A Post-analysis of the Association Rule Mining Applied to Internet Shopping Mall

Jae Kyeong Kim
School of Business Administration, KyungHee University, Seoul, Korea

Hee Seok Song
Graduate School of Management, KAIST, Seoul, Korea

Abstract

Understanding and adapting to changes of customer behavior is an important aspect for a company to survive in continuously changing environment. The aim of this paper is to develop a methodology which detects changes of customer behavior automatically from customer profiles and sales data at different time snapshots. For this purpose, we first define three types of changes as emerging pattern, unexpected change and the added / perished rule. Then, we develop similarity and difference measures for rule matching to detect all types of change. Finally, the degree of change is evaluated to detect significantly changed rules. Our proposed methodology can evaluate degree of changes as well as detect all kinds of change automatically from different time snapshot data. A case study for evaluation and practical business implications for this methodology are also provided.

Key Words : Data Mining, Association Rule Mining, Change Mining

1. Introduction

Understanding and adapting to changes of customer behavior is an important aspect of surviving in a continuously changing environment. Especially for businesses, knowing what is changing and how it has been changed is of crucial importance because it allows businesses to provide the right products and services to suit the changing market needs (Liu et al. 2000). Data mining is the process of exploration and analysis of large quantities of data in order to discover meaningful patterns and rules. But much of existing data mining research has been focused on devising techniques to build accurate models and to discover rules. Relatively little attention has been paid to mining changes in databases collected over time (Liu et al. 2000). In this paper, we develop a methodology which detects changes automatically from customer profiles and sales data at different periods of time. The most common approach to discover changes between two datasets is to generate rules from each dataset and directly compare the rules by rule matching. But this is not a simple process because of 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 emerging pattern, unexpected change and the added / perished rule. Then we develop similarity and difference measures for rule matching to detect all types of change. Finally, the degree of change is evaluated to detect significantly changed rules. The proposed methodology can evaluate degree of changes as well as detect all kinds of changes automatically from different time snapshot data. Detected changes can be usefully applied to plan various niche marketing campaigns. For example, if a manager can find out that a certain customer's preference has moved from a medium-size car to a large-size car, then that manager can establish a trade-in plan for customers who have a medium-size car and have the intention of buying a large-size car for replacement. Association rule mining finds interesting association relationships among a large set of data items (Agrawal et al. 1993). With massive amounts of data continuously being collected and stored, many industries are becoming interested in mining association rules from their databases. Association rule mining is used as a basic mining methodology in our research.

2. Background

2.1 Association Rule Mining

A typical association rule has an implication of the form $A \Rightarrow B$ where A is an itemset and B is an itemset that contains only a single atomic condition. The *support* of an association rule is the percentage of records containing itemsets A and B together. The *confidence* of a rule is the percentage of records containing itemset A that also contain itemset B . Support represents the usefulness of discovered rule and the confidence represents certainty of the detected association rule. Figure 1 shows two association rules of which support is the same but the confidence of Rule 2 is larger than that of Rule 1.

Record ID	Items Bought
2000	A, B, C
1000	A, C
4000	A, D
5000	B, E, F
Discovered Association Rule Rule 1 : A $\Rightarrow$ C (sup: 50 %, conf: 66.6 %) Rule 2 : C $\Rightarrow$ A (sup: 50 %, conf: 100 %)	

Figure 1. Dataset and discovered association rules

Association rule mining finds all collections of items in a database whose confidence and support meet or exceed pre-specified threshold value. Apriori algorithm is one of the prevalent techniques used to find association rules (Agrawal et al. 1993). Apriori operates in two phases. In the first phase, all large itemsets are generated. This phase utilizes the downward closure property of support. The second phase of the algorithm generates rules from the set of all large itemsets. Please refer to the study of Agrawal et al. (1993) for a more detail.

2.2 Data mining in a changing environment

There are existing works that have been done on learning and mining (Bay and Pazzani 1999; Ganti et al. 1999; Han and Kamber 2001; Liu et al. 2000) in a changing environment. All the following related works focus on dynamic aspects or comparison between two different datasets or rules. The first research trend related to our work is to discover Emerging Patterns (Agrawal and Psaila 1995; Dong and Li 1999; Li et al. 2000). Their research tries to find Emerging Patterns (EPs) which are defined as itemsets whose supports increase significantly from one dataset to another. EPs can capture emerging trends in timestamped 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 growth / decrease rate of consumption over time but cannot detect any unexpected changes such as a change from coffee $\Rightarrow$ tea to coffee $\Rightarrow$ milk. The second research field is mining class comparisons to discriminate between different classes (Bay and Pazzani 1999; Ganti et al. 1999; Han and Kamber 2001). Ganti et al. (1999) presents 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. It provides deviations measure between two mining model or focused regions but cannot be directly applied to detect customer behavior changes because it does not provide which aspects are 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) presents a technique for change mining by overlapping two decision trees which are generated from different time snapshots. But the change mining technique using decision trees cannot detect complete sets of change. Since decision tree techniques run within a specified objective class, only changes about that designated consequent attribute can be detected. This approach can be used only in cases which have a specific research question. Also, this technique does not provide any information for the type of change and the degree of change. Besides of the above studies, the researches for rule maintenance and time series also handle dynamic situations but we omit the explanations because of page limitation.

3. Problem Definition

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

D^t, D^{t+k} : datasets at time t, t+k

R^t, R^{t+k} : discovered association rulesets at time t, t+k

r_i^t, r_j^{t+k} : each rule from corresponding ruleset R^t, R^{t+k}

,where $i = 1,2,\cdots,|R^t|$, $j = 1,2,\cdots,|R^{t+k}|$

$Sup^t(r_i)$: support of r_i in time t dataset

Dong and Li (1999) introduced *Emerging Patterns* concept which captures significant changes and differences between datasets. Emerging patterns are defined as itemsets whose supports increase significantly from one dataset to another. We bring from 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 Pattern 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 two rules are significantly different

Example 1]

r_i^t : Income = High, Age = High $\Rightarrow$ Model = Large (Support = 0.1), r_j^{t+k} : Income = High, Age = High $\Rightarrow$ Model = Large (Support = 0.13). In this case, r_j^{t+k} is the emerging pattern with respect to r_i^t if we specify minimum growth rate to be 0.2. This is because the two rules have 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 and Hsu 1996; Liu et al. 1997; Padmanabhan and Tuzhilin 1999; Silberschatz and Tuzhilin 1996; Suzuki 1997). Liu and Hsu (1996) defined unexpected changes as rule similarity and difference aspects. They distinguished unexpected changes to unexpected condition changes and unexpected consequent changes based on a syntactic comparison between a rule and a belief. But we only adapt unexpected consequent changes because most unexpected condition changes usually make no sense. These unexpected consequent changes are the second type of change to detect which has a different rule structure over time. Therefore we redefine the term unexpected changes like the following from the study of Liu and Hsu (1996).

Definition 2] 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 : Income = High, Age = High $\Rightarrow$ Model = Large, r_j^{t+k} : Income = High, Age = High $\Rightarrow$ Model = Medium. In this case, r_j^{t+k} is 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 added rules and 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 be found only in the past but not the present. We define added and perished rule 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. Those terms are used to compare two rules in syntactic aspects and to judge degree of similarity and difference. But the terms "similar" and "quite different" are quite subjective and different from each individual. Therefore we define *Rule Matching Threshold (RMT)* which can be differently decided by individual user. Figure 2 explains the concept of RMT and provides how the different types of change can be distinguished by RMT.

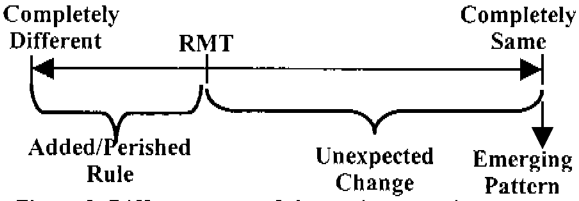


Figure 2. Different types of change in syntactic aspects

Finally, we define *the degree of change* as the measure of how much change has occurred. The degree of change has to be evaluated differently by each type of change because of different characteristics. The main way of evaluating 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 datasets which are collected from different periods and ranking the changed rules in each type by the degree of change.

4. Methodology

4.1 Overall Architecture

Now we suggest the methodology for the change detection problem. The methodology consists of the following three phases in Figure 3. In phase I, two association rulesets are generated from each dataset by using Apriori algorithm (Agrawal et al. 1993). In phase II, the changed ruleset is generated by using the rule matching method which compares two rules selected from each ruleset. We adapted the rule matching method developed by Liu and Hsu (1996) and modified it to distinguish between the above three types of change. For efficient rule matching, similarity and difference measures are developed. Our rule matching method can detect all types of changed rules including emerging patterns, unexpected changes, added and perished rules. In phase III, various changed rules detected in phase II are ranked according to the predefined degree of change which is a measure to evaluate how much change has occurred.

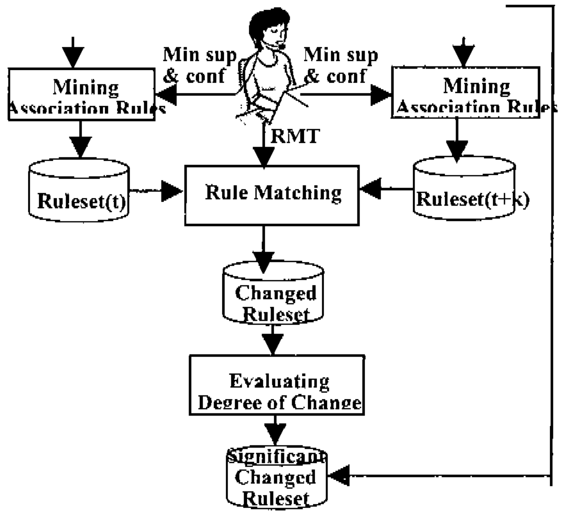


Figure 3. An architecture for detecting the change of customer behavior

4.2 Discovery of Association Rule

For the Apriori algorithm, we have to perform the discretization process (Hussain et al. 1999) to discover association rules. In this paper, all the values in the dataset are assumed to be discretized for the simplicity of explanation. We need two datasets collected at different times, minimum support levels and minimum confidence levels as inputs. In our experience, a lower minimum support level is preferred to discover association rules. If the minimum support level is set very high, we may lose the opportunity to detect the emerging patterns which have large growth (or decrease) rates but are rare items.

4.3 Discovery of changed rule

In this phase, various types of changed rules are detected using the rule matching method. The inputs of phase II are rulesets at time t and t+k, and the Rule Matching Threshold (RMT) which is specified by the user. Phase II is composed of the following three steps.

Step [1] Calculate the maximum similarity value for each rule in time t and t+k.
Step [2] For each rule r_i^t, calculate the difference measure between r_i^t, r_j^{t+k}.
Step [3] Classify the type of change for the rules using the maximum similarity value and the difference measure.

For the explanation of each step, some notations are briefly defined.

δ_{ij} : *Difference measure.*
Degree of difference between r_i^t and r_j^{t+k}
$(-1 \le \delta_{ij} \le 1,\ 0 \le |\delta_{ij}| \le 1)$

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

c_{ij} : *Degree of attribute match of the consequent parts*

$|A_{ij}|$: *Number of attributes common to both conditional parts of* 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}

y_{ij} : *Degree of value match of the consequent attribute*

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

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

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

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

In s_{ij}, $\ell_{ij} \times \sum_{k \in A_{ij}} x_{ijk} / |A_{ij}|$ represents a similarity of conditional part, and $c_{ij} \times y_{ij}$ represents a similarity of 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. To detect added and perished rules, the *maximum similarity value* is provided as follows.

$$s_i = \max(s_{i1}, \cdots, s_{i|R_{t+k}|}) \text{ ; } Maximum\,Similarity\,Value\,of\ r_i^t$$
$$s_i = \max(s_{1j}, \cdots, s_{|R_t|j}) \text{ ; } Maximum\,Similarity\,Value\,of\ r_j^{t+k}$$

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

Example 3] Assume the following rules are generated from each dataset D^t and D^{t+k}.

r_1^t : Income = High $\Longrightarrow$ Sales = High
r_2^t : Age=High,Preference=Price $\Longrightarrow$ Sales=High
r_1^{t+k} : Income = High $\Longrightarrow$ Sales = High
r_2^{t+k} : Age = High $\Longrightarrow$ Sales = High
r_3^{t+k} : Income=High,Preference=Price $\Longrightarrow$ Sales=Low

We can compute the similarity measure between r_2^t, r_2^{t+k} and the maximum similarity value of r_2^t as follows.

$$s_{22} = \frac{\frac{1}{2} \times 1 \times 1 \times 1}{1} = 0.5, \qquad s_2^t = \max(0,0.5,0) = 0.5$$

In the same manner, we can compute the maximum similarity value of each rule.

$$s_1^t = \max(1,0,0) = 1 \qquad s_2^t = \max(0,0.5,0) = 0.5$$
$$s_1^{t+k} = \max(1,0) = 1 \qquad s_2^{t+k} = \max(0,0.5) = 0.5$$
$$s_3^{t+k} = \max(0,0) = 0$$

If we specify RMT to be 0.4, then we can conclude that only r_3^{t+k} is an added rule.

As we can see from example 3, the maximum similarity value in step [1] is used to discover added rules or perished rules. The purpose of step [2] is to detect unexpected changes and emerging patterns. To detect unexpected change, a *difference measure* is provided as follows.

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

As defined above in the problem definition section, if conditional parts are similar but consequent parts are different, then this rule is called as an unexpected consequent. It means that the similarity of the conditional part is greater than that of the consequent part. Based on this measure, we can judge whether the rule r_j^{t+k} is an unexpected consequent change with respect to r_i^t. In summary, 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 two rules r_i^t and r_j^{t+k} are the same rules or completely different rules. Therefore additional measures such as ℓ_{ij}, y_{ij}, etc should be provided in case of $\delta_{ij} = 0$. If these values are 1 then we can directly find that two rules are same. We compute difference measures only in the case of $c_{ij} = 1$. If attributes of consequent parts between the two rules are different, it makes no sense to compare the degree of difference because these two rules are completely different rules. The step [3] classifies the rules as three types of change. To classify the type of change, additional computation is needed. 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_m^t which has the same structure with r_j^{t+k}. In this case, r_j^{t+k} should be classified into an emerging pattern and not to be classified 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.

$$\delta'_{ij} = |\delta_{ij}| - k_{ij}, \qquad where\ k_{ij} = \begin{cases} 1 & ,if \max(s_i, s_j) = 1 \\ 0 & ,otherwise \end{cases}$$

The fact that s_i (or s_j) is equal to 1 means that the same rule exists in another ruleset. That means r_j^{t+k} is likely to be classified into an emerging pattern. If $\delta_{ij}^{'}$ 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.

Example 4]

r_1^t : Income=High,Preference=Price $\Rightarrow$ Sales=Low
r_2^t : Age = High, Preference = Price $\Rightarrow$ Sales = High
r_1^{t+k} : Income = High $\Rightarrow$ Sales = High
r_2^{t+k} : Age = High $\Rightarrow$ Sales = High
r_3^{t+k} : Income=High,Preference=Price $\Rightarrow$ Sales=Low

With the association ruleset, we can compute the difference and modified difference measure between r_2^t and r_3^{t+k} as follows.

$$\delta_{23} = 0.5 \qquad \delta_{23}^{'} = 0.5 - 1 = -0.5$$

If we specify that RMT is equal to 0.4, we cannot conclude that r_3^{t+k} is an unexpected consequent change with respect to r_2^t because r_3^{t+k} has a same rule structure with r_1^t. Therefore, we can conclude that r_3^{t+k} is an emerging pattern of r_1^t. And r_3^{t+k} is not thought to be an unexpected consequent change with respect to r_1^t.

Table 1 summarizes the value of each measure for each type of change.

Table 1. Value of measure for each type of change

Type of Change	Value of measure to classify
Emerging Pattern	$\delta_{ij} = 0, \quad \sum_{k \in A_{ij}} x_{ijk} > 0$
Unexpected Consequent	$\delta_{ij} > 0, \quad \delta_{ij}^{'} \geq RMT$
Added Rule (Perished Rule)	$s_j < RMT \ (s_i < RMT)$

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's consider the unexpected change. The following example presents why additional measures should be required.

r_i^t : Income = High, Age = High $\Rightarrow$ Model = Large

r_j^{t+k} : Preference= Price, Age= High $\Rightarrow$ Model = Small

If RMT value 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 exist two problems to conclude whether this change is significant. First, we cannot capture this change easily because conditional parts are not same. Second, although we can understand this change, we 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 (Hussain et al. 2000; Suzuki 1997) as follows.

Definition 5] Unexpectedness

If an association rule $A \Rightarrow B$ is unexpected with respect to the belief $X \Rightarrow Y$, then the following must hold.

(1) $B \ And \ Y = False$

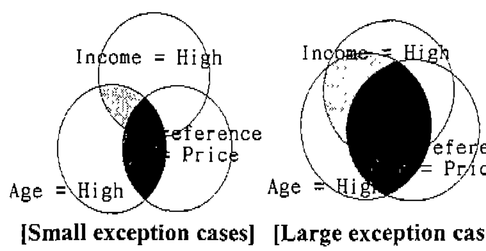
(2) The rule $X, A \Rightarrow B$ holds.

A new measure for the degree of change of 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 above (1) condition of definition 5 because of definition 2. Furthermore the support value of the conjunction rule should be evaluated to check whether (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 $\Rightarrow$ Model = Small

If the above conjunction rule, $r_{i \cap j}$, is statistically large (i.e. has 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 to 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. Additional conditions which should be included is that the support value of $r_{i \cap j}$ should be enough large 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 $r_{i \cap j}$. Figure 4 illustrates the above situation. In Figure 4, as more exceptional cases occur for certain existing beliefs or rules, we consider that more unexpected consequent changes have occurred. Now we are ready to provide the following measure for the degree of unexpected consequent change.

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



[Small exception cases] [Large exception cases]

Customers who bought small sized car at period t+k (New Discovered Rule)

Customers who have tendency to buy large sized car (Existing Belief)

Exception Cases : customers who expected to buy large sized car but bought small sized car

Figure 4. Concept for degree of change of unexpected change

In the case of emerging pattern, it is more simple to evaluate the significance level than the case of unexpected change. The growth or decrease rate are used as the measure for this type of change. To evaluate the degree of change for added and perished rule cases, the support value of those rules and the maximum similarity value are used. As mentioned before, the maximum similarity value of the rule represents the degree of similarity of the most similar rule to the other ruleset. If there is a situation that the support values of two added rules are 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.

$$\alpha_{ij} = \begin{cases} \dfrac{Sup^{t+k}(r_i) - Sup^{t}(r_i)}{Sup^{t}(r_i)} & ,\textit{emerging pattern case} \\[2ex] \dfrac{Sup^{t+k}(r_{i\cap j})}{Sup^{t+k}(r_j)} & ,\textit{unexpected change case} \\[2ex] (1 - s_i) \times Sup^{t}(r_i) & ,\textit{perished rule case} \\[1ex] (1 - s_j) \times Sup^{t+k}(r_j) & ,\textit{added rule case} \end{cases}$$

5. Evaluation

For the evaluation of proposed methodology, the system is implemented using Visual Basic 6.0. The case study has been conducted to evaluate how well the system performs its intended task of detecting significant changes. The dataset is prepared from an Korean online shopping mall which sells various consumer goods. The dataset contains customer profiles and purchasing history such as age, job, sex, address, registration year, cyber money, number of purchases, total purchase amount, number of visits, payment method during one year. We constructed a data warehouse which aggregated historical data by individual customer. We prepared two dataset to detect significant changes of purchasing behavior by their customers. The first dataset contains profiles and purchasing history information of certain customers who had bought more than one cosmetics from Feb/1/2000 to Jun/30/2000. The second dataset contains the same information but includes customers who had made one additional purchase of cosmetics from Jul/1/2000 to Jan/5/2001. After preprocessing the data for cleansing and discretization, an Apriori technique was applied to discover the association rules from each dataset. We selected the number of purchases and total sales amount as output variables. In the condition of 1 % minimum support, 80 % minimum confidence and maximum itemset of size 3, the system found 127 association rules for the first dataset and 104 association rules for the second one. Given a 0.4 Rule Matching Threshold (RMT), the system found 101 changed rules and 24 significantly changed rules. The number of changed rules for each type of change is provided in Table 2.

Significant emerging patterns, unexpected changes, added/perished rules are summarized in Table 3, 4, and 5. From changed rules 4) and 5) in Table 3, we can see the rapid growth (90 % growth) in sales for customers who are specialists and visit the mall frequently. Although the support value for them is low (0.021, 0.04), those customers have the possibility to become loyal customers in the near future because of high growth rate. Therefore, a marketing campaign to invoke the revisiting by those customers should be developed. We can also identify above trends in changed rule 1) of Table 3. From the changed rule 6), 7) and 8) in Table 3, we can see the rapid growth in sales for customers who live in Daegu, Pusan city and visit the mall frequently. Without the change detection methodology, the marketing manager may understand that customers who live in Daegu, Pusan city and visit the mall frequently are not important because of the low support value.

Table 2. Number of changed rules for each type of change

Type of change	Number of Changed rules	Number of significant changed rules
Emerging Patterns	92	17 (Degree of change > 0.4)
Unexpected Changes	6	4 (Degree of change > 0.3)
Added Perished Rules	3	3 (Degree of change > .01)

Table 3. Significant Emerging Patterns (DoC > 0.4)

Emerging Pattrens	Rule Support (T/T+K)		DoC
1) Visit=Low, Job=Specialist ➔ OrdCnt=Low	0.037	0.078	1.11
2) Visit=Low,ReservedMoney=Low➔OrdCnt=Low	0.177	0.368	1.08
3) Visit=Low, ReservedMoney=Low ➔ Sales=Low	0.177	0.368	1.08
4) Visit=High, Job=Specialist ➔ OrdCnt=High	0.021	0.04	0.90
5) Visit=High, Job=Specialist ➔ Sales=High	0.021	0.04	0.90
6) Visit=High, Addr=Daegu ➔ OrdCnt=High	0.01	0.017	0.70
7) Visit-High, Addr=Pusan ➔ OrdCnt=High	0.015	0.025	0.67
8) Visit=High, Addr=Pusan ➔ Sales=High	0.015	0.025	0.67
9) ReservedMoney=Low,Job=Student➔Sales=Low	0.011	0.018	0.64
....	...	...	...
17) Visit=Low, Addr=ChungBuk ➔ OrdCnt=Low	0.01	0.014	0.40

With regard to unexpected changes, we identified 4 significant changes. From the changed rule 1) of Table 4, we can find that sales for female customers who live in KyungNam are low from the first dataset. But in the second dataset, we can see that sales for female customers who visit the mall frequently are high even if they are female customers who live in KyungNam. It means that the importance of customers who live in KyungNam and visit the mall frequently is gradually increasing. Therefore, a modification for the existing marketing strategy and plan is required. Changed rules 2), 3) and 4) in Table 4 can be interpreted similarly. Finally, we three perished rules are found in Table 5. From Feb. to Jun. in 2000, most of their customers were aged twentith and sales for the other customers were very low. But nowadays, we can find a trend that the age of their customers covers a wider range. Therefore additional services and products for elders and teenagers should be also developed.

Table 4. Significant Unexpected Changes (DoC > 0.3)

Time T	Time T+K	δ_{ij}	δ'_{ij}	α_{ij}
1) Sex=F, Addr=KyungNam → OrdCnt=Low	Visit=High, Addr=KyungNam → OrdCnt=High	0.5	0.5	0.85
2) Registday=This_year, Addr=KyungNam → OrdCnt=Low	Visit=High, Addr=KyungNam → OrdCnt=High	0.5	0.5	0.79
3) Payment=Cash, Addr=KyungNam → OrdCnt=Low	Visit=High, Addr=KyungNam → OrdCnt=High	0.5	0.5	0.58
4) ReservedMoney=Low, Addr=KyungNam → OrdCnt=Low	Visit=High, Addr=KyungNam → OrdCnt=High	0.5	0.5	0.31

Table 5. Significant Added/Perished Rules (DoC > 0.01)

Time T	MSV	Sup	α_{ij}
1) Age=Teen → Sales=Low	0	0.018	0.018
2) Sex=F, Age=Teen → Sales=Low	0	0.015	0.015
3) Age=Thirtith, Addr=Pusan → Sales=Low	0	0.012	0.012

6. Conclusion

In this paper, we developed a methodology which detects changes of customer behavior automatically from customer profiles and sales data at different time snapshots. The practical applications and opportunities to use for the methodology are as follows. First, in macro aspects, business mangers can follow the changing trends using change detection methodology. They need to analyze their customer's changing behaviors in order to provide products and services that suit the changing needs of the customers. Second, in micro aspects, it can be possible for a business manager to understand customer needs more deeply and design additional niche marketing campaigns using this methodology. Knowing the history of customer behavior can give a better understanding of customer behavior. Some limitations of suggested methodology can be described as follows. With regard to the number of target datasets which should be compared, the methodology is suggested to only two datasets. If there are three or more datasets to be compared over time, then another methodology will have to be developed. The methodology is run on the datasets which have descretized values. If there is a dataset which has continuous values, then a pre-processing step for discretization is needed. Various techniques for discretization are summarized in the study of Hussain et al. (1999). The rules for the methodology is come from association rule mining. We do not consider rules generated from another rule induction method such as a decision tree. But these assumptions are easily loosened if we prepare functions for processing continuous variables. Finally, more sophisticated evaluation is needed. But we have plan to evaluate proposed measures and test significance of discovered change rule using statistical method. As a further research area, we plan to extend our methodology to discover changes of a more general nature than association rules. It will be also promising to setup the campaign management planning based on our suggested methodology. And it will be also interesting to check the effectiveness of the campaign. We have also plan to apply our methodology to detect the changes of individual level. We believe that the change detection problem will become more and more important as more data mining applications are implemented.

References

1. Agrawal, R., Imielinski, T., and Swami, A. (1993). "Mining association rules between sets of items in large databases," *Proceedings of the ACM SIGMOD Conference on Management of Data.*
2. Agrawal, R., and Psaila, G. (1995). "Active data mining," *Proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD-95).*
3. Bay, S. D., and Pazzani, M. J. (1999). "Detecting Change in Categorical Data: Mining Contrast Sets," *Proceedings of the Fifth International Conference on Knowledge Discovery and Data Mining (KDD-99).*
4. Dong, G., and Li, J. (1999). "Efficient Mining of Emerging Patterns : Discovering Trends and Differences," *Proceedings of the Fifth International Conference on Knowledge Discovery and Data Mining (KDD-99).*
5. Ganti, V., Gehrke, J., and Ramakrishnan, R. (1999). "A framework for measuring changes in data characteristics," *Proceedings of the Eighteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (PODS-99).*
6. Han, J., and Kamber, M. (2001). *Data Mining : Concepts and Techniques*, Morgan Kaufmann Publishers, San Francisco, 200-207.
7. Hussain, F., Liu, H., Suzuki, E., and Lu, H. (2000). "Exception Rule Mining with a Relative Interestingness Measure," *Proceedings of Pacific Asia Conference on Knowledge Discovery in Databases (PAKDD)*, 86-97.
8. Hussain, F., Liu, H., Tan, C. L., and Dash, M. (1999). "Discretization: An Enabling Technique," *The National University of Singapore Technical Report TRC6/99*, June.
9. Lanquillon, C. (1999). "Information filtering in changing domains," *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI99)*, 41-48.
10. Li, J., Dong, G., and Ramamohanarao, K. (2000). "Making Use of the Most Expressive Jumping Emerging Patterns for Classification," *Proceedings of Pacific Asia Conference on Knowledge Discovery in Databases (PAKDD).*
11. Liu, B., and Hsu, W. (1996). "Post-Analysis of Learned Rules," *Proceedings of the Thirteenth National Conference on Artificial Intelligence (AAAI-96).*
12. Liu, B., Hsu, W., and Chen, S. (1997). "Using general impressions to analyze discovered classification rules," *Proceedings of the Third International*

Conference on Knowledge Discovery and Data Mining (KDD-97), 1997.

13. Liu, B., Hsu, W., Han, H. S., and Xia, Y. (2000). "Mining Changes for Real-Life Applications," in *Publishing in the Second International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2000),* Y. Kambayashi, M.K. Mohania, A.M. Tjoa (eds.).
14. Nakhaeizadeh, G., Taylor, C., and Lanquillon, C. (1998). "Evaluating usefulness of dynamic classification," *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98).*
15. Padmanabhan, B., and Tuzhilin, A. (1999). "Unexpectedness as a measure of interestingness in knowledge discovery," *Decision Support Systems* (27), 303-318.
16. Silberschatz, A., and Tuzhilin, A. (1996). "What makes patterns interesting in knowledge discovery systems," *IEEE Trans. On Knowledge and Data Engineering* 8(6), 970-974.
17. Suzuki, E. (1997). "Autonomous discovery of reliable exception rules," *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining (KDD-97).*