k} are the same, then the degree of similarity becomes 1. The similarity measure can take any value between 0 and 1. Also the *maximum similarity value* is defined as follows.

$$s_i = \max(s_{i1}, s_{i2}, \ldots, s_{i|R_{t+k}|})$$

maximum similarity value of r_i^t

$$s_j = \max(s_{1j}, s_{2j}, \ldots, s_{|R_t|j})$$

maximum similarity value of r_j^{t+k}

The maximum similarity value indicates whether the rule is added or perished. If $s_i < \text{RMT}$, then r_i^t is recognized as a perished rule. If $s_j < \text{RMT}$, then the rule r_j^{t+k} becomes an added rule.

Unexpected change consists of unexpected conditions and unexpected consequents. To detect an unexpected change, a *difference measure* is provided as follows.

$$\delta_{ij} = \begin{cases} \dfrac{\ell_{ij} \sum_{k \in A_{ij}} x_{ijk}}{|A_{ij}|} - y_{ij} & \text{if } |A_{ij}| \neq 0, c_{ij} = 1 \\ -y_{ij} & \text{if } |A_{ij}| = 0, c_{ij} = 1 \end{cases}$$

As defined in the problem definition section, if conditional parts are similar but consequent parts are different, then this rule is called an unexpected consequent. If the rule r_j^{t+k} becomes an unexpected consequent change with respect to r_i^t, the similarity in the conditional part and the difference in the consequent part should be large. Therefore, if $\delta_{ij} > 0$, then rule r_j^{t+k} is an unexpected consequent change with respect to r_i^t. If $\delta_{ij} < 0$, then rule r_j^{t+k} is an unexpected condition change with respect to r_i^t. If $\delta_{ij} = 0$, then the two rules r_i^t and r_j^{t+k} are the same rules or completely different rules. Additional measures such as ℓ_{ij}, y_{ij} etc. should be provided in the case $\delta_{ij} = 0$. If these values are 1 then we can directly find that the two rules are the same. We compute difference measures between two rules r_i^t and r_j^{t+k} only in the case of $c_{ij} = 1$. If attributes of consequent parts between two rules are different, it makes no sense to compare the degree of difference because these two rules are completely different rules.

Finally, we should resolve duplication of type. For example, although r_j^{t+k} is judged to be an unexpected change with regard to r_i^t by the difference measure, we cannot conclude directly whether it is an unexpected change or not, because r_j^{t+k} can be an emerging pattern with regard to r_j^t which has the same structure as r_j^{t+k}. In this case, r_j^{t+k} should be classified as an emerging pattern and not as an unexpected change. As we cannot conclude based on δ_{ij} alone whether r_j^{t+k} is an unexpected change or an emerging pattern, we provide the following modified difference measure.

$$\delta'_{ij} = |\delta_{ij}| - k_{ij}$$

where

$$k_{ij} = \begin{cases} 1 & \text{if } \max(s_i, s_j) = 1 \\ 0 & \text{otherwise} \end{cases}$$

The fact that s_i (or s_j) is equal to 1 means that the same rule exists in another rule set. That means that r_j^{t+k} is likely to be classified as an emerging pattern. If δ'_{ij} is greater than the pre-specified RMT, then the rule r_j^{t+k} is concluded to be an unexpected change with respect to r_i^t. Table 1 summarizes the value of each measure for each type of change.

4.4. Evaluating the degree of change

All the changed rules have to be ranked by the degree of change. We will explain the idea of evaluating the degree of change for each type of change. First, let us consider an unexpected change. The following example presents why additional measures are required.

r_i^t: If Income = High, Age = High, then Sales amount = High

r_j^{t+k}: If Preference = Price, Age = High, then Sales amount = Low

If RMT is set equal to 0.4, then the rule r_j^{t+k} becomes an unexpected consequent change with respect to r_i^t as $\delta_{ij} = 0.5$. But there are two problems in concluding whether this change is significant or not. First, we cannot capture the change easily because the conditional parts are not the same. Second, although we can understand the change, we

Table 1: *Value of measure for each type of change*

Type of change	*Value of measure to classify*
Emerging pattern	$\delta_{ij} = 0, (\sum_{k \in A_{ij}} x_{ijk} > 0$ or $y_{ij} > 0$ or $\ell_{ij} > 0)$
Unexpected consequent	$\delta_{ij} > 0, \delta'_{ij} \geq$ RMT
Unexpected condition	$\delta_{ij} < 0, \delta'_{ij} \geq$ RMT
Added rule	$s_j <$ RMT
Perished rule	$s_i <$ RMT

do not know how much change has occurred. Therefore additional logical judgement is required to conclude whether the degree of change is significant or not. For this purpose, we adapt the unexpectedness concept from the study of Padmanabhan and Tuzhilin (1999). They define unexpectedness using the exception rule concept (Suzuki, 1997; Hussain *et al.*, 2000) as follows.

Definition 5: Unexpectedness If a rule from a decision tree $A \Rightarrow B$ is unexpected with respect to the belief $X \Rightarrow Y$, then the following must hold.

(1) B and Y = False.
(2) The rule X, $A \Rightarrow B$ holds.

A new measure for the degree of change of an unexpected consequent change is defined using Definition 5. To measure the degree of unexpected consequent change, r_i^t is assumed to be a belief or existing knowledge. Every unexpected consequent change satisfies condition (1) of Definition 5 because of Definition 2. Furthermore the support value of the conjunction rule should be evaluated to check whether condition (2) of Definition 5 holds or not. For example, a conjunction rule of the above example is as follows.

$r_{i \cap j}$: Income = High, Age = High, Preference = Price
$\Rightarrow$ Sales amount = Low

If the above conjunction rule, $r_{i \cap j}$, is statistically large (i.e. has a large support value), then we can conclude that r_j^{t+k} is an unexpected consequent change with respect to r_i^t by condition (2) of Definition 5. Therefore, the support value of the conjunction rule can be regarded as the degree of change for unexpected consequent change. But the two conditions of Definition 5 are not sufficient. If the support value of the conjunction rule is relatively small by comparison with the support value of r_j^{t+k}, then we cannot conclude that r_j^{t+k} is a significant unexpected consequent change with respect to r_i^t. An additional condition that should be included is that the support value of $r_{i \cap j}$ should be large enough to represent r_j^{t+k}. Therefore, the degree of change for unexpected consequent change should be composed of the support value of r_j^{t+k} and $r_{i \cap j}$. Now we provide the following measure for the degree of unexpected consequent change.

$$\alpha_{ij} = \frac{\sup^{t+k}(r_{i \cap j})}{\sup^{t+k}(r_j)}$$

In the case of an emerging pattern, it is simpler to evaluate the significance level than the case of unexpected change. The growth rate or rate of decrease are used as the measure for this type of change. To evaluate the degree of change for the added and perished rule cases, the support value of these rules and the maximum similarity value are used. As mentioned before, the maximum similarity value of a rule represents the degree of similarity of the most similar rule to the other rule set. If there is a situation that the support values of two added rules are the same, we naturally place more importance on the rule which has less maximum similarity value. Such a rule gives more significance than the other rule. The measure of the degree of change, α_{ij}, is summarized as follows. Based on the value of α_{ij}, we can rank the changed rules in each type of change.

$$\alpha_{ij} = \begin{cases} \dfrac{\sup^{t+k}(r_i) - \sup^t(r_i)}{\sup^t(r_i)} & \text{emerging pattern case} \\ \dfrac{\sup^{t+k}(r_{i \cap j})}{\sup^{t+k}(r_j)} & \text{unexpected change case} \\ (1 - s_i)\sup^t(r_i) & \text{perished rule case} \\ (1 - s_j)\sup^{t+k}(r_j) & \text{added rule case} \end{cases}$$

5. Experiments and applications

A case study has been conducted to evaluate how well the procedure performs its intended task of detecting significant changes. The data set was prepared from a Korean online shopping mall which sells various consumer goods. The data set contains customer profiles and purchasing history such as age, job, sex, address, registration year, cyber money, number of purchases, number of visits, payment method during one year, and total purchase amount. We prepared two data sets to detect significant changes of purchasing behavior by the customers. The first data set contains profiles and purchasing history information of certain customers who had bought more than one cosmetic from 1 February 2000 to 30 June 2000. The second data set contains the same information but includes customers who had made an additional purchase of cosmetics from 1 July 2000 to 5 January 2001. After preprocessing the data for cleansing and discretization, we built the decision tree. The splitting criterion is the χ^2 test. The procedure to detect change is explained according to three categories.

5.1. Same attribute and same cut at time t

We processed the data at time t for cleansing and discretization and built the decision tree and corresponding rules with the SAS 8.0 enterprise miner program. And with the same attribute and same cut, we built the decision tree and corresponding rules of the data set at time $t+k$. Then, we compared the rules of time t and $t+k$ and discovered the emerging pattern, the unexpected changes and the added/perished rules. We evaluated the changed rules to find what rules had changed most. Figures 3 and 4 show the decision trees of time t and time $t+k$. The emerging patterns, unexpected changes and added/perished rules are shown in Table 2. Significant emerging patterns, unexpected changes and added/perished rules are summarized in Tables 3, 4 and 5.

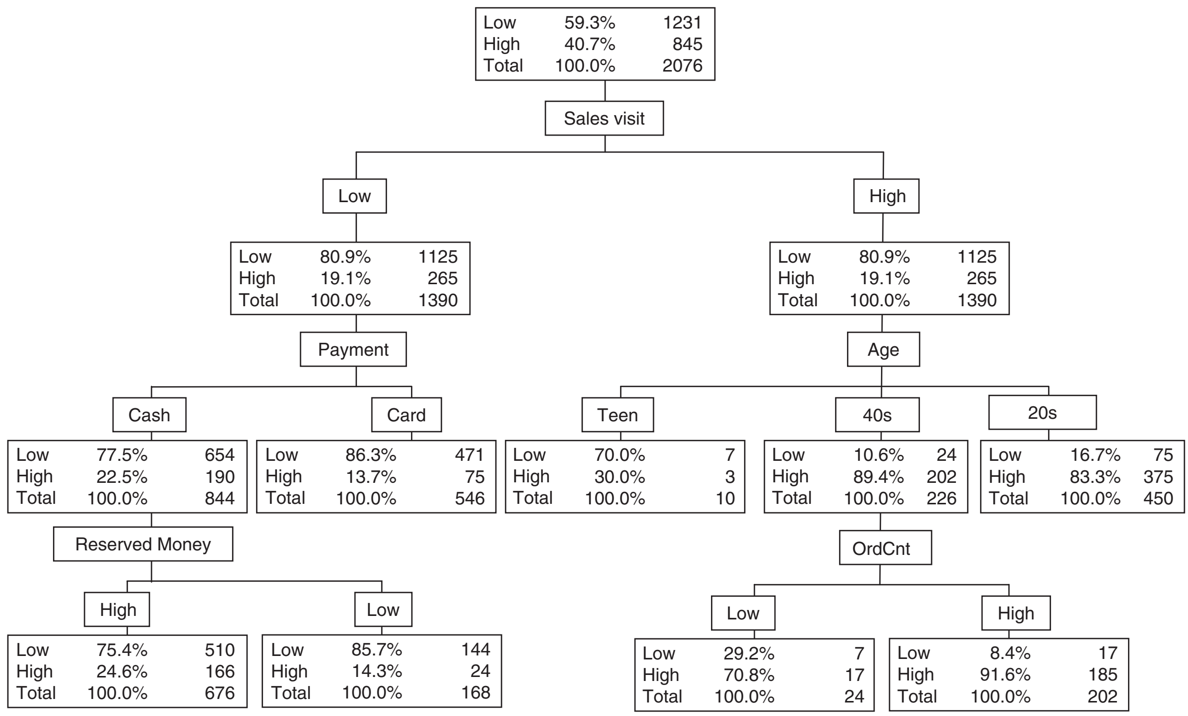

Figure 3: *Decision tree of the data set at time t in the same attribute and the same cut.*

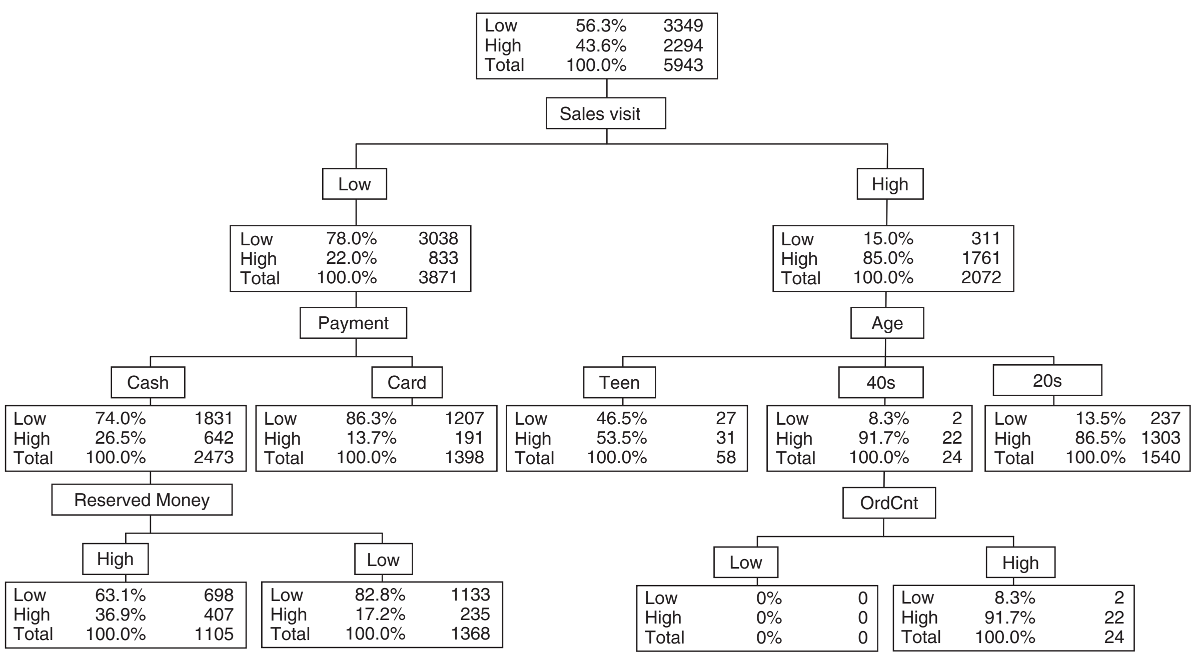

Figure 4: *Decision tree of the data set at time t + k in the same attribute and the same cut.*

From changed rule (1) in Table 3, we can see the rapid decrease of customers with a high purchasing rate (96% rate of decrease) in sales for customers who are in their 40s, order frequently and visit the mall frequently. Because of the large rate of decrease, this segment of customers will disappear in the near future. Therefore special careful consideration of these customers will be necessary. From changed rule (2) in Table 3, we find a rapid growth of customer groups with a high purchasing rate in sales for customers who are in their 20s and visit the mall frequently. This tendency is obviously desirable; therefore a marketing campaign to encourage the revisiting of these customers should be developed. With regard to the unexpected changes of Table 4, the sales for customers who are in their teens and visit the mall frequently are low from the first data set. But in the second data set, we can see that the

Table 2: *Number of changed rules in the same attribute and the same cut at time t*

Type of change	*Number of changed rules*	*Number of significant changed rules*
Emerging patterns	4	3 (degree of change > 0.1)
Unexpected changes	1	1 (degree of change > 0.1)
Added/Perished rules	1	1 (degree of change > 0.001)

Table 4: *Significant unexpected changes (degree of change > 0.1) in the same attribute and the same cut at time t*

	r_i^t	r_j^{t+k}	δ_{ij}	δ'_{ij}	α_{ij}
1	If Sales visit = High, Age = Teen, then Sales amount Low Support = 0.003	If Sales visit = High, Age = Teen, then Sales amount = High Support = 0.005	0.5	0.5	0.6

Table 5: *Significant perished rules (degree of change > 0.001) in the same attribute and the same cut at time t*

	r_i^t	*MSV*	*Support*	α_{ij}
1	If Sales visit = High, Age = 40s, Order count = Low, then Sales amount = High	0	0.008	0.008

Table 3: *Significant emerging patterns (degree of change > 0.1) in the same attribute and the same cut at time t*

No.	*Rules*	$\sup^t(r_i)$	$\sup^{t+k}(r_j)$	α_{ij}
1	If Sales visit = High, Age = 40s, Order count = High, then Sales amount = High	0.09	0.00	0.9585
2	If Sales visit = High, Age = 20s, then Sales amount = High	0.18	0.22	0.2138
3	If Sales visit = Low, Payment = Card, then Sales amount = Low	0.23	0.20	0.1048

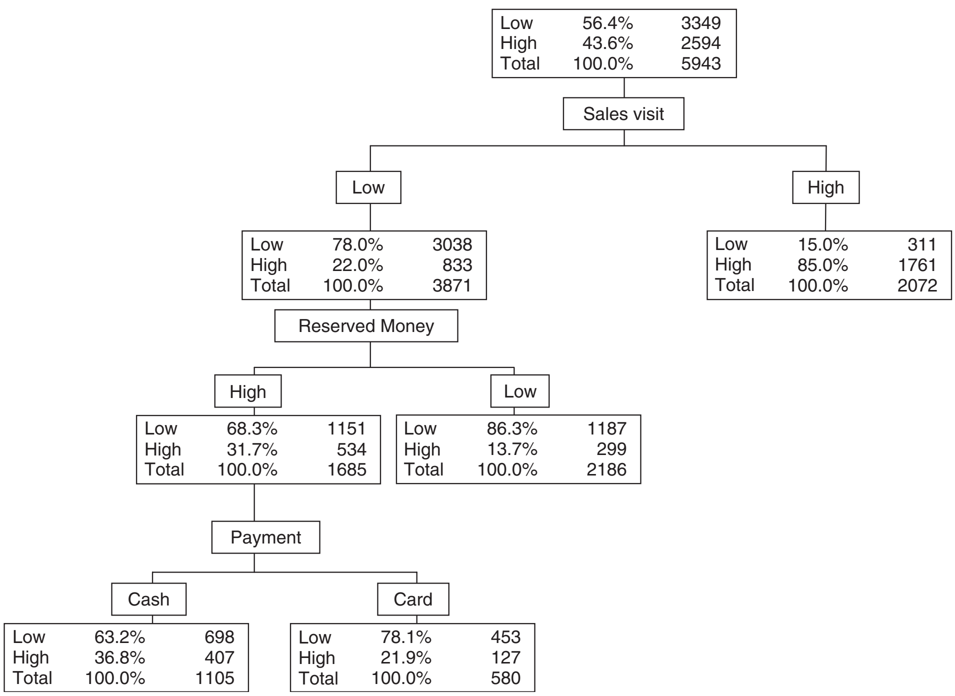

Figure 5: *Decision tree of the data set at time t + k in the same attribute and the same cut.*

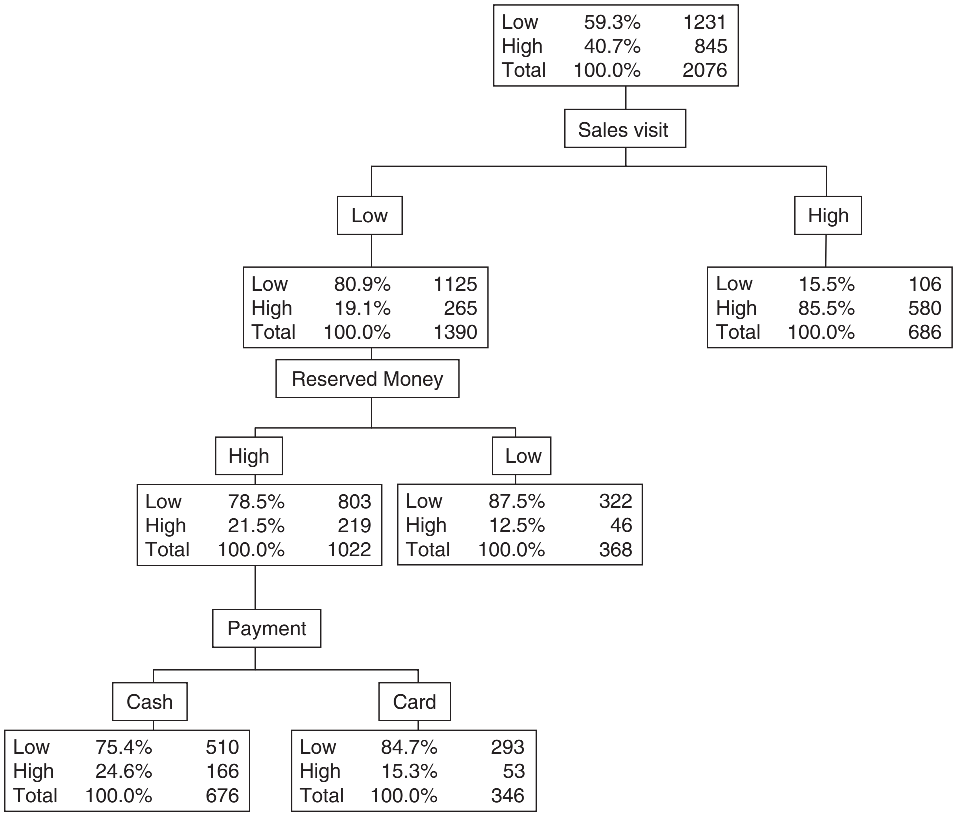

Figure 6: *Decision tree of the data set at time t in the same attribute and the same cut.*

Table 6: *Number of changed rules in the same attribute and the same cut at time $t+k$*

Type of change	*Number of changed rules*	*Number of significant changed rules*
Emerging patterns	4	3 (degree of change >0.1)
Unexpected changes	0	0 (degree of change >0.1)
Added/Perished rules	0	0 (degree of change >0.001)

Table 7: *Significant emerging patterns (degree of change >0.1) in the same attribute and the same cut at time $t+k$*

		$\sup^t(r_i)$	$\sup^{t+k}(r_j)$	α_{ij}
1	If Sales visit = Low, Reserved money = Low, then Sales amount = Low	0.57	0.19	0.6666
2	If Sales visit = Low, Reserved money = High, Payment = Cash, then Sales amount = Low	0.25	0.12	0.5219
3	If Sales visit = Low, Reserved money = High, Payment = Card, then Sales amount = Low	0.14	0.08	0.4599
4	If Sales visit = High, then Sales amount = High	0.33	0.35	0.0551

sales for these customers are high. This means that the importance of such customers is gradually increasing. Therefore, a modification of the existing marketing strategy and plan for the customers is required. Finally, one perished rule is found in Table 5. From February to June in 2000, the sales of frequently visiting customers, in their 40s and who order infrequently were high. But after June 2000, we cannot find these rules anymore. Therefore we should decide whether additional services and products for elders need to be developed.

5.2. Same attribute and same cut at time $t+k$

As with the case of the same attribute and same cut at time t, we built the decision tree and rules of the data set at time $t+k$ with the SAS 8.0 enterprise miner program. And with the same attribute and same cut of the tree, we built the decision tree and rules of the data set at time t. Then, we compared the two rule sets of time $t+k$ and t to discover the emerging pattern, unexpected changes and added/perished rules. Figures 5 and 6 show the decision trees of time $t+k$ and time t respectively. The emerging patterns, unexpected changes and added/perished rules of the same attribute and the same cut at time t are summarized in Table 6. Significant emerging patterns are summarized in Table 7.

From changed rule (1) in Table 7, we can see the rapid decrease of sales in the customer group who have a little cyber money and visit the mall infrequently. In other words, these customers are likely to have disappeared in the near future because of the high rate of decrease. Therefore an urgent market strategy is necessary to prevent the rapid sales decrease of these customers. With regard to unexpected changes and added/perished rules, we identified no significant changes.

Table 8: *Number of changed rules in the new decision tree at time t and time $t+k$*

Type of change	*Number of changed rules*	*Number of significant changed rules*
Emerging patterns	1	1 (degree of change >0.1)
Unexpected changes	0	0 (degree of change >0.1)
Added/Perished rules	9	9 (degree of change >0.001)

Table 9: *Significant emerging patterns (degree of change >0.1) in the new decision tree at time t and time $t+k$*

		$\sup^{t}(r_i)$	$\sup^{t+k}(r_j)$	α_{ij}
1	If Sales visit = Low, Payment = Cash, Reserved money = High, then Sales amount = Low	0.24	0.11	0.5417

5.3. *New decision tree at time t and time $t+k$*

The decision trees are built with the data sets of time t and time $t+k$ independently using the SAS 8.0 enterprise miner program. And two rule sets are constructed along with the pathway of each decision tree. Then, we compared the rules of time t and $t+k$ and discovered the emerging patterns, unexpected changes and added/perished rules. We evaluated the changed rules to find what rules had changed most. Figures 3 and 5 represent the new decision trees of time t and time $t+k$ respectively. Table 8 shows the emerging patterns, unexpected changes and added/perished rules of the decision trees at time t and time $t+k$. Compared to the previous cases, new decision trees at each time generate more added rules and perished rules. Significant emerging patterns and added/perished rules at time t and time $t+k$ are summarized in Tables 9 and 10.

With regard to added/perished rules, we identified nine significant changes. From Table 10, we find that sales of customers who are in their 40s are increased in the second data set from rule 5 and rule 8. This means that the importance of customer groups who are in their 40s is increased. Therefore additional services and products for such customers should be developed.

6. Conclusion

In this paper, we developed a decision-tree-based methodology to detect changes of customer behavior automatically from customer profiles and sales data at different time snapshots. For this purpose, we found rule sets from decision trees and defined three types of change. Then we developed the similarity and difference measures for rule matching to detect all types of change in syntactic aspects. Additionally, the degree of change was evaluated to detect significantly changed rules in semantic aspects.

We summarize the opportunities of using this methodology and various applications in practical business perspectives as follows. First, with regard to macro aspects, business managers can follow the change trends using our suggested change detection methodology. They need to analyze their customers' changing behavior in order to

Table 10: *Significant added/perished rules (degree of change >0.001) in the new decision tree at time t and time $t+k$*

	r_i^t	*MSV*	*Support*	α_{ij}
1	If Sales visit = High, then Sales amount = High	0	0.2965	0.2965
2	If Sales visit = Low, Payment = Card, then Sales amount = Low	0	0.2269	0.2269
3	If Sales visit = Low, Reserved money = Low, then Sales amount = Low	0	0.1997	0.1997
4	If Sales visit = High, Age = 20s, then Sales amount = High	0	0.1806	0.1806
5	If Sales visit = High, Age = 40s, Order count = High, then Sales amount = High	0	0.0891	0.0891
6	If Sales visit = Low, Reserved money = Low, Payment = Card, then Sales amount = Low	0	0.0762	0.0762
7	If Sales visit = Low, Payment = Cash, Reserved money = Low, then Sales amount = Low	0	0.0694	0.0694
8	If Sales visit = High, Age = 40s, Order count = Low, then Sales amount = High	0	0.0082	0.0082
9	If Sales visit = High, Age = Teen, then Sales amount = Low	0	0.0034	0.0034

provide products and services that suit the changing needs of the customers. Second, with regard to micro aspects, it is possible for a manager to understand customer needs more deeply and design additional niche-marketing campaigns based on the rule sets of the suggested methodology. Knowing the purchasing history of a certain customer segment can give a better understanding of the behavior of the segment.

Change detection is more suitable in domains where the environment is relatively dynamic and there is much human intervention. Besides understanding customer behavior change, another promising application is analyzing the effectiveness of a marketing campaign. If a manager generates rules from the sales data set before and after a campaign, he/she can evaluate the effectiveness of his/her marketing campaign by comparing two rule sets using the suggested methodology. Furthermore, the suggested methodology can be used recursively in a time-series data set to analyze the change of classified customer segments.

Some limitations of the suggested methodology are as follows. With regard to the number of target data sets which should be compared, the methodology is suitable for only two. If there are three or more data sets to be compared over time, then a recursive methodology will have to be developed. The methodology is run on data sets which have discrete values. If there is a data set which has continuous values, then a preprocessing step to make discrete values is needed.

As a further research area, we plan to extend our methodology to discover changes over three or more data sets. It will be a promising research area to set up a campaign management plan based on our suggested methodology, and it will also be interesting to check the effectiveness of the campaign.

Acknowledgement

Hee Seok Song contributed as a corresponding author for this paper.

References

AGRAWAL, R. and G. PSAILA (1995) Active data mining, in *Proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD-95)*, 3–8.

BAY, S.D. and M.J. PAZZANI (1999) Detecting change in categorical data: mining contrast sets, in *Proceedings of the Fifth International Conference on Knowledge Discovery and Data Mining (KDD-99)*, 302–306.

CHEUNG, D.W., J. HAN, V.T. NG and C.Y. WONG (1996) Maintenance of discovered association rules in large databases: an incremental updating technique, in *Proceedings of the Twelfth International Conference on Data Engineering (ICDE-96)*, 106–114.

DAS, G., D. GUNOPULOUS and H. MANNILA (1997) Finding similar time series, in *Proceedings of the First European Symposium: Principles of Data Mining and Knowledge Discovery (PKDD '97)*, 88–100.

DAS, G., K. LIN, H. MANNILA, G. RENGANATHAN and P. SMYTH (1998) Rule discovery from time series, in *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98)*, 16–22.

DONG, G. and J. LI (1999) Efficient mining of emerging patterns: discovering trends and differences, in *Proceedings of the Fifth International Conference on Knowledge Discovery and Data Mining (KDD-99)*, 43–52.

FELDMAN, R., Y. AUMANN, A. AMIR and H. MANILA (1997) Efficient algorithms for discovering frequent sets in incremental databases, in *SIGMOD '97 Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD '97)*, 59–66.

FREUND, Y. and Y. MANSOUR (1997) Learning under persistent drift, in *Computational Learning Theory: Third European Conference*, 109–118.

GANTI, V., J. GEHRKE and R. RAMAKRISHNAN (1999) A framework for measuring changes in data characteristics, in *Proceedings of the Eighteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (PODS-99)*, 126–137.

HAN, J. and M. KAMBER (2001) *Data Mining: Concepts and Techniques*, San Francisco, CA: Morgan Kaufmann, 200–207.

HAN, J., G. DONG and Y. YIN (1999) Efficient mining of partial periodic patterns in time series database, in *Proceedings of the Fifteenth International Conference on Data Engineering (ICDE 1999)*, 106–115.

HELMBOLD, D.P. and P.M. LONG (1994) Tracking drifting concepts by minimizing disagreements, *Machine Learning*, **14**, 27–45.

HUSSAIN, F., H. LIU, E. SUZUKI and H. LU (2000) Exception rule mining with a relative interestingness measure, in *Proceedings of Pacific Asia Conference on Knowledge Discovery in Databases (PAKDD)*, 86–97.

LANQUILLON, C. (1999) Information filtering in changing domains, in *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI99)*, 41–48.

LI, J., G. DONG and K. RAMAMOHANARAO (2000) Making use of the most expressive jumping emerging patterns for classification, in *Proceedings of Pacific Asia Conference on Knowledge Discovery in Databases (PAKDD)*, 220–232.

LIU, B. and W. HSU (1996) Post-analysis of learned rules, in *Proceedings of the Thirteenth National Conference on Artificial Intelligence (AAAI-96)*, 828–834.

LIU, B., W. HSU and S. CHEN (1997) Using general impressions to analyze discovered classification rules, in *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining (KDD-97)*, 31–36.

LIU, B., W. HSU, H.S. HAN and Y. XIA (2000) Mining changes for real-life applications, in *Proceedings of the Second International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2000)*, Y. Kambayashi, M.K. Mohania and A.M. Tjoa (eds), 337–346.

PADMANABHAN, B. and A. TUZHILIN (1999) Unexpectedness as a measure of interestingness in knowledge discovery, *Decision Support Systems*, **27**, 303–318.

SILBERSCHATZ, A. and A. TUZHILIN (1996) What makes patterns interesting in knowledge discovery systems, *IEEE Transactions on Knowledge and Data Engineering*, **8** (6), 970–974.

SONG, H.S., J.K. KIM and S.H. KIM (2001) Mining the change of customer behavior in an Internet shopping mall, *Expert Systems with Applications*, **21** (3), 157–168.

SUZUKI, E. (1997) Autonomous discovery of reliable exception rules, in *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining (KDD-97)*, 259–262.

THOMAS, S., S. BODAGALA, K. ALSABTI and S. RANKA (1997) An efficient algorithm for the incremental update of association rules in large databases, in *Proceedings of the Third International Conference on Knowledge Discovery in Databases*, 263–266.

WIDMER, G. (1996) Learning in the presence of concept drift and hidden contexts, *Machine Learning*, **23**, 69–101.

The authors

Jae Kyeong Kim

Jae Kyeong Kim is a professor of e-business in the School of Business Administration in the Kyung Hee University in Seoul, Korea. He received his BS degree from Seoul National University, and both his MS degree and his PhD in industrial engineering from Korea Advanced Institute of Science and Technology. His areas of interest are data mining, e-customer relationship management and intelligent information systems. He has published numerous papers, which have appeared in *Applied Artificial Intelligence, Artificial Intelligence Review, Computers and Industrial Engineering, Computers and Operations Research, Electronic Commerce Research and Applications, European Journal of Operational Research, Expert Systems, Expert Systems with Applications, Group Decision and Negotiations, Intelligent Systems on Finance Management and Accounting, International Journal of Human–Computer Studies, International Journal of Internet and Enterprise Management, International Journal of Production Economics, Journal of Decision Systems, Lecture Notes in Artificial Intelligence, Lecture Notes in Computer Science* etc.

Hee Seok Song

Hee Seok Song is a professor in the Management Information System Department in Hannam University, Korea. He received his MA in management science and PhD in management engineering from Korea Advanced Institute of Science and Technology. He has been working for Daewoo Information Systems since 1989 and has worked on various system integration projects about customer relationship management and electronic commerce. His main research interests include data mining and customer relationship management. He has published in a number of international conferences and journals on artificial intelligence and data mining such as *Expert Systems with Applications*, *Artificial Intelligence Review* and the *Pacific Asian Conference on Intelligent Systems*.

Tae Seong Kim

Tae Seong Kim is a researcher of the e-Business Information Center at Kyung Hee University in Korea. He received a BS in international economics and an MS in management information systems from Kyung Hee University in Korea. His research interests include data mining, customer relationship management and e-business models.

Hyea Kyeong Kim

Hyea Kyeong Kim is a PhD student at Kyung Hee University in Korea. She received a BS in physics and an MS in e-business from Kyung Hee University. Her research interests include recommender systems, peer-to-peer networks and ubiquitous computing. She has published numerous papers, which have appeared in *International Conference on Electronic Business, Lecture Notes in Computer Science*, the *Journal of MIS Research* and so on.