

Mining the change of event trends for decision support in environmental scanning

Duen-Ren Liu ^{a,*}, Meng-Jung Shih ^a, Churn-Jung Liao ^b, Chin-Hui Lai ^a

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

^b *Institute of Information Science, Academia Sinica, Taipei, Taiwan*

Abstract

As the business environment has become increasingly complex, the demand for environmental scanning to assist company managers plan strategies and responses has grown significantly. The conventional technique for supporting environmental scanning is event detection from text documents such as news stories. Event detection methods recognize events, but neglect to discover the changes brought about by the events. In this work, we propose an event change detection (ECD) approach that combines association rule mining and change mining techniques. The approach detects changes caused by events to help managers respond rapidly to changes in the external environment. Association rule mining is used to discover event trends (the subject patterns of events) from news stories. The changes can be identified by comparing event trends in different time periods. The empirical evaluation showed that the discovered event changes can support decision-makers by providing up-to-date information about the business environment, which enables them to make appropriate decisions. The proposed approach is practical for business managers to be aware of environmental changes and adjust their business strategies accordingly.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Information processing and management; Association rule mining; Change mining; Environmental scanning; Event detection; Event tracking

1. Introduction

With the rapid growth of the Internet, the business environment has become more complex and dynamic, and environmental scanning by businesses has received considerable attention to assist managers plan an organization's future strategies (Choo, 1999). Because of their comprehensive content, news stories retrieved from News Web-sites on the Internet are the best sources of information for businesses to obtain environmental data (events). "Events" are happenings of interest that occur spontaneously at specific points in time. Event detection and tracking techniques support environmental scanning by identifying new events and tracking subsequent news stories that discuss the event of interest (Allen, Papka, & Lavrenko, 1998; Brants, Chen, & Farahat, 2003; Wei & Lee, 2004; Yang et al., 1999; Yang,

Pierce, & Carbonell, 1998), so that organization managers can be aware of new events in their environment. However, it is insufficient for business administrators to be notified only when a new event happens. Event detection and tracking methods are only concerned with recognizing new events and tracking events in news stories and neglect to discover the changes that occur between the news stories. From the view point of environmental scanning, it is important not only to identify events, but also to identify changes in event trends (the subject pattern of events) so that business managers can respond rapidly and appropriately to changes in the external environment. Thus, detecting event changes is critical for businesses.

To capture event changes, we must first determine the event trend. An event trend is a pattern found in most news stories about the same event, and can be characterized by the relationships between the 4Ws: when, who, where, and what. In this research, we employ association rule mining to identify the subject patterns of events in news stories.

* Corresponding author. Tel.: +886 3 5131245; fax: +886 3 5723792.
E-mail address: dliu@iim.nctu.edu.tw (D.-R. Liu).

Association rules discovered in news stories about the same event are regarded as the subject pattern of that event. For instance, if most news stories about the event “Telecom Services” report that telecommunication companies provide recreational services (the “what” property), this would represent the subject pattern of the event (event trend). An event change is a change in event trends in two time periods. For example, the market trend of the mobile telecommunication industry before the 2nd season of 2005 was to provide GSM services. From July of 2005, GSM services were replaced by 3G services, and 3G services became a hot topic in telecommunication markets. Such event change can be discovered from news stories during year 2003 to year 2007.

Information needs usually vary, depending on the time, situation, and people involved. Business managers may require different levels of information in order to develop different strategies. A decision-maker not only needs to know about the business operations of his/her own company, but also the operations of the industry that the company belongs to. Different levels of information form a concept hierarchy. To meet various information needs of business managers in environmental scanning, it is necessary to construct a concept hierarchy that describes the hierarchical relationships between event properties (attributes) according to the content of news stories.

Motivated by the need to capture event changes, this work adopts the association rule change mining technique (Song, Kim, & Kim, 2001) to develop an ECD (event change detection) technique. The change mining technique has been successfully used in transaction data to discover the changes of customer behaviors. However, the conventional change mining technique does not consider a concept hierarchy or unstructured data, such as news stories. To fill this gap, we modify the conventional change mining technique to discover the changes in event trends by combining event properties and a concept hierarchy to improve the quality of change detection. The proposed technique can provide useful information about environmental changes to enhance environmental scanning on the Internet.

The remainder of this paper is organized as follows. We first review literature relevant to this research, including environmental scanning, event detection and tracking, association rule mining, and change mining technologies. We then give an overview of our event change detection (ECD) technique. This is followed by detailing the methods for detecting event changes. Finally, we report the experimental results and conclude the paper.

2. Background and related work

2.1. Environmental scanning

As the business environment becomes more complex and dynamic, more unexpected situations may occur. Managers can adapt to this environment and develop effective responses to secure or improve a company’s position by

using environment scanning (Choo, 1999; Jennings & Lumpkin, 1992), which provides information about the external business environment. It is essential that managers scan the external business environment to make appropriate decisions.

2.2. Event detection and tracking

Event detection and tracking techniques support the detection of new events and track subsequent news stories of existing events. Event detection focuses on identifying new events from news stories. The goal of event detection is to identify news stories that discuss new events (Allen et al., 1998; Brants et al., 2003; Wei & Lee, 2004; Yang et al., 1999; Yang, Pierce et al., 1998), so that managers can be aware of new events in the business environment. Event tracking starts with a set of pre-classified news stories, and searches for all subsequent stories that discuss the event of interest (Allen et al., 1998; Yang et al., 1999). The goal of event tracking is to find follow-up stories related to the event. Event detection and tracking methods are only concerned with tracking events and recognizing new events from news stories and neglect to discover the changes that occur between these news stories.

Event tracking identifies which event a news story belongs to according to the feature sets of events. Feature extraction and selection are often used to identify the feature set for each event from a set of pre-classified training news documents (Yang et al., 1999). The feature extraction phase parses each training news story and produces a list of terms referred to as features. After feature extraction, feature selection condenses the size of the event feature set. This phase removes unnecessary terms from the set produced in the previous phase. Several feature selection methods have been proposed in the literature, including *tf* and *tf-idf* (Salton & Buckley, 1988).

A news story or event can be represented as a feature set of weighted terms using a term weighting approach. The weight of a term (feature) indicates its degree of importance in representing the document. The well-known *tf-idf* approach, which is often used for term weighting (Porter, 1980), considers that frequently occurring terms are better discriminators to represent a document, especially when they do not appear frequently in other documents. In addition, event properties have been used to improve event tracking and detection (Allen et al., 1998; Wei & Lee, 2004).

2.3. Association rule mining

Data mining techniques have been broadly used in various fields of information science (Chen & Liu, 2004; Lu, Kou, Zhao, & Chen, 2007; Kuo, Lin, & Shih, 2007; Yen & Lee, 2006). Association rule mining is a data mining technique widely used in various applications, such as market basket analysis. The technique searches for interesting associations or relationships among items in a large data

set (Han & Kamber, 2001). Different association rules express different regularities that exist in a dataset. Two measures, support and confidence, are used to determine whether a mined rule is a regular pattern (Han & Kamber, 2001; Ian & Eibe, 2000). The support measure determines the probability that a transaction contains both the conditional and consequent parts of a rule, while the confidence measure is the conditional probability that a transaction containing the conditional part of a rule also contains the consequent part. The *apriori* algorithm (Agrawal & Srikant, 1994) is typically used to find association rules by discovering frequent *itemsets* (sets of items), which are considered to be frequent if their support exceeds a user-specified minimum support threshold. Association rules that meet a user-specified minimum confidence can then be generated from the frequent *itemsets*.

As mentioned before, the purpose of this work is to discover event trends. Discovering the subject patterns of an event is helpful to determine the relationships between subjects. Association rule mining can discover frequent patterns, which represent the major behaviors of subjects, and are valuable for environmental monitoring. In this work, we apply association rule mining to news data to find the subject patterns (rule patterns) of events. The formats of news data and transaction data are very different. Transaction data is structured, and its attributes and values are often fixed. In contrast, news data is unstructured, i.e., a free text format. Therefore, mining association rules from news data is quite different to that of transaction data. In this study, the subject patterns of events are extracted according to user-defined event properties and the concept hierarchy of event properties.

2.4. Change mining

The objective of change mining is to discover the changes of data (e.g. customer behaviors) between two datasets from different time periods. The approaches to change mining can be classified as follows:

- *Decision tree models*: This method constructs decision trees for two datasets, and then derives the differences by comparing the two decision trees (Liu & Hsu, 1996; Liu, Hsu, Han, & Xia, 2000).
- *Association rules*: This method determines changes by comparing the association rules mined from two datasets (Chen, Chiu, & Chang, 2005; Liu, Hau, & Ma, 2001; Song et al., 2001). Users can decide the type of rule changes according to the similarities and differences between the rules in the datasets. There are several types of possible mining change pattern (Chen et al., 2005; Dong & Li, 1999; Lanquillon, 1999; Liu & Hsu, 1996; Song et al., 2001).
- *Emerging patterns*: The concept of emerging patterns captures significant changes between datasets. An emerging pattern is a rule pattern whose support increases significantly from one dataset to another.

- *Unexpected consequent changes*: These changes can be found in a newly discovered association rule whose consequent parts differ from the previous rule patterns.
- *Added rules*: These are new rules that only exist in the present dataset.
- *Perished rules*: These are rules that only exist in the previous dataset.

Association rule change mining techniques are used to analyze transaction data in order to discover changes in customer behavior. This work identifies event changes from news data. Conventional change mining techniques pay a little attention to consider unstructured data or rules with multiple-attributes in the consequent part. To overcome such difficulties, we modify the association rule change mining technique to process rules with a more general format that has multiple-attributes in the consequent part. An event's properties and concept hierarchy are also considered to improve the quality of change detection.

3. Event change detection technique

In a dynamic business environment, events are constantly evolving. It is important, therefore, for business managers to be aware of environmental changes and adjust their business strategies accordingly. The proposed event change detection (ECD) technique comprises three processes: event identification, tracking, and change detection, as shown in Fig. 1.

From a training set of news stories, the event identification process identifies the feature set of an event. The event tracking process identifies to which event a news story belongs according to the feature set of the event. News stories of an event are classified into different time periods of news datasets based on their reporting time. The change detection process identifies event changes. First, an event's properties and the concept hierarchy of the properties are identified. The properties are used to extract the important content of a news story. News datasets are transformed into property datasets based on event properties, after which association rule mining is used to extract the subject patterns of the event (event patterns) from the event property datasets. The extracted patterns are expressed in rule format representing the frequent association of event properties. The process then analyzes the event patterns to identify event changes.

3.1. News fetcher

The news fetcher obtains and processes news stories from news providers on the Internet.

3.1.1. News searching

This feature looks for up-to-date news stories from online news providers. News searching checks whether any new stories have been published. If a new story is discovered, news fetching is activated.

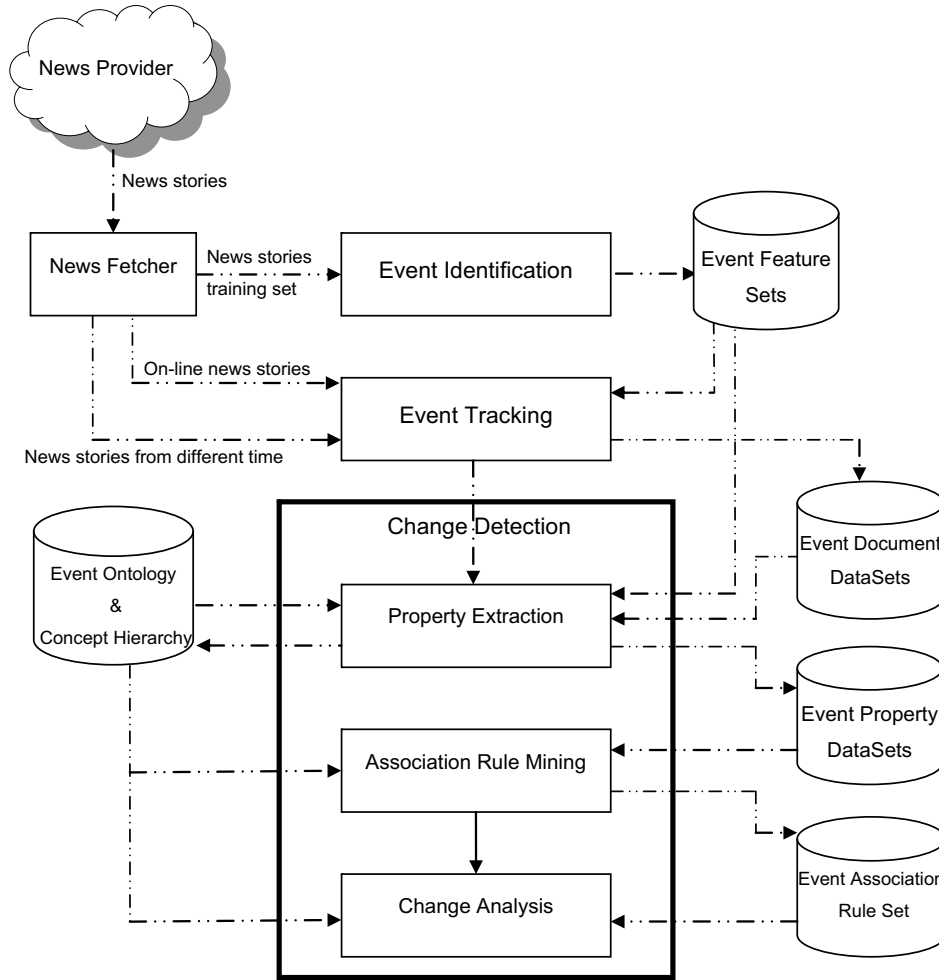


Fig. 1. Overview of the ECD technique.

3.1.2. News fetching

If a news story is found, news fetching acquires the news story (in HTML format) from the news website and stores it in the database.

3.1.3. Transforming

This step transforms raw news stories from HTML format into text format, filters out non-news data (i.e., advertisements, website links, etc.), and extracts the following news content related data: the report’s date, reporter, title, main body of news, and source.

3.2. Event identification

The objective of event identification is to identify a feature set for each event from a set of pre-classified training news documents. All training news stories are labeled to indicate which event they belong to. This step also determines the feature sets that will be used to represent the news stories. The process is comprised of two phases: feature extraction and feature selection.

The first phase extracts a set of terms from the news stories. To conduct feature extraction, we use a Chinese Dic-

tionary (Chinese Dictionary, 2003), which contains one hundred and sixty thousands terms, to parse each training news story and produce a list of terms referred to as features. After feature extraction, feature selection condenses the event feature set. We select the features by *tf-idf* approach (Salton & Buckley, 1988).

Let F_D and F_e be the feature sets of a news story D and an event e , respectively. w_{Dk} is the weight of the representative term, t_k , of news story D . Let the term frequency tf_{Dk} be the occurrence frequency of term t_k in D , and let the document frequency df_k represent the number of news stories that contain the term t_k . The importance of term t_k to D is proportional to the term frequency and inversely proportional to the document frequency, which is expressed as equation:

$$w_{Dk} = \frac{1}{\sqrt{\sum_{t_k} (tf_{Dk} \times (\log \frac{N}{df_k} + 1))^2}} tf_{Dk} \times \left(\log \frac{N}{df_k} + 1 \right) \tag{1}$$

where N is the total number of news stories in the training set and the denominator on the right-hand side of the equation is a normalization factor to normalize the

weight of the term. The top-N features with the highest term weights are selected to represent each news story.

The feature set for each event can be selected from a set of labeled training news stories. We adopt the *tf-idf* approach for document classification (Langari & Tompa, 2001) to derive the feature set for each event. Let w_{ek} be the weight of the representative term, t_k , of an event e . Let the term frequency tf_{ek} be the occurrence frequency of term t_k , in event e , and let the event frequency ef_k represent the number of events that contain at least one occurrence of term t_k . The importance of term t_k to event e is proportional to the term frequency and inversely proportional to the event frequency, which is expressed as equation:

$$w_{ek} = \frac{1}{\sqrt{\sum_{t_k} (tf_{ek} \times (\log \frac{M}{ef_k} + 1))^2}} tf_{ek} \times \left(\log \frac{M}{ef_k} + 1 \right) \quad (2)$$

where M is the total the number of events and the denominator on the right-hand side of the equation is a normalization factor to normalize the weight of the term. The top-N features with the highest term weights are selected to represent each event.

3.3. Event tracking and detection

The event tracking process identifies to which event a news story belongs according to the feature set of events. The process can also be used to determine whether a news story discusses a new event. We focus on detecting the changes in the trends of existing events. Thus, event tracking is used to collect news stories of events of interest, rather than detecting new events. The process comprises three steps: news document representation, similarity comparison, and event assignment.

3.3.1. News document representation

Each news document is represented by its features.

3.3.2. Similarity comparison

This step calculates the similarity between the news story D and all known events. The cosine distance is used to compute the similarity as equation:

$$\text{Sim}(D, e) = \frac{\sum_{t_k} w_{Dk} \times w_{ek}}{\sqrt{\sum_{t_k} w_{Dk}^2} \sqrt{\sum_{t_k} w_{ek}^2}}, t_k \in F_D \cup F_e \quad (3)$$

where F_D and F_e are the feature sets of document D and an event e , respectively; w_{Dk} is the weight of the representative feature t_k , of D ; and w_{ek} is the weight of the representative feature t_k of e . Note that w_{Dk} is zero if $t_k \notin F_D$; and w_{ek} is zero if $t_k \notin F_e$.

3.3.3. Event assignment

After obtaining the similarity scores for the new news story D and all known events, we set a pre-defined threshold and assign event labels according to the similarity scores. If the maximal similarity score for D and the known

events is below the threshold, the new document is labeled as the first story of a novel event; otherwise, we assign the news document to the event with the maximal similarity.

3.4. Change detection

The change detection process, which identifies event changes, comprises three steps: property extraction, association rule mining and change analysis.

3.4.1. Property extraction

Event feature sets, produced by the event identification process to represent specific events and news stories, are used for event detection and tracking. Although event features are useful for identifying which event a news story belongs to, some features cannot represent the important content of a news story. For example, a feature “Andy Chen” is extracted because Andy Chen is a famous reporter of the active event. However, the reporter is irrelevant to the main content of the news story. We focus on identifying the changes of event trends that denote the important content of a news story. Accordingly, user-defined event properties are selected from the event feature sets to improve the accuracy of content representation. The event properties are defined by users to extract the important semantic meaning (attributes) of new stories.

A knowledge-engineering process identifies event properties and the concept hierarchy of the properties. Event properties are classified into the four categories (4Ws) (Mayeux, 1996): (1) When: date, time; (2) Who: person, organization; (3) Where: location; and (4) What: action, claim, standpoint, statement. The concept hierarchies are constructed from the defined event properties, and consist of the hierarchical relationships between the properties. Once the properties of each news story have been extracted, the news document datasets can be transformed into event property datasets based on the extracted event properties.

3.4.2. Association rule mining

Association rule mining is used to extract the subject patterns of events from the event property datasets. The extracted patterns represent the frequent associations of event properties, namely the event trends, and are expressed in a rule format. Note that associations of different concept levels of event properties can be extracted according to the concept hierarchy. The event patterns are stored in the event association rule set for further change analysis.

3.4.3. Change analysis

The process then analyzes the event patterns to identify event changes, as described in the following section.

4. Detection of event change

An event change is the change in an event’s trends over two time periods. The trends are detected from news stories

of the same event in different periods. Song et al. (2001) studied the problem of mining changes in customer behavior, and proposed a methodology to detect changes in data in different time periods. Their approach has the following features: (1) the methodology is applied to transaction data. (2) The format of compared rules is specialized for transaction data – there is only one attribute in the consequent part. (3) Although they define three types of change, they focus on unexpected consequent changes. (4) They compute the similarity degree of matched attributes without considering a concept hierarchy.

In general, there may be multiple-attributes in both the conditional and the consequent parts of a rule. Furthermore, business decision-makers may require different levels of information (a concept hierarchy) to develop different strategies. To generalize the range of applications, we extend the association rule change mining technique by considering a multiple-attribute rule format and a concept hierarchy to enhance the event change detection technique. The detection of event changes is illustrated in Fig. 2.

4.1. Types of event changes

Based on past research and business requirements, we define five types of possible change in event patterns.

4.1.1. Emerging event patterns

An emerging event pattern is a rule pattern whose support increases significantly from one dataset to another.

According to the definition in (Song et al., 2001), the essential condition of an emerging pattern is that the two rules are the same. But in the real world, attributes (concepts) at different levels of a concept hierarchy have some degree of similarity that are worth analyzing and exploring as emerging patterns. Accordingly, we consider the concept hierarchy when computing the degree of similarity for comparison of event patterns (rules).

4.1.2. Unexpected consequent changes of event patterns

Unexpected consequent changes of event patterns can be found in newly discovered event patterns whose consequent parts are different from previous event patterns. In previous research (Song et al., 2001), unexpected consequent change detection considered a single attribute in the consequent part without considering the concept hierarchy, which limited the number of detected changes. We have combined concept hierarchies to compute similarities when detecting unexpected consequent changes.

4.1.3. Unexpected condition changes of event patterns

Unlike unexpected consequent changes of event patterns, unexpected condition changes are newly discovered association rules whose conditional parts differ from previous rules. Previous studies did not focus on detecting unexpected condition changes. Our approach detects unexpected condition changes.

4.1.4. Added event patterns

An added event pattern is a new rule, i.e., a rule not found in previous patterns.

4.1.5. Perished event patterns

A perished event pattern is the opposite of an added rule, as it is only found in past events.

4.2. Discovery of event change patterns

The objective of ECD is to detect the five types of event change in different time periods. We start by dividing a temporally ordered stream of news stories about the same event into several groups. From these datasets, the system mines association rules, which represent the subject patterns of events in different time periods. We then compute the similarity measures and difference measures of the event patterns in different time periods. Finally, based on these measures, we can identify five types of event changes.

In this study, we adapt the rule matching method proposed by Song et al. (2001) to fit event change detection by considering a concept hierarchy. The rule expressions are shown in Fig. 3. The left-hand side of each expression represents the conditional part of the rule, while the right-hand side represents the consequent part. Several notations are defined in Section 4.2.1 to identify the five types of event changes. The modified rule matching method is detailed in Section 4.2.2.

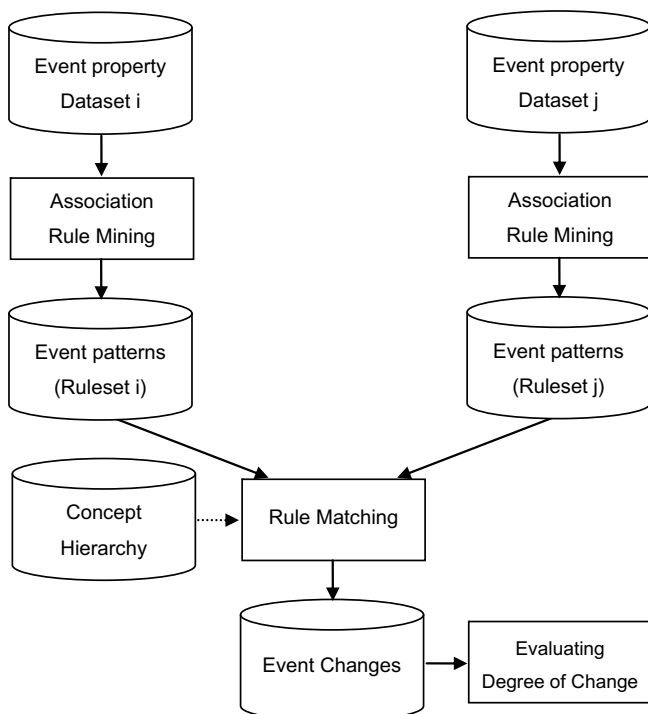


Fig. 2. The process of event change detection.

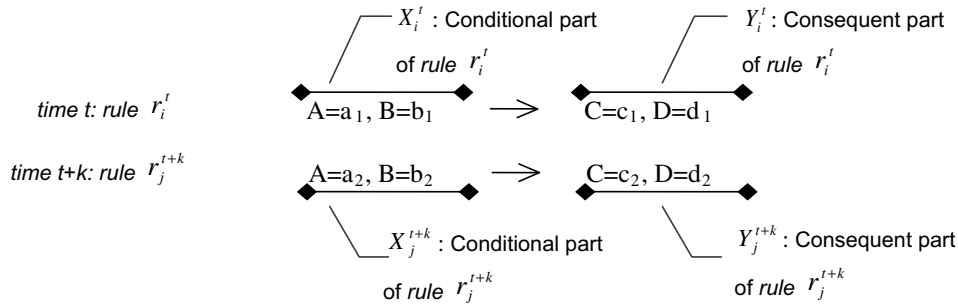


Fig. 3. Rule expressions.

4.2.1. Definitions and conventions

We use the following notations to represent the elements in the calculation process, which computes the similarity measures and difference measures of the event patterns r_i^t and r_j^{t+k} in time t and time $t + k$, respectively.

4.2.1.1. Conditional part of rules

- p_{ij} degree of attribute match of the conditional part
- $P_{ij} = \frac{|A_{ij}|}{\max(|X_i^t|, |X_j^{t+k}|)}$
- A_{ij} set of attributes common to the conditional parts of r_i^t and r_j^{t+k}
- X_i^t set of attributes in the conditional parts of r_i^t
- X_j^{t+k} set of attributes in the conditional parts of r_j^{t+k}
- A_{ijk} the k th attribute in A_{ij}
- $v(r_i^t, A_{ijk})$ value of the k th attribute in A_{ij} of r_i^t
- $v(r_j^{t+k}, A_{ijk})$ value of the k th attribute in A_{ij} of r_j^{t+k}
- l_{ijk} degree of value match of the k th matching attribute in A_{ij}
- $l_{ijk} = 1 - \eta_H(v(r_i^t, A_{ijk}), v(r_j^{t+k}, A_{ijk}))$
- $\eta_H(v(r_i^t, A_{ijk}), v(r_j^{t+k}, A_{ijk}))$ is calculated according to the formulation of the node difference in the concept hierarchy (Eq. (4))
- C_{ij} similarity degree of the conditional parts of r_i^t and r_j^{t+k}
- $C_{ij} = \frac{p_{ij} \times \sum_{k=1}^{|A_{ij}|} l_{ijk}}{|A_{ij}|}$

4.2.1.2. Consequent part of rules

- q_{ij} degree of attribute match of the consequent part
- $Q_{ij} = \frac{|B_{ij}|}{\max(|Y_i^t|, |Y_j^{t+k}|)}$
- B_{ij} set of attributes common to the consequent parts of r_i^t and r_j^{t+k}
- Y_i^t set of attributes in the consequent parts of r_i^t
- Y_j^{t+k} set of attributes in the consequent parts of r_j^{t+k}
- B_{ijm} the m th attribute in B_{ij}
- $v(r_i^t, B_{ijm})$ value of the m th attribute in B_{ij} of r_i^t
- $v(r_j^{t+k}, B_{ijm})$ value of the m th attribute in B_{ij} of r_j^{t+k}
- f_{ijm} degree of value match of the m th matching attribute in B_{ij} (calculate the difference of two values according to their positions in the concept hierarchy)
- $f_{ijm} = 1 - \eta_H(v(r_i^t, B_{ijm}), v(r_j^{t+k}, B_{ijm}))$

$\eta_H(v(r_i^t, B_{ijm}), v(r_j^{t+k}, B_{ijm}))$ is calculated according to the formulation of the node difference in the concept hierarchy (Eq. (4))

Q_{ij} similarity degree of consequent parts of r_i^t and r_j^{t+k}

$$Q_{ij} = \frac{q_{ij} \times \sum_{m=1}^{|B_{ij}|} f_{ijm}}{|B_{ij}|}$$

4.2.1.3. Similarity measure and difference measure. The similarity measure S_{ij} between r_i^t and r_j^{t+k} is calculated using the following formula: ($0 \leq S_{ij} \leq 1$)

$$S_{ij} = \begin{cases} C_{ij} \times Q_{ij}, & \text{if } |A_{ij}| \neq 0 \text{ and } |B_{ij}| \neq 0 \\ 0, & \text{otherwise} \end{cases}$$

The maximum similarity measure of r_i^t : $\zeta_i = \max S_{ij}$.

The maximum similarity measure of r_j^{t+k} : $\zeta_j = \max S_{ij}$.

The second judged factor, i.e., the difference measure δ_{ij} between r_i^t and r_j^{t+k} is given by $\delta_{ij} = C_{ij} - Q_{ij}$ ($-1 \leq \delta_{ij} \leq 1$, $|\delta_{ij}| \leq 1$).

4.2.2. Steps of identifying event change

The rule matching method computes the similarity measures and difference measures of the event patterns r_i^t and r_j^{t+k} in time t and time $t + k$, respectively. The modified rule matching method comprises four steps.

- Step 1. Calculate the similarity degree of the conditional/consequent parts of two rules in different time periods.
- Step 2. Calculate the similarity measure S_{ij} between two rules. The measure is derived by multiplying the similarity degree of the conditional parts (C_{ij}) by the similarity degree of the consequent parts (Q_{ij}).
- Step 3. Calculate the difference measure δ_{ij} between two rules. The measure is the similarity degree of the conditional parts minus the similarity degree of the consequent parts.
- Step 4. Determine the type of event changes according to the similarity measures and difference measures.

Steps 1–3 are detailed in the following. Step 4 is detailed in Section 4.3.

Step 1. Calculate the similarity degree C_{ij}/Q_{ij} of the conditional/consequent parts of two rules.

The similarity degree of the conditional parts, C_{ij} , is the similarity between the conditional parts of rule r_i^t and rule r_j^{t+k} , derived by matching the values of the attributes of the conditional parts of the two rules. The similarity degree of the consequent parts, Q_{ij} , is the similarity between the consequent parts of rule r_i^t and rule r_j^{t+k} , derived by matching the values of the attributes of the conditional parts of the two rules. The detailed formulations for calculating the similarity degree are specified in Section 4.2.1.

Song et al.’s matching method computes the similarity degree based on binary matching. For example, if the value of the attribute “Service” in Rule_{*i*} is “Download MIDI Ringtone” and the value of the attribute “Service” in Rule_{*j*} is “Download MP3 Ringtone”, the similarity degree of these two values is counted as 0. But “Download MIDI Ringtone” and “Download MP3 Ringtone” both belong to “Download Ringtone”, so their similarity degree is high based on the concept hierarchy. Accordingly, we may be unable to find meaningful event changes with Song’s method. To identify such changes and generalize rule matching method, we consider concept hierarchies when calculating similarity degrees. The details are presented in the following.

Deriving the similarity degree of attributes based on concept hierarchies. Different levels of information form a concept hierarchy, which we use to explore all possible information needs and define the hierarchical relationships between event properties (attribute values).

Fig. 4 shows an example of a concept hierarchy. During the event change detection process, we need to determine the difference between the attribute values of two event patterns. These values are the nodes of the concept hierarchy. The difference between two attribute values in the concept hierarchy is derived by the equation:

$$\eta_H(A, B) = \frac{\text{Max}(\sum_{L_i \in P_A} WL_i, \sum_{L_j \in P_B} WL_j) - \sum_{L_k \in P_{\text{comm}}(A, B)} WL_k}{\text{Max}(\sum_{L_i \in P_A} WL_i, \sum_{L_j \in P_B} WL_j)} \quad (4)$$

where A and B are nodes in the concept hierarchy; P_A is the path from the root to node A ; P_B is the path from the root to node B ; $P_{\text{comm}}(A, B)$ is the common path between P_A and P_B ; L_i is a link i in P_A ; L_j is a link j in P_B ; WL_i is the weight on the level of link i ; and WL_j is the weight on the level of link j .

The similarity degree of the attribute match value is equal to $1 - \eta_H(A, B)$. To illustrate Eq. (4), the weight on the level 1 link is set to 1; on the level 2 link, it is set to 0.5; and on the level 3 link, it is set to 0.3. The difference between “MIDI” and “MP3” is: $\eta_H(\text{MIDI}, \text{MP3}) = \frac{0.3}{1.8} = 0.167$.

Both $\sum_{L_i \in P_{\text{MIDI}}} WL_i$ and $\sum_{L_j \in P_{\text{MP3}}} WL_j$ are equal to 1.8, and $\sum_{L_k \in P_{\text{comm}}(\text{MIDI}, \text{MP3})} WL_k$ is equal to 1.5. The similarity degree of “MIDI” and “MP3” is $1 - 0.167 = 0.833$.

Step 2. Calculate the similarity measure S_{ij} between two rules.

The measure is derived by multiplying the similarity degree of the conditional parts (C_{ij}) by the similarity degree of the consequent parts (Q_{ij}). $S_{ij} = C_{ij} \times Q_{ij}$. If the conditional (resp. consequent) parts of rule r_i^t and rule r_j^{t+k} are the same, the similarity degree of the parts will be 1. However, if the conditional (resp. consequent) parts of the two rules are completely different, the similarity degree of the parts will be 0. The similarity measure shows the similarity of the two compared rules by considering their conditional parts and consequent parts; the larger the similarity measure, the more similar the two rules will be.

Step 3. Calculate the difference measure δ_{ij} between two rules.

The measure is the similarity degree of the conditional parts (C_{ij}) minus the similarity degree of the consequent parts (Q_{ij}). $\delta_{ij} = C_{ij} - Q_{ij}$. The measure shows the change between the two rules. If it is greater than 0, the conditional parts of the two rules are alike, but the consequent parts are quite different. If the difference measure is less than 0, the conditional parts are different, but the consequent parts are similar.

4.3. Identifying the type of event changes

In the final step of rule matching, we identify the type of changes in event patterns according to the judged factors, i.e., the similarity measure S_{ij} and the difference measure δ_{ij} .

4.3.1. Emerging event patterns

An emerging event pattern is an event pattern in time t that also appears in time $t + k$. The mined rule r_i^t is similar to r_j^{t+k} ; thus the similarity degree of both the conditional and the consequent parts between r_i^t and r_j^{t+k} is high. Because the definitions of ‘similar’ and ‘different’ are subjective, the parameter θ_{em} is a threshold used to determine whether the two rules are similar or not. The rule r_i^t is classified as an emerging pattern with respect to r_j^{t+k} when the similarity measure between r_i^t and r_j^{t+k} (denoted by S_{ij}) is greater than θ_{em} .

4.3.2. Perished event patterns

When a subject pattern in time t is very different from the event patterns in time $t + k$, it is classified as a perished event pattern, which means that the mined rule r_i^t in time period t is quite different from all rules in time $t + k$. A perished event pattern is identified if the maximum similarity measure (denoted by ζ_i) between r_i^t and all rules in time period $t+k$ is less than $\theta_{\text{a/p}}$. Note that $\zeta_i = \max S_{ij}$. The parameter $\theta_{\text{a/p}}$ is a threshold used to determine whether there are any rules similar to the target rule.

4.3.3. Added event pattern

An added event pattern is an event pattern in time $t + k$ that is quite different from the event patterns in time t . This

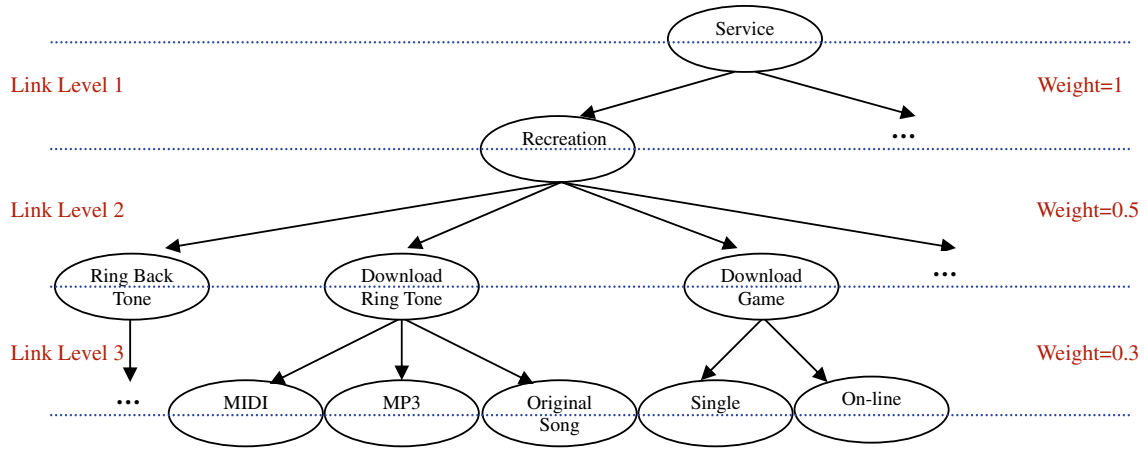


Fig. 4. An example of a concept hierarchy.

means that the mined rule r_j^{t+k} from time period $t+k$ is quite different from all rules in time t . Thus, if the maximum similarity measure between r_j^{t+k} and all rules in time period t (denoted by ς_j) is less than $\theta_{a/p}$, an added event pattern occurs. Note that $\varsigma_j = \max_i S_{ij}$.

4.3.4. Unexpected consequent changes of event patterns

According to the definition of unexpected consequent changes of event patterns, the conditional parts of rule r_i^t in time t and rule r_j^{t+k} in time $t+k$ are similar, but the consequent parts are different. It seems reasonable to assume that if the difference measure ∂_{ij} between r_i^t and r_j^{t+k} is greater than 0, an unexpected consequent change of an event pattern has occurred. But before determining whether such an event pattern has occurred in time period $t+k$, we must confirm that there is no similar event pattern in time period $t+k(t)$. For example, the difference measure ∂_{ij} between rule r_i^t in time t and rule r_j^{t+k} in time $t+k$ is 0.76, but the similarity measure, S_{ij} between r_i^t and r_j^{t+k} (another rule r_i^{t+k} in time $t+k$) is 0.85. The high similarity measure shows r_i^t is similar to r_i^{t+k} ; thus, r_i^t is classified as an emerging event pattern with respect to r_i^{t+k} . Therefore, r_i^t and r_j^{t+k} cannot be regarded as unexpected consequent changes. In this example, it is important to recognize the maximum similarity measure for r_i^t and r_j^{t+k} (denoted by $\text{Max}(\varsigma_i, \varsigma_j)$; $\varsigma_i = \max_j S_{ij}$; $\varsigma_j = \max_i S_{ij}$). If $\text{Max}(\varsigma_i, \varsigma_j) < \theta_{em}$, there is no similar rule to r_i^t or r_j^{t+k} ; thus, if the difference measure is large enough, the event pattern can be identified as an unexpected consequent change. To decide the type of change, we first eliminate emerging event patterns based on the maximum similarity measure. The parameter θ_{un} is a threshold used to determine whether the rules are sufficiently different ($\partial_{ij} > \theta_{un}$) to be classified as unexpected consequent changes.

4.3.5. Unexpected condition changes of event pattern

Similar to the unexpected consequent change of event patterns, we first eliminate emerging patterns ($\text{Max}(\varsigma_i, \varsigma_j) < \theta_{em}$), and then determine whether the rules

are unexpected condition changes based on the difference measure. An unexpected condition change of event patterns occurs when the consequent parts of rule r_i^t in time t and rule r_j^{t+k} in time $t+k$ are similar, but the conditional parts are dissimilar. To determine whether such a change has occurred, we must consider both the difference measure and the absolute value of the difference measure. If the difference measure is less than 0, the consequent parts are similar and the conditional parts are quite different. If the absolute value of difference measure is greater than $\theta_{un}(|\partial_{ij}| > \theta_{un})$, the rules r_i^t and r_j^{t+k} are sufficiently different to be classified as unexpected condition changes.

Table 1 shows the measurement for determining each type of event change, which is adopted and modified from (Song et al., 2001) by adding the measurement for unexpected condition change and three event change thresholds – θ_{em} , θ_{un} , and $\theta_{a/p}$. The five types of event change can be classified according to the two judged factors and three predefined thresholds: θ_{em} for emerging patterns, θ_{un} for unexpected consequent and unexpected condition changes, and $\theta_{a/p}$ for added and perished rules. Note that $\theta_{em} > \theta_{un} > \theta_{a/p}$.

Table 1 Measurement for each type of event change

Type of change (r_i^t, r_j^{t+k})	Measurement
Emerging event pattern	$S_{ij} \geq \theta_{em} (S_{ij} = C_{ij} \times Q_{ij})$ (C_{ij} : similarity degree of the conditional parts, Q_{ij} : similarity degree of the consequent parts)
Unexpected consequent change of event pattern	$\text{Max}(\varsigma_i, \varsigma_j) < \theta_{em}, \partial_{ij} > \theta_{un} (\partial_{ij} = C_{ij} - Q_{ij})$
Unexpected condition change of event pattern	$\text{Max}(\varsigma_i, \varsigma_j) < \theta_{em}, \partial_{ij} < 0, \partial_{ij} > \theta_{un}$
Added event pattern	$\varsigma_j < \theta_{a/p} (\varsigma_j = \max_i S_{ij})$
Perished event pattern	$\varsigma_i < \theta_{a/p} (\varsigma_i = \max_j S_{ij})$

In the process of determining the types of event changes, there is a pre-determined sequence. First, we decide emerging event patterns. If the similarity measure S_{ij} is greater than or equal to θ_{em} , it means that the two rules are similar and rule r_j^{t+k} can be regarded as an emerging event pattern. If the maximum similarity measure $\text{Max}(\zeta_i, \zeta_j)$ is less than θ_{em} and the difference measure δ_{ij} is greater than θ_{un} , we regard rule r_j^{t+k} as an unexpected consequent change of the event pattern. Note that $\zeta_i = \max S_{ij}$; $\zeta_j = \max S_{ij}$. If the difference measure δ_{ij} is less than 0 and the absolute value of the difference measure is greater than θ_{un} , we regard rule r_j^{t+k} as an unexpected condition change of event pattern. Finally, if ζ_j is less than $\theta_{a/p}$, rule r_j^{t+k} is identified as an added event pattern; and if ζ_i is less than $\theta_{a/p}$, rule r_i^t is identified as a perished event pattern.

4.4. Evaluating the degree of event changes

As a large number of changes occur in the business environment, managers need to focus on the essential ones. To achieve this goal, it is important to evaluate the degree of change, and rank changed rules according to their importance. Song et al. (2001) propose a method to calculate the degree of change between two rules. Their approach is based on the features of their data, i.e., the consequent part can only have one attribute. They focus on unexpected consequent changes without considering unexpected condition changes; therefore, they used the theory of unexpectedness to express the degree of change of unexpected changes. In our study, we consider a general case – multi-attributes in both the conditional and consequent parts. To assess the degree of unexpected changes, we compute it according to the change in the ratio of the rules' support values. Table 2 shows the simple formulations for measuring the degree of change. The formulations are adopted and modified from (Song et al., 2001) to measure the degree of unexpected changes for a general case – multi-attributes in both the conditional and consequent parts.

Let $\text{support}^t(r_i)$ and $\text{support}^{t+k}(r_j)$ represent the support value of r_i at time t and r_j at time $t+k$, respectively. The degree of change of an emerging event pattern shows the

change ratio of the support value between time t and time $t+k$. The degree of change for an unexpected consequent/condition change is the change ratio of r_i multiplied by the support value of r_j at time $t+k$. If $\text{support}^{t+k}(r_i)$ is large, the event pattern r_i^{t+k} will be found at time $t+k$, which means the event pattern r_i still exists at time $t+k$. r_i^{t+k} is regarded as an emerging rule to r_i^t . If r_i does not exist at time $t+k$, $\text{support}^{t+k}(r_i)$ is less than the user-defined minimum support and $\text{support}^{t+k}(r_i)$ is less than $\text{support}^t(r_i)$. In the worst case, $\text{support}^{t+k}(r_i)$ is equal to 0; thus, the larger $\text{support}^{t+k}(r_j)$ is, the larger the degree of change will be. The degree of change is affected by both the change ratio of the support value of r_i at time t and $t+k$ multiplied by $\text{support}^{t+k}(r_j)$.

The degree of change for a perished (resp. added) rule is obtained from the support value of the perished (added) rule multiplied by 1 minus the maximum similarity measure (ζ_i or ζ_j). The degree of change will be larger, if the perished (added) rule has less maximum similarity measure. After calculation of the degrees of change, the essential changes will be notified to business managers, who can then analyze the trends of event changes in different time periods, and use the information to understand business directions and plan appropriate strategies.

5. Empirical evaluation

We applied the proposed methodology to detect the changes of event patterns in a dataset “Telecom Services provided by Taiwan mobile telecommunication companies (Telecom Service)” collected from news websites on the Internet.

This section reports the empirical result of the proposed ECD technique. The dataset of “Telecom Services” is divided into two period datasets based on different time periods. The first part contained 156 news stories from April 2003 to March 2005 which are collected from a news website, ETtoday, <http://www.ettoday.com>. The second part contained 175 news stories from April 2005 to March 2007 which are collected from the same news website. The event properties and concept hierarchy as required by the ECD technique, were manually identified. The attributes of “Telecom Services” are: Company (telecom company), Tech (technology), User (target user of service), Service (service provided by telecom company), Co-Company (cooperative company). The concept hierarchies of the attributes are: Company {TWM, CHT, Fareastone, APBW, VIBO...}, Tech {2G [GSM, TDMA, cdmaone], 2.5G [GPRS], 3G [WCDMA, cdma2000, TDS-CDMA]}, User {Enterprise, Individual}, Service {Information [meteorology, news...], Educational [English Learning...], Recreational [Download Ringtone and Picture Download Ringtone [MIDI, Mp3, Original song], Download Game [Single, on-line], Messaging [SMS, Voice, Voice/Vedio]...}, Co-Company {Educational [StudioClassroom...], Bank [Chinatrust...], Mass medium [TV (TTV, CTITV...), News paper (AppleDaily, UDN...)]...}. We note that

Table 2
Measuring the degree of change in event trends

Type of change	Degree of change
Emerging event pattern	$\frac{\text{support}^{t+k}(r_j) - \text{support}^t(r_i)}{\text{support}^t(r_i)}$
Unexpected consequent change of an event pattern	$\frac{\text{support}^t(r_i) - \text{support}^{t+k}(r_i)}{\text{support}^t(r_i)} \times \text{support}^{t+k}(r_j)$
Unexpected condition change of an event pattern	$\frac{\text{support}^t(r_i) - \text{support}^{t+k}(r_i)}{\text{support}^t(r_i)} \times \text{support}^{t+k}(r_j)$
Added event pattern	$(1 - \zeta_j) \times \text{support}^{t+k}(r_j)$
Perished event pattern	$(1 - \zeta_i) \times \text{support}^t(r_i)$

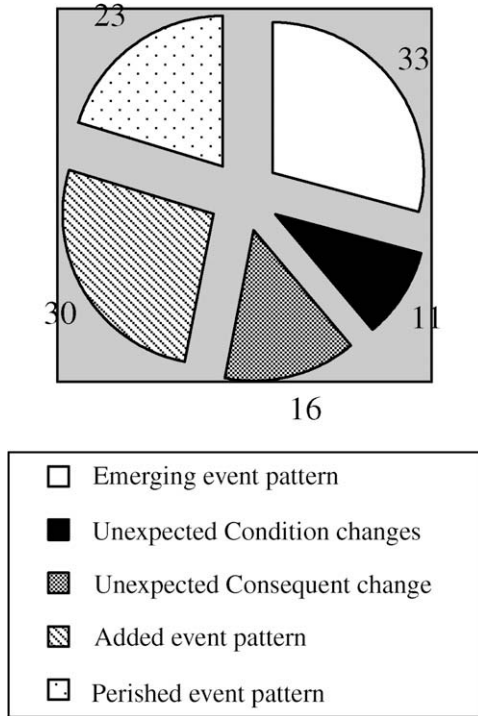


Fig. 5. The results of five types of changed event patterns (Telecom Services).

the parentheses are used to distinguish the hierarchical relationships. For example, Tech includes three kinds of techniques {2G, 2.5G, 3G}, while the 2G includes [GSM, TDMA, cdmaone].

Given $\theta_{em} = 0.75$, $\theta_{un} = 0.5$ and $\theta_{a/p} = 0.3$, we found 113 changed rules, including 33 emerging event patterns, 11 unexpected condition changes, 16 unexpected consequent changes, 30 added event patterns, and 23 perished event patterns (see Fig. 5). Some of which are listed in Table 3.

From the changed pattern (1) in Table 3, we can see the rapid growth (85.1%) in providing download original ringtone service to individual user. This information shows that providing download ringtone is one of the popular services for individual users between 2003 and 2005. And it became more important for mobile telecommunication companies to retain customers between 2005 and 2007. It is interesting to note that pattern (1) cannot be discovered according to traditional change mining methods, since the literal meanings of the two rules are different (Download Ringtone and Download Original Ringtone are different). But in the concept hierarchy, “Download Original Ringtone” belongs to “Download Ringtone”. The similarity between “Download Original Ringtone” and “Download Ringtone” in the concept hierarchy is high (similarity = 0.83). Therefore, rule “User = Individual → Service = Download Ringtone” and “User = Individual → Service = Download Original Ringtone” can be regarded as similar rules. From this emerging pattern, we know that download ringtone is a hot service for individual users, so the mobile communication companies can increase their ringtone content continually to attract new customers and retain customers.

From the changed pattern (3) (unexpected condition change) in Table 3, we find that APBW is a famous

Table 3
Some examples of changed event patterns (Telecom Services)

r_i^l	r_j^{l+k}	$Sup^l(r_i)$	$Sup^{l+k}(r_j)$	Degree
<i>Emerging event patterns</i>				
(1) User = Individual → Service = Download Ringtone	User = Individual → Service = Download Original Ringtone	0.141	0.261	0.851
(2) User = Enterprise → Service = Message broadcasting	User = Enterprise → Service = Message broadcasting	0.103	0.166	0.612
<i>Unexpected condition changes of event patterns</i>				
(3) Company = APBW → User = Individual, Service = Recent News	Company = VIBO → User = Individual, Service = TV news	0.103	0.16	0.089
(4) User = Individual, Co-Comp. = Bank → Service = finance and economics	User = Individual, Co-Comp. = TV → Service = finance and economics	0.122	0.147	0.086
<i>Unexpected consequent changes of event patterns</i>				
(5) Company = TWM, User = Individual → Service = voice mailbox	Company = TWM, User = Individual → Service = video mailbox	0.109	0.131	0.096
(6) Company = CHT, User = Enterprise → Service = mail	Company = CHT, User = Enterprise → Service = Push Mail	0.135	0.183	0.114
<i>Added event patterns</i>				
(7) User = Individual → Tech = 3G			0.486	0.486
(8) Company = VIBO, user = Individual → Service = Remote Surveillance			0.131	0.131
<i>Perished event patterns</i>				
(9) User = Individual → Tech = GSM		0.333		0.333
(10) Company = Mobitai → Service = Recreation		0.109		0.109

company that provided recent news to individual user in 2003–2005. But in 2005–2007, the new company VIBO replaced the position of APBW to provide news to individual users. Moreover, VIBO allowed its individual users to watch TV news program directly via their phones. We note that “Recent News” and “TV news” are different subjects in conventional rule matching method. The concept hierarchy is used to overcome the problem of exact-match. This changed pattern indicates that individual users like receiving real time news via TV news. The marketing managers of telecom companies should improve their news channel by incorporating TV to attract customers.

The changed pattern (5) (unexpected consequent change) shows that the individual mailbox service of TWN had changed from voice to video. This change can be regarded as a marketing trend. In the period between 2005–2007, most mobile telecommunication companies upgraded their communication technology from 2G to 3G. 3G (The third Generation Mobile System) is the solution to satisfy new mobile communication requirements. The system is based on new technologies of wireless communication with a very high speed access to the Internet services. The launching of 3G influences mobile telecommunication services, and the discovered pattern (5) is a typical example. The pattern suggest that marketing managers should spend marketing efforts in developing more efficient 3G services, such as video mailbox.

The added event patterns (7) in Table 3 indicate that 3G service became a hot topic during 2005–2007. This pattern is in accordance with the development of 3G service market in Taiwan. 3G service started in year 2002 in Taiwan. But only one company launched the market until the end of 2nd Season of 2005. From July of 2005, other telecommunication companies started to launch the 3G service market, and the action makes 3G become a hot topic.

Finally, several perished event patterns are also shown in Table 3. The changed rule (9) shows a perished trend of GSM service in which GSM service decreased during 2005–2007 gradually. The perished event pattern (10) points out that the company Mobitai did not provide recreational services during 2005–2007. The result agrees with the fact that “TWM (Taiwan Mobile)” was merged with “Mobitai telecommunication” in January 2006.

To evaluate the quality of the proposed ECD technique, we invited an expert in Telecom Services field to measure the empirical results. The expert assessed the quality of the discovered 113 changed event patterns based on whether the event pattern is meaningful in Telecom industry. We used hit ratio as the evaluation metric to represent the accuracy of empirical results. The hit ratio is computed as the ratio of the number of meaningful event patterns to the number of discovered event patterns. The higher hit ratio value indicates the higher quality of change detection. Table 4 shows that 22 emerging event patterns are determined by the expert as meaningful event patterns in the 33 discovered emerging patterns, and the hit ratio of emerging event pattern is 0.667.

Table 4
Quality measure of changed event patterns (Telecom Services)

	# of discovered event patterns	# of meaningful event patterns	Hit ratio
Emerging event patterns	33	22	0.667
Unexpected condition changes of event patterns	11	5	0.455
Unexpected consequent changes of event patterns	16	10	0.625
Added event patterns	30	23	0.767
Perished event patterns	23	12	0.522
Total	113	72	0.637

6. Conclusion

As the business environment becomes increasingly complicated, capturing environmental changes and being sensitive to the business environment is crucial to success in business. Current research into environmental scanning emphasizes event detection and tracking, the main purpose of which is to identify which event a news story describes. It is insufficient for business managers to obtain environmental change information besides knowing what event occurs. We have proposed an event change detection (ECD) technique to capture the changes of event trends. The proposed technique combines a change mining approach and a concept hierarchy to detect environmental changes to enhance environmental scanning on the Internet.

Our empirical evaluation showed that the discovered event changes can support decision-makers by providing up-to-date information about the business environment, which enables them to make appropriate decisions. The proposed approach is practical for business managers to be aware of environmental changes and adjust their business strategies accordingly. The event trends discovered in this work are expressed as association rules. Besides association rules, sequential patterns are also valuable knowledge that are worth to be explored. Our future work will investigate the mining of sequential patterns from news documents and the discovery of changes of event trends based on sequential patterns.

Acknowledgements

This research was supported in part by the National Science Council of the Taiwan (Republic of China) under the Grant NSC 95-2416-H-009-002.

References

- Agrawal, R. & Srikant, R. (1994). Fast algorithms for mining association rules. In *Proceedings of the 1994 International Conference Very Large Data Bases (VLDB'94)*. Santiago, Chile.
- Allen, J., Papka, R., & Lavrenko, V. (1998). On-line new event detection and tracking. In *Proceedings of the 21st annual international ACM SIGIR conference on research and development in information retrieval* (pp. 37–45). Melbourne, Australia: ACM Press.
- Brants, T., Chen, F., & Farahat, A. (2003). A system for event detection. In *Proceedings of the 26th annual international ACM SIGIR conference*

- on research and development in information retrieval. Toronto, Canada: ACM Press.
- Chen, M.-C., Chiu, A.-L., & Chang, H.-H. (2005). Mining changes in customer behavior in retail marketing. *Expert Systems with Applications*, 28(4), 773–781.
- Chen, S. Y., & Liu, X. (2004). The contribution of data mining to information science. *Journal of Information Science*, 30(6), 550–558.
- Chinese Dictionary. (2003). Available at: william.cswiz.org/techreport/moeedict. Accessed 5.10.2003.
- Choo, C. W. (1999). The art of scanning the environment. *Bulletin of the American Society for information Science and Technology*, 25(3), 21–24.
- Dong, G. & Li, J. (1999). Efficient mining of emerging patterns: Discovering trends and differences. In *KDD-99: Proceedings of the fifth international conference on knowledge discovery and data mining*. San Diego, CA.
- Han, J., & Kamber, M. (2001). *Data mining: Concepts and techniques*. USA: Morgan Kaufmann Publishers.
- Ian, H. W., & Eibe, F. (2000). *Data mining*. USA: Morgan Kaufmann Publishers.
- Jennings, D., & Lumpkin, J. (1992). Insights between environmental scanning activities and porter's generic strategies: An empirical analysis. *Journal of Management*, 18(4), 791–803.
- Kuo, R. J., Lin, S. Y., & Shih, C. W. (2007). Mining association rules through integration of clustering analysis and ant colony system for health insurance database in Taiwan. *Expert Systems with Applications*, 33(3), 794–808.
- Langari, Z. & Tompa, F.W. (2001). Subject classification in the oxford English dictionary. In *Proceedings of the IEEE international conference on data mining* (pp. 329–336). San Jose, CA.
- Lanquillon, C. (1999). Information filtering in changing domains. In *Proceedings of the international joint conference on artificial intelligence*. Stockholm, Sweden.
- Liu, B. & Hsu, W. (1996). Post-analysis of learned rules. In *Proceedings of the 13th national conference on artificial intelligence* (pp. 828–834). Menlo Park, California.
- Liu, B., Hau, W. & Ma, Y. (2001). Discovering the set of fundamental rule changes. In *Proceedings of the seventh ACM international conference on knowledge discovery and data mining* (pp. 335–340). San Francisco, California.
- Liu, B., Hsu, W., Han, H.-S. & Xia, Y. (2000). Mining changes for real-life applications. In *Proceedings of the second international conference on data warehousing and knowledge discovery* (pp. 337–346). London.
- Lu, C.-T., Kou, Y., Zhao, J., & Chen, Li. (2007). Detecting and tracking regional outliers in meteorological data. *Information Sciences*, 177(7), 1609–1632.
- Mayeux, P. (1996). *Broadcast news: Writing and reporting* (2nd ed.). Guilford, CT: Brown & Benchmark Publishers.
- Porter, M. F. (1980). An algorithm for suffix stripping. *Program*, 14(3), 130–137.
- Salton, G., & Buckley, C. (1988). Term weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5), 513–523.
- Song, H. S., Kim, J. K., & Kim, S. H. (2001). Mining the change of customer behavior in an internet shopping mall. *Expert Systems with Applications*, 21(3), 157–168.
- Wei, C.-P., & Lee, Y.-H. (2004). Event detection from online news documents for supporting environmental scanning. *Decision Support Systems*, 36(4), 385–401.
- Yang, Y., Carbonell, J. G., Brown, R. D., Pierce, T., Archibald, B. T., & Liu, X. (1999). Learning approaches for detecting and tracking news events. *IEEE Intelligent Systems*, 14(4), 32–43.
- Yang, Y., Pierce, T. & Carbonell, J.G. (1998). A study on retrospective and on-line event detection. In *Proceedings of the 21 annual international ACM SIGIR conference on research and development in information retrieval* (pp. 28–36). Melbourne, Australia: ACM Press.
- Yen, S.-J., & Lee, Y.-S. (2006). An efficient data mining approach for discovering interesting knowledge from customer transactions. *Expert Systems with Applications*, 30(4), 650–657.