PERGAMON

Expert Systems with Applications 21 (2001) 157–168

Expert Systems with Applications

www.elsevier.com/locate/eswa

Mining the change of customer behavior in an internet shopping mall

Hee Seok Song[a,*], Jae kyeong Kim[b], Soung Hie Kim[a]

[a]Graduate School of Management, Korea Advanced Institute of Science and Technology, 207-43, Cheongryangri, Dongdaemun, Seoul, 130-012, South Korea

[b]School of Business Administration, KyungHee University, no. 1, Hoeki-Dong, Dongdaemoon, Seoul, 130-701, South Korea

Abstract

Understanding and adapting to changes of customer behavior is an important aspect for a internet-based company to survive in a 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 the 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 the degree of changes as well as detect all kinds of change automatically from different time snapshot data. A case study on an internet shopping mall for evaluation of this methodology is also provided. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: 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 internet-based companies, 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, Hsu, Han & Xia, 2000). More specifically, most decision makers in internet-based companies have a strong need to know and adapt to the answers to following questions about their customers.

- Which customer group's sales are gradually increasing?
- Which customer groups moved from product A to product B over the years?
- Whether certain groups of customers had gradually emerged to be the major buyers over the years?

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 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). Most data mining techniques such as association rules, decision trees and neural networks cannot be applied alone to answer the above research questions, because they cannot handle dynamic situations well. Also, most data mining techniques usually ignore rare items which have a small frequency of occurrence, but rare items which have a large growth rate or decreasing rate may give very significant implications to managers in changing environment. These are the reasons why we are motivated to develop other methodology to detect change. Association rule mining finds interesting association relationships among a large set of data items (Agrawal, Imielinski & Swami, 1993a,b; Agrawal & Srikant, 1994). 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.

Detected changes can be usefully applied to plan various niche marketing campaigns. For example, in a shop, people used to buy beer and snacks together—now they still buy beer, but seldom buy snacks. The shop manager needs to know this information so that he/she can find out the reason for this and design some catalysts to attract customers to buy snacks again. As an another 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. In this paper, we develop a methodology

* Corresponding author. Tel.: +82-2-958-3670; fax: +82-2-958-3604.
E-mail address: hssong@kgsm.kaist.ac.kr (H.S. Song).

PII: S0957-4174(01)00037-9

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. Furthermore, the methodology can also be applicable for discovery of different characteristics from different categorical data.

We begin by reviewing the concept of association rules which are a prerequisite of our research and the discussion of related works in Section 2. We define types of change and change detection problems to clarify our objectives in Section 3. In Section 4, we provide our methodology. A case study for evaluation and its business implications are presented in Section 5 and Section 6. Finally we summarize our contributions and outline areas for further research in the Conclusion section.

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. The support represents the usefulness of the discovered rule and the confidence represents certainty of the detected association rule. Fig. 1 shows two association rules of which the support is the same but the confidence of Rule 2 is larger than that of Rule 1.

Association rule mining finds all collections of items in a database whose confidence and support meet or exceed a prespecified threshold value. Apriori algorithm is one of the prevalent techniques used to find association rules (Agrawal et al., 1993a,b; Agrawal & Srikant, 1994; Agrawal, Srikant & Vu, 1997). Apriori operates in two phases. In the first phase, all large itemsets are generated. This phase utilizes the downward closure property of support. In other words, if k size (or length k) itemset is a large itemset, then all the itemsets below $(k-1)$ size must also be large itemsets.

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

Fig. 1. Dataset and discovered association rules.

Using this property, candidate itemsets of length k are generated from the set of large itemsets of length $(k-1)$ by imposing the constraint that all subsets of length $(k-1)$ of any candidate itemset must be present in the set of large itemsets of length $(k-1)$. The second phase of the algorithm generates rules from the set of all large itemsets. Please refer to the study of Agrawal et al. (1993a,b) for a more detail.

In this paper we use profile association rules to explain examples. A profile association rule is one type of association rule in which the left hand side of the rule consists of customer profile information, such as age, salary, years of education, and social status. The right hand side of the rules consists of customer behavior information, such as buying milk, diaper, beer, etc (Aggarwal, Sun & Yu, 1998). Since profile association rules give various chances to establish a target marketing strategy, we use the profile association rule in each example and case study, but the proposed methodology is not limited to special type of association rules.

2.2. Data mining in a changing environment

There are existing works that have been carried out on learning (Freund & Mansour, 1997; Helmbold & Long, 1994; Widmer, 1996) and mining (Bay & Pazzani, 1999; Ganti, Gehrke & Ramakrishnan, 1999; Han & Kamber, 2001; Liu et al., 2000; Nakhaeizadeh, Taylor & Lanquillon, 1998) in a changing environment. All the following related works focus on dynamic aspects or comparison between two different datasets 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, Han, Ng & Wong, 1996a; Cheung, Ng & Tam, 1996b; Feldman, Aumann, Amir & Manila, 1997; Thomas, Bodagala, Alsabti & Ranka, 1997). The purpose of these studies is improving accuracy in a changing environment. For example, in the study of Cheung et al. (1996a,b), 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 related to our work is to discover Emerging Patterns (Agrawal & Psaila, 1995; Dong & Li, 1999; Li, Dong & Ramamohanarao, 2000). This 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 → tea to coffee → milk. Another related research is subjective interestingness in data mining (Liu & Hsu, 1996; Liu et al., 1997; Liu, Hsu, Ma & Chen, 1999; Padmanabhan & Tuzhilin, 1999; Silberschatz & Tuzhilin, 1996; Suzuki, 1997). These researches 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 rule to find degrees of difference, and 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 an increasing interest to discover regularity from time series data (Das, Gunopulous & Mannila, 1997; Das, Lin, Mannila, Renganathan & Smyth, 1998; Han, Dong & Yin, 1999). Das et al. (1998) considers 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, and Han et al. (1999) presents several algorithms for efficient mining of partial periodic patterns, by exploring some interesting properties related to partial periodicity. Das et al. (1997) also presents 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 from 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) presents a general framework for measuring changes in two models. Essentially, the difference between the 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 itemsets, decision tree classifiers, and clusters, and captures standard measures of deviation such as misclassification rate and the chi-squared metric as special cases. It provides deviation measures between the 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 which kind of changes have occurred. Bay and Pazzani (1999); 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 degree of change.

3. Problem

In this section, we examine all the possible types of change based on past research and business requirements (Dong & Li, 1999; Lanquillon, 1999; Liu & Hsu, 1996; Liu et al., 1997; Padmanabhan & Tuzhilin, 1999; Suzuki, 1997). Following that, each type of change and change detection problem is defined. Let us 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 $i = 1,2,\ldots,|R^t|$, $j = 1,2,\ldots,|R^{t+k}|$
$\mathrm{Sup}^t(r_i)$: support of r_i at time t dataset.

Dong and Li (1999) introduced the *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. More specifically, emerging patterns are itemsets whose growth rates are larger than a given threshold value. When applied to time stamped databases, 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 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 → Model = Large (Support = 0.1) r_j^{t+k}: Income = High, Age = High → 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 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 on discovering interesting patterns (Liu & Hsu, 1996; Liu et al., 1997, 1999; Padmanabhan & Tuzhilin, 1999; Silberschatz & Tuzhilin, 1996; Suzuki, 1997). Unexpected changes can be found from newly discovered association rules which are different from user's existing beliefs. Liu and Hsu (1996) defined 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 either the consequents of the rule and the belief are 'similar', but the conditions are 'far apart' or 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. In other words, they distinguished unexpected changes to unexpected condition changes and unexpected consequent changes, but we only adapt unexpected consequent changes because most unexpected condition changes are usually unrelated. These unexpected consequent changes are the second type of change to detect which has a different rule structure overtime. Therefore, we redefine the term 'unexpected changes' like the following from the study of 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: 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 the 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. Added and perished rules are different from emerging patterns and unexpected changes because these rules have completely different structures when compared to any rule of any other ruleset. Therefore, additionally, we define the 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 the 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 the individual user. Fig. 2 explains the concept of RMT and provides how the different types of change can be distinguished by RMT.

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 the degree of change will be explained in the next section. 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 at a different time and ranking the changed rules in each type by the degree of change.

4. Methodology

4.1. Overall architecture

The change detection problem is defined in a

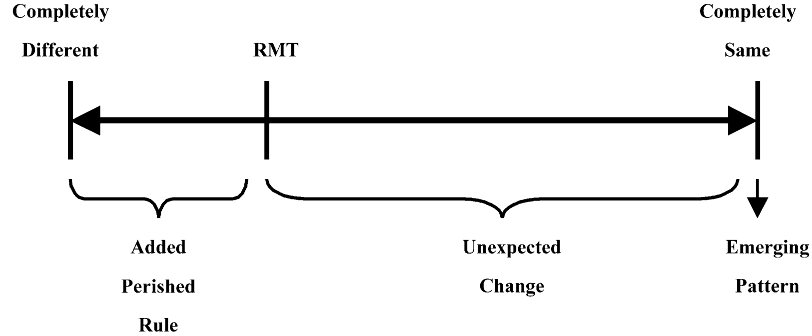

Fig. 2. Different types of change in syntactic aspects.

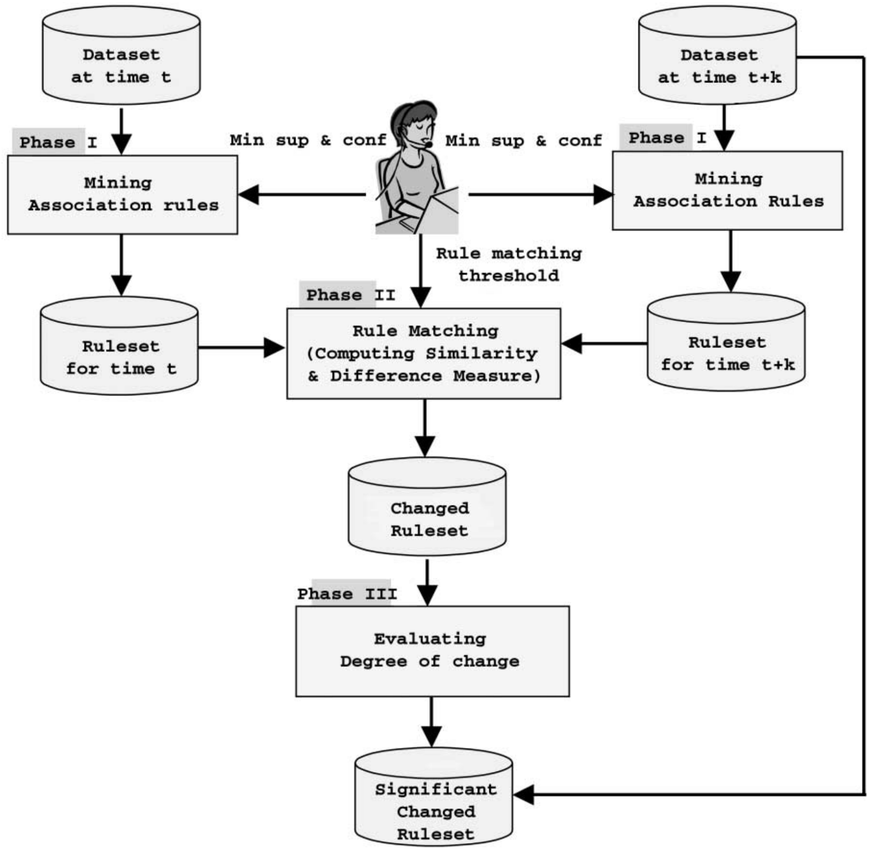

Fig. 3. An architecture for detecting the change of customer behavior.

previous section. We now suggest the methodology for the change detection problem. The methodology consists of the following architecture that has three phases in Fig. 3.

In phase I, two association rulesets are generated from each dataset by using Apriori algorithm (Agrawal et al., 1993a,b; Agrawal & Srikant, 1994). 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. Liu and Hsu (1996) provide a technique for finding unexpected rules with respect to the user's existing knowledge, but this technique cannot be used for detecting changes, as it only compares each newly generated rule with each existing belief, it does not find which aspects have changed and which kinds of change have occurred. Therefore, their measures are modified to classify 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. In the case of emerging patterns, the growth or decrease rate of the *support* value of the changed rule is computed and then they are sorted by the absolute value of the rate. The growth or decrease rate 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). They define unexpectedness using the exception rule concept (Hussain, Liu, Suzuki & Lu, 2000; Suzuki, 1997). For the understanding of the exception rule concept, we explain it by the following example: there is an existing belief such as 'birds can fly'. Given this belief, a new rule 'penguin (which is a kind of bird) cannot fly' is discovered. In this case, the new rule becomes an exception rule with respect to the existing belief. If the exception rule becomes an unexpected rule (or unexpected change in our paper), the penguin has to be a statistically large subset in the dataset. We apply this concept to evaluate the degree of change for unexpected change. If more exception cases occur for certain existing beliefs or rules, we consider that a more unexpected consequent change has occurred. The degree of exception for certain existing rules plays the role of degree of change in 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 the later section, and more detailed explanations for the measures and procedures are provided in the continuing subsections step by step.

4.2. Discovery of association rule

For the Apriori algorithm, we have to perform the discretization process (Hussain, Liu, Tan & Dash, 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 input of phase II is discovered 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 measures 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 measures.

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

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

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

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

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

In

$$s_{ij}, \frac{\ell_{ij} \times \sum_{k \in A_{ij}} x_{ijk}}{|A_{ij}|}$$

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:

$s_i = \max(s_{i1}, s_{i2}, \cdots, s_{i|R_{t+k}|})$; maximum similarity value of r_i^t

$s_j = \max(s_{1j}, s_{2j}, \cdots, s_{|R_t|j})$; maximum similarity value of r_j^{t+k}

The maximum similarity value indicates whether the rule is added or perished. If $s_i <$ 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 → Sales = High
r_2^t: Age = High, Preference = Price → Sales = High
r_1^{t+k}: Income = High → Sales = High
r_2^{t+k}: Age = High → Sales = High
r_3^{t+k}: Income = High, Preference = Design → 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:

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

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

$$s_1^t = \max(1, 0, 0) = 1\ s_2^t = \max(0, 0.5, 0) = 0.5$$

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

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

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

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. Therefore, 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. 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 or unexpected condition 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 $c_{ij} = 1$ between two rules r_i^t and r_j^{t+k}. 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.

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:

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

The fact that s_i (or s_j) is close 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 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.

Example 4.

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

$$\delta_{23} = 0.5\ \delta'_{23} = 0.5 - 1 = -0.5$$

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 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.

4.4. Evaluating degree of change

All the changed rules have to be ranked by the degree of

Table 1
Value of measure for each type of change

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

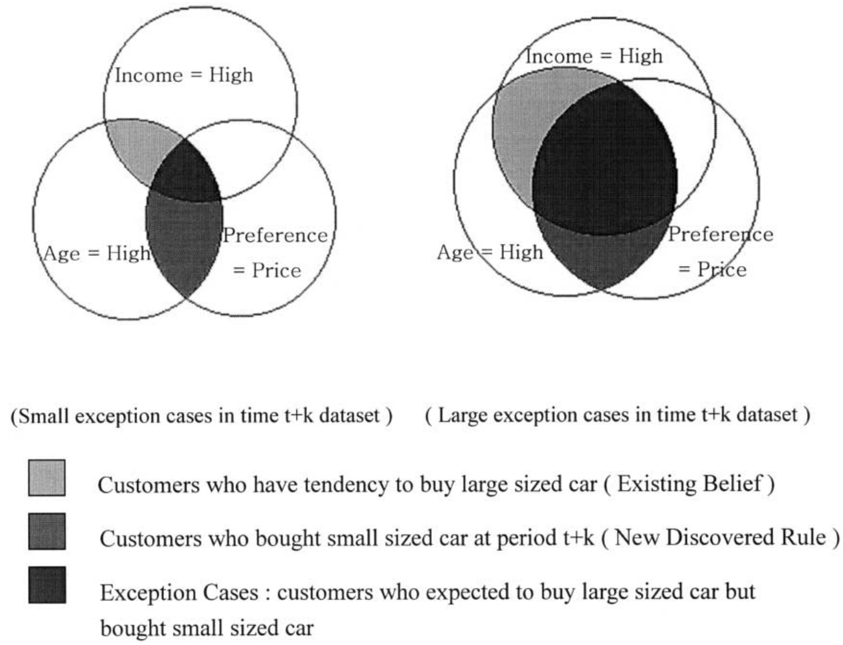

Fig. 4. Concept for significant test of unexpected consequent.

change. We will explain the idea of evaluating the degree of change for each type of change. First, let us consider the unexpected change. The following example presents why additional measures should be required:

r_i^t: Income = High, Age = High → Model = Large
r_j^{t+k}: Preference = Price, Age = High → 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 two problems exist 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 → 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 $r_{i\cap j}$. Fig. 4 illustrates the above situation.

In Fig. 4, as more exceptional cases occur for certain existing beliefs or rules, we consider that more unexpected consequent changes have occurred. We are now ready to provide the following measure for the degree of unexpected consequent change.

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

In the case of the emerging pattern, it is simpler to evaluate the significance level than the case of unexpected change. The growth or decrease rate is used as the measure

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 > 0.01)

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 is 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 the same, we naturally place more importance on the rule which has the 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.

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

5. Evaluation

For the evaluation of the 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 a Korean online shopping mall which sells various consumer goods. The dataset contains customer profiles and purchasing information such as age, job, sex, address, registration year, cyber money, number of purchases, total purchase amount, number of visits and the payment method during 1 year. We constructed a data warehouse which aggregated historical data by individual customer. We prepared two datasets to detect significant changes of purchasing behavior by their customers. The first dataset contains profiles and purchasing information of certain customers who had bought more than one cosmetic from 1st February 2000 to 30th June 2000. The second dataset contains the same information but includes customers who had made one additional purchase of cosmetics from 1st July 2000 to 5th January 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 Tables 3–5.

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 rules (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. With regard to unexpected changes, we identified four 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, three perished rules are found in Table 5. From February to June in 2000, most of their customers were in their twenties and sales for the other customers were very low. Nowadays, however, 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.

Our suggested methodology focuses on finding the changes between different time snapshot datasets, but the methodology can be used in another applications, too. At the same time, it has some limitations in its use. From the perspective of application, the methodology can be applied for comparison of categorical data as well as dynamically

Table 3
Significant Emerging Patterns (Degree of change > 0.4)

r_i^t (Or r_j^{t+k})	Rule Support		α_{ij} (>0.4)
	Sup^t (r_i)	Sup^{t+k} (r_j)	
(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
(10) Visit = High, Job = ETC → Sales = High	0.063	0.097	0.54
(11) Visit = High, Job = ETC → OrdCnt = High	0.063	0.097	0.54
(12) Visit = High, ReservedMoney = Low → OrdCnt = High	0.1	0.146	0.46
(13) Visit = Low, Registday = Last_year → OrdCnt = Low	0.141	0.079	−0.44
(14) Visit = High, Registday = Last_year → OrdCnt = High	0.102	0.059	−0.42
(15) Visit = High, Regist_day = Last_year → Sales = High	0.102	0.059	−0.42
(16) ReservedMoney = High, Visit = Low → OrdCnt = Low	0.492	0.284	−0.42
17) Visit = Low, Addr = ChungBuk → OrdCnt = Low	0.01	0.014	0.40

Table 4
Significant Unexpected Changes (Degree of change > 0.3)

r_i^t	r_j^{t+k}	δ_{ij}	δ_{ij}'	α_{ij}
(1) Sex = F, Addr = KyungNam → OrdCnt = Low(Support: 0.034)	Visit = High, Addr = KyungNam → OrdCnt = High(Support: 0.015)	0.5	0.5	0.85
(2) Registday = This_year, Addr = KyungNam → OrdCnt = Low(Support: 0.032)	Visit = High, Addr = KyungNam → OrdCnt = High(Support: 0.015)	0.5	0.5	0.79
(3) Payment = Cash, Addr = KyungNam → OrdCnt = Low(Support: 0.021)	Visit = High, Addr = KyungNam → OrdCnt = High(Support: 0.015)	0.5	0.5	0.58
(4) ReservedMoney = Low, Addr = KyungNam → OrdCnt = Low(Support: 0.012)	Visit = High, Addr = KyungNam → OrdCnt = High(Support: 0.015)	0.5	0.5	0.31

Table 5
Significant added/perished rules (degree of change > 0.01)

r_i^t	MSV	Support	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

changed data. The methodology is run on the datasets which have discretized 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). Also, the rules for the methodology came 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, in the perspective of rule structure, it is assumed that every rule has a single attribute consequent, which is very common in data mining research.

6. Business implications

In this section, we summarize the opportunities of using this methodology and provide various applications in practical business perspectives. 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 (Liu et al., 2000). For example, if a manager finds the trend that the age of customer for a certain product is decreasing, then he can develop additional services and product specifications for youths. On the other hand, undesirable changes can be properly controlled. If the manager knows the decreasing trend of co-purchase of two certain products, then he/she can examine the reason and establish a reaction plan to prevent that trend. 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. For example, although the satisfaction level of two different customer groups are both 'poor' for a certain product, we can find great differences between two groups if we know that satisfaction of customer group A is 'great' and the other group B is 'poor' in the past year. In this case, the manager can establish various niche marketing campaigns based on the examination of reasons for group A's decline. To give another example, if a manger finds a certain customer group which is very small but has a large growth rate, then he can design a direct mailing campaign for customer groups who have similar properties. Through this campaign, he can invoke the purchasing of their products. Furthermore, if a manager can find out that a certain customer's preference has moved from medium- to large-size cars, then that manager can establish a trade-in plan. 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 most promising application for change detection is analyzing the effectiveness of a marketing campaign. Using this methodology, if the manager generates rules from sales dataset before and after a campaign, he/she can evaluate the effectiveness of his/her marketing campaign by seeing whether it operated correctly with respect to the original intention. In manufacturing, change detection can be also applied to monitor changes and control the quality factor. Changes of various measures of product quality can be easily detected by this methodology and then undesirable changes can be properly controlled. In some cases, additional considerations may be involved to meet real time monitoring and control needs. Even in a static environment, our change detection methodology can be used. The most common area in a static environment is comparison between the two categorical datasets. Also, it can be used as a basis of comparison of two mining models to evaluate accuracy and effectiveness.

7. 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. For this purpose, we defined types of change as emerging pattern, unexpected change and the added/perished rule. We then developed similarity and difference measures for rule matching to detect all types of change in syntactic aspects. Additionally, the degree of change is evaluated to detect significantly changed rules in semantic aspects. We also suggested practical applications and opportunities to use for our methodology. 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 set up the campaign management planning based on our suggested methodology and it will be also interesting to check the effectiveness of the campaign. The contribution of this research is that the proposed methodology can evaluate the degree of changes as well as detecting all kinds of changes automatically from different time snapshot data. For such a purpose, we developed new measures for changes. We believe that the change detection problem will become more and more important as more data mining applications are implemented.

References

Aggarwal, C. C., Sun, Z., & Yu, P. S. (1998). Online algorithms for finding profile association rules. *Proceedings of the 1998 ACM CIKM International Conference on Information and Knowledge Management*, (pp. 86–95).

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).

Agrawal, R., Imielinski, T., & Swami, A. (1993b). Database mining: a performance perspective. *IEEE Trans. Knowledge and Data Engineering*, *5* (6), 914–925.

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

Agrawal, R., & Srikant, R. (1994). Fast algorithm for mining association rules. *Proceedings of the International Conference on Very Large Databases (VLDB-94)*, (pp. 487–499).

Agrawal, R., Srikant, R., & Vu, Q. (1997). Mining association rules with item constraints. *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining (KDD-97)*, (pp. 67–73).

Bay, S. D., & 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)*, (pp. 302–306).

Cheung, D. W., Han, J., Ng, V. T., & Wong, C. Y. (1996a). Maintenance of discovered association rules in large databases: an incremental updating technique. *Proceedings of the Twelth International Conference on Data Engineering (ICDE-96)*, 106–114.

Cheung, D. W., Ng, V. T., & Tam, B. W. (1996b). Maintenance of discovered knowledge: a case in multi-level association rules. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 307–310.

Das, G., Gunopulous, D., & Mannila, H. (1997). Finding similar time series. *Proceedings of the First European Symposium: Principles of Data Mining and Knowledge Discovery (PKDD '97)*, (pp. 88–100).

Das, G., Lin, K., Mannila, H., Renganathan, G., & Smyth, P. (1998). Rule discovery from time series. *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98)*, (pp. 16–22).

Dong, G., & 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)*, (pp. 43–52).

Feldman, R., Aumann, Y., Amir, A., & Manila, P. (1997). Efficient algorithms for discovering frequent sets in incremental databases. *SIGMOD'97 Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD'97)*, (pp. 59–66).

Freund, Y., & Mansour, Y. (1997). Learning under persistent drift. *Computational learning theory: Third European conference*, (pp. 109–118).

Ganti, V., Gehrke, J., & 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)*, (pp. 126–137).

Han, J., Dong, G., & Yin, Y. (1999). Efficient mining of partial periodic patterns in time series database. *Proceedings of the Fifteenth International Conference on Data Engineering (ICDE 1999)*, (pp. 106–115).

Han, J., & Kamber, M. (2001). *Data mining: concepts and techniques* San Francisco: Morgan Kaufmann Publishers (pp. 200–207).

Helmbold, D. P., & Long, P. M. (1994). Tracking drifting concepts by minimizing disagreements. *Machine Learning*, *14*, 27.

Hussain, F., Liu, H., Tan, C. L., & Dash, M. (1999). Discretization: An Enabling Technique. The National University of Singapore Technical Report TRC6/99, June.

Hussain, F., Liu, H., Suzuki, E., & Lu, H. (2000). Exception rule mining with a relative interestingness measure. *Proceedings of Pacific Asia Conference on Knowledge Discovery in Databases (PAKDD)*, 86–97.

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

Li, J., Dong, G., & 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)*, (pp. 220–232).

Liu, B., & Hsu, W. (1996). Post-analysis of learned rules. *Proceedings of the Thirteenth National Conference on Artificial Intelligence (AAAI-96)*, (pp. 828–834).

Liu, B., Hsu, W., & 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)*, (pp. 31–36).

Liu, B., Hsu, W., Ma, Y., & Chen, S. (1999). Mining Interesting Knowledge Using DM-II. *Proceedings of the Fifth International Conference on Knowledge Discovery and Data Mining (KDD-99)*, (pp. 430–434).

Liu, B., Hsu, W., Han, H.S., & Xia, Y. (2000). Mining Changes for Real-Life Applications. *Second International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2000)*, (pp. 337–346).

Nakhaeizadeh, G., Taylor, C., & Lanquillon, C. (1998). Evaluating usefulness of dynamic classification. *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98)*.

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

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

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

Thomas, S., Bodagala, S., Alsabti, K., & Ranka, S. (1997). An efficient algorithm for the incremental updation of association rules in large databases. *Proceedings of the Third International Conference on Knowledge Discovery in Databases*, (pp. 263–266).

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