# 2010 TOPCO 崇越論文大賞

論文題目

**Mining Customer Behavior Change Model for Fuzzy Quantitative Sequential Patterns**

報名編號：C0030

# Abstract

Fuzzy quantitative sequential pattern mining is one of the serviceable data-mining techniques to discover customer behavioral patterns over time and purchased quantities. A shopping example, <{(Beer, *Low*), (Milk, *High*)} (Cola, *Middle*)>, means that customers initially buy Beer and Milk in Low and High quantity and then purchase Cola in Middle quantity in the next trip, where *Low*, *Middle*, and *High* are predetermined linguistic terms given by managers. It reveals more general and concise knowledge for managers, which help them make quick-response decisions, especially in business.

However, no research, to our knowledge, has ever addressed the change issue of fuzzy quantitative sequential patterns. To approach the problem, we propose a novel change mining model, *FuzzChgMining*, to detect the change in fuzzy quantitative sequential patterns. Experiments are carried out by using secondary data collected from a retail chain in Taiwan, to evaluate the proposed model.

Some findings are as follows. The impact of three parameters (*min_sup*、$\beta$、$\phi_{min}$) on the number of fuzzy sequential patterns is significant. New shopping findings of three types of changed pattern have been formed during 2001 and 2002. Seasonal issue is a critical factor in consuming trends. And Health-care issue has raised people's attention in recent years.

***Keywords***: *data mining; change mining; sequential patterns; quantitative data; fuzzy sets*

# 1. Introduction

In response to the issue of the changing behaviors, adopting some practical tools to discover customer behavioral patterns is the primary task. As a result, many data-mining techniques proposed to discover useful information in customer behaviors and market trends have brought about some critical applications, such as product bundling (Yang and Lai 2006), RFM sequential patterns (Chen et al. 2009), product recommendation (Huang and Huang 2009, Kim et al. 2002), cross-selling (Lin et al. 2003), and fuzzy quantitative sequential patterns (Chen and Huang 2006). In the numerous data-mining techniques applied in business, mining sequential A supermarket shopping example of a sequential pattern is a customer who returns to buy Cola after buying Beer and Milk. This pattern, however, does not consider the purchased quantities associated with items. Therefore, Kim et al. (2004) proposed a model to discover quantitative sequential patterns. A quantitative sequential pattern may look as follows.

> After buying '10-15' Beers and '1-3' Milks, a customer at a supermarket will return to buy '5-8' Colas.

The above quantitative sequential pattern is informative, which reveals not only the order of items but also the range of their quantities.

Although quantitative sequential patterns can reveal items' quantity intervals, such approach may encounter a sharp boundary problem. For instance, let a quantity interval of $qty\text{-}int_1$ be $2 \leq qty < 4$ and that of $qty\text{-}int_2$ be $4 \leq qty < 6$, where $qty$ is the purchased quantity of an item. Then if the quantity is near 4, either a little larger or smaller, it is difficult to say whether the quantity is in $qty\text{-}int_1$ or in $qty\text{-}int_2$. Hence, the case can only be one hundred percent in $qty\text{-}int_1$ or in $qty\text{-}int_2$. This difficulty can be adequately tackled by using fuzzy techniques, since fuzzy set theory allows this quantity to be 50% in $qty\text{-}int_1$ and at the same time 50% in $qty\text{-}int_2$.

A fuzzy extension, called *fuzzy quantitative sequential patterns* (FQSP), was proposed by Hong et al. (1999) to find fuzzy purchased quantities with items in sequential patterns. A fuzzy quantitative sequential pattern has a form like

> Customers first buy Beer and Milk in High and Low quantities and then purchase Cola in Middle quantities in the next trip

The pattern can be represented by <{(Beer, *High*), (Milk, *Low*)} (Cola, *Middle*)> as well. This simple example indicates that the fuzzy concept is better than the partition method because fuzzy sets provide smooth transitions between members and non-members of a set.

As the above introduction, we know that there are several advantages to apply fuzzy quantitative sequential patterns in business. First, the knowledge of decision making can be shown in a more intuitive way by using fuzzy logic, which makes it easier for managers to understand the context of the mining results. Second, it is widely acknowledged that many real-life situations are intrinsically fuzzy. Third, using linguistic terms is simple and easy for managers. In the nature of executive work activities, managers often take care of strategic issues and long-term trends (Watson and Frolick 1993). Therefore, verbal communications with those linguistic terms are preferred for the exchange of soft knowledge.

Although applying the technique to mining FQSP is workable, it still not considers customer behavioral changes in the fast-changing business environment. For example, the fuzzy quantitative sequential pattern <{(Beer, *High*), (Milk, *Low*)} (Cola, *Middle*)> is available in the last year. The pattern, however, is not a trend in this year, substituted by <(Beer, *Low*), {(Cola, *High*) (Milk, *Middle*)}>. If managers cannot capture the customer behavioral change in time, two failed beliefs between the last year and this year will still exist in their mind, including:

(1) They still believe that a customer will buy Beer and Milk together before buying Cola. In fact, Milk is no longer to be purchased with Beer simultaneously yet is Cola.

(2) The purchased quantities of Beer, Milk, and Cola are not *High*, *Low*, and *Middle* but *Low*, *Middle*, and *High*.

Without renewing this knowledge, managers might map out the inappropriate marketing plans of products or services and dated inventory strategies with respect to time and quantities. No research, to our knowledge, has ever addressed the change issue of fuzzy quantitative sequential patterns. To address the research gap, we propose a novel change mining model, *FuzzChgMining*, to detect the FQSP change. The remainder of this paper is organized as follows. Section 2 reviews related works. Section 3 defines a similarity measurement for fuzzy quantitative sequential patterns. Section 4 presents the *FuzzChgMining* model for mining the change in fuzzy quantitative sequential patterns. Section 5 shows the experimental results by the proposed model. Conclusions are drawn in Section 6.

## 2. Related works

Previous works addressed the change problem are comparing different databases to discover the change patterns (Chen et al. 2005, Cho et al. 2005, Liu et al. 2000, Song et al. 2001, Tsai and Shieh 2009).

To recognize changes between different databases, Liu et al. (2000) devised an approach of change mining in the context of decision trees for predicting changes in customer behavior. Cho et al. (2005) proposed a new methodology for enhancing the quality of collaborative filtering recommendation that uses customer purchase sequences evolving over time. Song et al. (2001) developed a methodology which detects changes of customer behavior in association rules from databases automatically. They defined three types of changes as emerging pattern, unexpected change, and the added/perished rule and used similarity and different measures to detect them. Chen et al. (2005) integrated customer behavioral variables (recency, frequency, and monetary), demographic variables, and transaction databases to establish a method of mining changes in customer behavior. They also followed the definitions of the three types of changes proposed in Song et al. (2001) 0to mine change patterns. To detect changes in sequential patterns, Tsai and Shieh (2009), therefore, developed a change detection framework to observe the dynamic alternation of sequential patterns between two time-periods. They proposed new three types of changes in sequential patterns, whose idea was stemmed from those in Song et al. (2001) as well.

To the best of our knowledge, there are no researches to address the second way in fuzzy quantitative sequential patterns. Therefore, our work fills the research gap by proposing a change mining model for this type of knowledge. We also modify the definitions of emerging pattern, unexpected change, and the added/perished rule so that we can redefine our change patterns to detect fuzzy quantitative sequential patterns.

# 3. Problem statement and definitions

## *3.1. Fuzzy quantitative sequential pattern mining*

Suppose we have a universe of discourse $X$ in a quantity domain. A fuzzy set $A$ is characterized by a membership function $\mu_A(x)$, which assigns a membership degree between 0 and 1 to $x$, where $x \in X$. Suppose we are given a set of linguistic terms $LT=\{lt_j \mid j=1, 2, \ldots, l\}$. Then, the degree that a quantity value $q$ can be represented by linguistic term $lt_j$ (where $j=1, 2, \ldots, l$), is determined by the fuzzy membership function associated with $lt_j$.

**Example 1.** Suppose we want to represent a purchased quantity using three linguistic terms: *Low*($L$), *Middle*($M$), and *High*($H$). Their membership functions can be represented as shown in Fig 1. (Chen and Huang 2006).

$$\mu_{Low}(q)=\begin{cases} 1.0, & q \le 1 \\ \dfrac{5-q}{4}, & 1<q<5 \\ 0.0, & q \ge 5 \end{cases} \qquad \mu_{Middle}(q)=\begin{cases} 0.0, & \text{either } 1 \ge q \text{ or } q \ge 10 \\ \dfrac{q-1}{4}, & 1<q \le 5 \\ \dfrac{10-q}{5}, & 5<q<10 \end{cases}$$

$$\mu_{High}(q)=\begin{cases}0.0, & q\le 5\\ \dfrac{q-5}{5}, & 5<q<10\\ 1.0, & q\ge 10\end{cases}$$

Fig 1. The fuzzy membership functions for purchased quantity ($q$) concept.

Applying the above fuzzy functions, the quantity value 6 can be assigned as 0.0/*Low*+0.8/*Middle*+0.2/*High*.

Based on the research of Chen and Huang (2006), a fuzzy quantitative sequential sequence (*fq*-sequence) $\alpha$, can be represented as $\alpha$=<$a_1a_2\ldots a_n$>, where $a_j$ is a *fq*-itemset, for $1\le j\le n$. Here, a *fq*-itemset is a subset of *fq*-items, where *fq*-item is a pair (*it*, *lg*) for $it\in I$ and $lg\in LT$. Generally, $a_j(k)$ denotes the $k$th *fq*-item in $a_j$, $a_j(k).item$ denotes the item of $a_j(k)$, and $a_j(k).lg$ denotes the linguistic term of $a_j(k)$, for $1\le k\le m$. For brevity, the brackets can be ignored if a *fq*-itemset has only one *fq*-item. For instance, (*b*, *Low*) is a *fq*-item, {(*b*, *Low*)(*c*, *High*)} is a *fq*-itemset, and <{(*b*, *Low*)(*r*, *Middle*)} {(*r*, *Middle*)} {(*b*, *Low*)(*c*, *High*)}> is a *fq*-sequence with three *fq*-itemsets, where $a_1$={(*b*, *Low*)(*r*, *Middle*)}, $a_2$={(*r*, *Middle*)}, and $a_3$={(*b*, *Low*)(*c*, *High*)}. In $a_3$, we have $a_3(1)$=(*b*, *Low*) and $a_3(2)$=(*c*, *High*). Also, we have $a_3(1).item=b$, $a_3(1).lg=Low$, $a_3(2).item=c$, and $a_3(2).lg=High$.

Given a sequence database $S$ and a user-specified minimum support *min_sup*, a fuzzy quantitative sequence $\alpha$ is a fuzzy quantitative sequential pattern in $S$ if $support_S(\alpha)\ge min\_sup$. Therefore, the problem of fuzzy quantitative sequential pattern mining is to discover the complete set of fuzzy quantitative sequential patterns of which supports are more than or equal to *min_sup*. The total number of *fq*-items in a *fq*-sequence is the *length* of the sequence. A *fq*-sequence whose length is $k$ is referred to as a $k$-*fq*-sequence. A *fq*-sequential pattern whose length is $k$ is referred to as a $k$-*fq*-sequential pattern. The *position* of a *fq*-item in a *fq*-sequence is one plus the number of *fq*-items preceding it. For example, $s$=<{(*b*, 2)(*r*, 5)} {(*r*, 6)} {(*b*, 4)(*c*, 7)}> is a 5-*fq*-sequence and a *fq*-sequence $\alpha$=<{(*c*, *Low*)(*r*, *Middle*)} {(*r*, *Middle*)} {(*b*, *Low*)(*c*, *High*)(*d*, *High*)}> is a 6-*fq*-sequential pattern. The positions of (*b*, 2), (*r*, 6), and ($c$, 7) in $s$ are 1, 3, and 5. The position of ($c$, *Low*), ($b$, *Low*), and ($d$, *High*) in $\alpha$ are 1, 4, and 6.

To find fuzzy quantitative sequential patterns, Hong et al. (1999) and Chen and Huang (2006) proposed two algorithms, *Hong et al.* and *divide-and-conquer fuzzy sequential mining* (DFSM) algorithms, to find patterns. Since the latter performance is more efficient than the former one in their experimental results (Chen and Huang 2006), we decided to adopt the DFSM algorithm in this study.

## *3.2 The fuzzy quantitative sequential pattern matching method*

The Similarity Computation Index ($SCI_{ij}$) computation could be shown in a very simple formula. However, it contains several similarity computation parts to construct the whole formula. The $SCI_{ij}$ contains $SI_{ijn}$, and $SI_{ijn}$ consists of $SD_{ijn}^{item}$ and $SD_{ijn}^{QLT}$, and all of which has different computation rules. Therefore, we try to make it easy by breaking them into some pieces, and then explain one after another with simple instances.

$$
\begin{cases}
SD_{ijn}^{item} = \dfrac{NumofComfq\_Items(s_i^t, s_j^{t+l})_n}{Max(\left|s_i^t\right|_{item}^n, \left|s_j^{t+l}\right|_{item}^n)} \text{.........................................................} Formula(1) \\[2ex]
SD_{ijn}^{QLT} = \dfrac{SimOfQLT(ComItem)_n}{Max(\left|s_i^t\right|_{QLT}^n, \left|s_j^{t+l}\right|_{QLT}^n)} \text{.................................................................} Formula(2) \\[2ex]
SI_{ijn} = \beta \times SD_{ijn}^{item} + (1-\beta) \times SD_{ijn}^{QLT} \text{...............................................................} Formula(3) \\[2ex]
SCI_{ij} = \dfrac{\sum_{n=1}^{k}\left(\dfrac{k-n+1}{k} \times SI_{ijn}\right)}{\sum_{n=1}^{k}\left(\dfrac{k-n+1}{k}\right)} \text{......................................................................} Formula(4)
\end{cases}
$$

Fig 2.The Similarity Computation Index ($SCI_{ij}$) Formulas

In Formula (1), $SD_{ijn}^{item}$ is the $fq$-item similarity degree of $fq$-itemset $n$ between $s_i^t$ and $s_j^{t+l}$, where $\left|s_i^t\right|_{item}^n$ is the number of $fq$-items of $fq$-itemset $n$ in $fq$-sequence $s_i^t$, representing the length of *fq*-itemsets $n$ in *fq*-sequence $s_i^t$, and $\left|s_j^{t+l}\right|_{item}^{n}$ is the number of *fq*-items of *fq*-itemset $n$ in *fq*-sequence $s_i^{t+l}$, representing the length of *fq*-itemsets $n$ in *fq*-sequence $s_i^{t+l}$. So $Max(\left|s_i^t\right|_{item}^{n}, \left|s_j^{t+l}\right|_{item}^{n})$ is the maximal length of *fq*-itemset $n$ between $s_i^t$ and $s_j^{t+l}$. Besides, $NumofComfq\_Items(s_i^t, s_j^{t+l})_n$ is the number of matching *fq*-items of *fq*-itemset n between $s_i^t$ and $s_j^{t+l}$.

In Formula (2), $SD_{ijn}^{QLT}$ is QLT similarity of *fq*-itemset $n$ between $s_i^t$ and $s_j^{t+l}$, where $SimOfQLT(ComItem)_n$ is the aggregating QLT similarity of common *fq*-items in *fq*-itemset $n$, $\left|s_i^t\right|_{QLT}^{n}$ is the number (length) of QLTs in *fq*-sequence $s_i^t$, and $\left|s_j^{t+l}\right|_{QLT}^{n}$ is the length of QLTs in *fq*-sequence $s_i^{t+l}$. Then $Max(\left|s_i^t\right|_{QLT}^{n}, \left|s_j^{t+l}\right|_{QLT}^{n})$ is maximal QLT length of *fq*-itemset $n$ between $s_i^t$ and $s_j^{t+l}$. In fact, $\left|s_i^t\right|_{QLT}^{n} = \left|s_i^t\right|_{item}^{n}$ and $\left|s_j^{t+l}\right|_{QLT}^{n} = \left|s_j^{t+l}\right|_{item}^{n}$ because *fq*-item and QLT exist in pairs in one *fq*-item, e.g. (b, Middle) is one *fq*-item. By the same rule, $Max(\left|s_i^t\right|_{QLT}^{n}, \left|s_j^{t+l}\right|_{QLT}^{n})$ is equal to $Max(\left|s_i^t\right|_{item}^{n}, \left|s_j^{t+l}\right|_{item}^{n})$. But we still show them in different ways to emphasize the disparity between *fq*-item similarity and QLT similarity. Therefore, in the algorithm implementation, we do not calculate them repeatedly.

In Formula (3), $SI_{ijn}$ is the similarity degree of *fq*-itemset n between $s_i^t$ and $s_j^{t+l}$, containing the aggregated and weighted similarity of *fq*-items and quantitative linguistic terms in *fq*-itemset $n$. It is presented by $\beta \times SD_{ijn}^{item} + (1-\beta) \times SD_{ijn}^{QLT}$, where $\beta$ is the weighted argument from 0 to 1.

Finally, in Formula (4), we can get the final similarity degree between $s_i^t$ and $s_j^{t+l}$ with the computation of $SCI_{ij} = \dfrac{\sum_{n=1}^{k}\left(\dfrac{k-n+1}{k} \times SI_{ijn}\right)}{\sum_{n=1}^{k}\left(\dfrac{k-n+1}{k}\right)}$, where $k$ is the maximal *fq*-sequence length between $s_i^t$ and $s_j^{t+1}$, and $n$ is the ordinal number of *fq*-itemsets in $s_i^t$

and $s_j^{t+l}$. Thus, $\left(\frac{k-n+1}{k}\right)$ implies the weighted index of *fq*-itemset *n*.

## *3.3 An illustrative example*

In this section, we would like to use some examples to demonstrate how the matching method works when computing the similarity. Given two fuzzy quantitative sequential pattern sets from time *t* and time *t+l* in Table 1, we enlarge on the whole process of similarity computation between certain two patterns by using formulas from (1) to (4) pointed out in Section 3.2, and then find out the changes. We discover two patterns, $s_1^t$ and $s_2^t$, at time *t* and five patterns, $s_1^{t+l}$, $s_2^{t+l}$, $s_3^{t+l}$, $s_4^{t+l}$, and $s_5^{t+l}$, at time *t+l* respectively, and we use the patterns in time *t* as base period to compute the similarities with their corresponding patterns in time *t+l*. We let β as 0.6 and set up a quantitative linguistic-term Matrix in Table 2; therefore, the similarities between $s_1^t$ and $s_1^{t+l}$, $s_2^{t+l}$, $s_3^{t+l}$, $s_4^{t+l}$, and $s_5^{t+l}$ are computed as follows.

We first illustrate the detailed process of similarity computation for $SCI_{11}$ and $SCI_{12}$, and the ensuing *SCIs* will follow the same calculating rules of Formula (1)~(4).

**1. $\mathbf{SCI_{11}}$:**

$S_1^t$:<{(f, M)(k, H)}{(c, M)(h, L)(m, L)}{(b, L)(d, M)(e, M)}{(a, L)(g, H)}>

$S_1^{t+l}$:<{(f,M)(k,H)}{(c,H)(h,L)(m,L)}{(b,M)(d,M)(e,L)}{(a,L)(g,M)}>

(1) *fq*-itemset 1

$$SD_{111}^{item} = \frac{2}{Max(2,2)} = \frac{2}{2} = 1 \quad , \quad SD_{111}^{QLT} = \frac{1+1}{Max(2,2)} = \frac{2}{2} = 1$$

$$SI_{111} = 0.6\times 1 + 0.4\times 1 = 1$$

(2) *fq*-itemset 2

$$SD_{112}^{item} = \frac{3}{Max(3,3)} = \frac{3}{3} = 1, \quad SD_{112}^{QLT} = \frac{0.66+1+1}{Max(3,3)} = \frac{2.66}{3} = 0.89$$

$$SI_{112} = 0.6\times 1 + 0.4\times 0.89 = 0.96$$

(3) *fq*-itemset 3

$$SD_{113}^{item}=\frac{3}{Max(3,3)}=\frac{3}{3}=1,\quad SD_{113}^{QLT}=\frac{0.66+1+0.66}{Max(3,3)}=\frac{2.32}{3}=0.77$$

$$SI_{113}=0.6\times 1+0.4\times 0.77=0.91$$

(4) *fq*-itemset 4

$$SD_{114}^{item}=\frac{2}{Max(2,2)}=\frac{2}{2}=1,\quad SD_{114}^{QLT}=\frac{1+0.66}{Max(2,2)}=\frac{1.66}{2}=0.83$$

$$SI_{114}=0.6\times 1+0.4\times 0.83=0.93$$

$$SCI_{11}=\frac{\frac{4-1+1}{4}\times 1+\frac{4-2+1}{4}\times 0.96+\frac{4-3+1}{4}\times 0.91+\frac{4-4+1}{4}\times 0.93}{\frac{4-1+1}{4}+\frac{4-2+1}{4}+\frac{4-3+1}{4}+\frac{4-4+1}{4}}=0.96$$

2. **SCI$_{12}$:**

$S_1^t$:<{(f, M)(k, H)}{(c, M)(h, L)(m, L)}{(b, L)(d, M)(e, M)}{(a, L)(g, H)}>

$S_2^{t+l}$:<{(b,M)(f,H)(h,H)}{(a,M)(m,H)(k,L)}{(e,M)}{(d,L)(g,H)}{(g,M)}{(j,M)}{(p,M)}>

$$SCI_{12}=\frac{\frac{7-1+1}{7}\times 0.29+\frac{7-2+1}{7}\times 0.24+\frac{7-3+1}{7}\times 0.33+\frac{7-4+1}{7}\times 0.5+\frac{7-5+1}{7}\times 0+\frac{7-6+1}{7}\times 0+\frac{7-7+1}{7}\times 0}{\frac{7-1+1}{7}+\frac{7-2+1}{7}+\frac{7-3+1}{7}+\frac{7-4+1}{7}+\frac{7-5+1}{7}+\frac{7-6+1}{7}+\frac{7-7+1}{7}}=0.25$$

3. **SCI$_{13}$**=(4/4×0.93+3/4×0.86+2/4×0.91+1/4×0.93)/ (4/4+3/4+2/4+1/4)=0.91
4. **SCI$_{14}$**=(4/4×0.62+3/4×0.53+2/4×0.96+1/4×0.86)/ (4/4+3/4+2/4+1/4)=0.69
5. **SCI$_{15}$**=(4/4×0.87+3/4×1+2/4×0.96+1/4×0.8)/ (4/4+3/4+2/4+1/4)=0.92

In the end, we could get $MaxSCI_i^t$=max (0.96, 0.25, 0.91, 0.69, 0.92) = 0.96 and realize the truth that pattern $s_1^{t+l}$ is most similar to pattern $s_1^t$ with highest similarity (0.96) among others. By the same token, the process of similarity computation between $s_2^t$ and $s_1^{t+l}$, $s_2^{t+l}$, $s_3^{t+l}$, $s_4^{t+l}$, and $s_5^{t+l}$ are calculated as follows.

1. **SCI$_{21}$=**(5/5×0.29+4/5×0.24+3/5×0.62+2/5×0.43+1/5×0)/(5/5+4/5+3/5+2/5+1/5)=0.34
2. **SCI$_{22}$=**7/7×0.33+6/7×0.33+5/7×0.5+4/7×0.37+3/7×0.33+2/7×0+1/7×0)/( 7/7+6/7+ 7+4/7+3/7+2/7+1/7)=0.33
3. **SCI$_{23}$=**5/5×0.33+4/5×0.33+3/5×0.58+2/5×0.43+1/5×0)/(5/5+4/5+3/5+2/5+1/5)=0.37

4. **$SCI_{24}$**=5/5×0.58+4/5×0.62+3/5×0.62+2/5×0.43+1/5×0)/(5/5+4/5+3/5+2/5+1/5)=0.54

5. **$SCI_{25}$**=(5/5×0.29+4/5×0.24+3/5×0.58+2/5×0.5+1/5×0)/(5/5+4/5+3/5+2/5+1/5)=0.34

In the end, we could get $MaxSCI_2^t$=max(0.34, 0.33, 0.37, 0.54, 0.34)= 0.54 and realize the truth that pattern $s_4^{t+l}$ is most similar to pattern $s_2^t$ with highest similarity (0.54) among others. Since the $SCI_j^{t+1}$ traceback processes for items and quantitative linguistic terms (QLT) are similar to those of $SCI_i^t$ , we would not illustrate the detailed traceback processes in this section.

Table 1 Two pattern sets from time-periods *t* and *t + l*

| Time-period | Pattern Set | Fuzzy Quantitative Sequential Pattern | Support |
|---|---|---|---|
| *t* | $PS^t$ | $S_1^t$:<{(f, M)(k, H)}{(c, M)(h, L)(m, L)}{(b, L)(d, M)(e, M)}{(a, L)(g, H)}> | 3% |
| | | $S_2^t$:<{(a,L)(f,H)(g,H)}{(b,M)(m,H)}{(b,M)(e,M)}{(g,L)}{(e,M)(f,M)(g,M)}> | 5% |
| *t + l* | $PS^{t+l}$ | $S_1^{t+l}$:<{(f,M)(k,H)}{(c,H)(h,L)(m,L)}{(b,M)(d,M)(e,L)}{(a,L)(g,M)}> | 6% |
| | | $S_2^{t+l}$:<{(b,M)(f,H)(h,H)}{(a,M)(m,H)(k,L)}{(e,M)}{(d,L)(g,H)}{(g,M)}{(j,M)}{(p,M)}> | 13% |
| | | $S_3^{t+l}$: <{(f, H)(k, H)}{(c, M)(h, M)(m, H)}{(b, L)(d, H)(e, H)}{(a, L)(g, M)}> | 8% |
| | | $S_4^{t+l}$:<{(a,M)(f,M)(k,M)}{(b,H)(h,M)(m,H)}{(b,L)(d,H)(e,M)}{(a,M)(g,M)}> | 15% |
| | | $S_5^{t+l}$: <{(f, M)(k, L)}{(c, M)(h, L)(m, L)}{(b, L)(d, M)(e, L)}{(a, M)(g, L)}> | 10% |

*Low is denoted as *L,* Middle is denoted as *M,* High is denoted as *H*

Table 2 The Quantitative Linguistic-term Matrix

| Quantitative Linguistic term | *Low* | *Middle* | *High* | *Null* |
|---|---|---|---|---|
| *Low* | 1 | 0.66 | 0.33 | 0 |
| *Middle* | 0.66 | 1 | 0.66 | 0 |
| *High* | 0.33 | 0.66 | 1 | 0 |
| *Null* | 0 | 0 | 0 | |

# 4. Mining the change of fuzzy quantitative sequential pattern

In this section, we entitle a model "*FuzzChgMining*" to mine the change of customer behavior in fuzzy quantitative sequential patterns. It contains three continuous phases shown in Fig.5. In Phase 1, $PS^t$ and $PS^{t+l}$ are two fuzzy quantitative sequential pattern sets from databases at time $t$ and time $t+l$ respectively. The relevant and detailed process of mining fuzzy quantitative sequential patterns can be found in Chen and Huang (2006). In Phase 2, we evaluate the matching degree of any two patterns from $PS^t$ and $PS^{t+1}$ by computing their common similarity with formulas (1), (2), (3), and (4). Then, we classify them into three different types of changed patterns according to specific rule, including emerging patterns (quantity change), unexpected change, and added/perished patterns. A large number of changed patterns might baffle the decision maker, thus various patterns found in Phase 2 should be evaluated in accordance with the significance of each type of changed pattern in Phase 3. In other words, we only pick on the changed patterns which have a significant change degree.

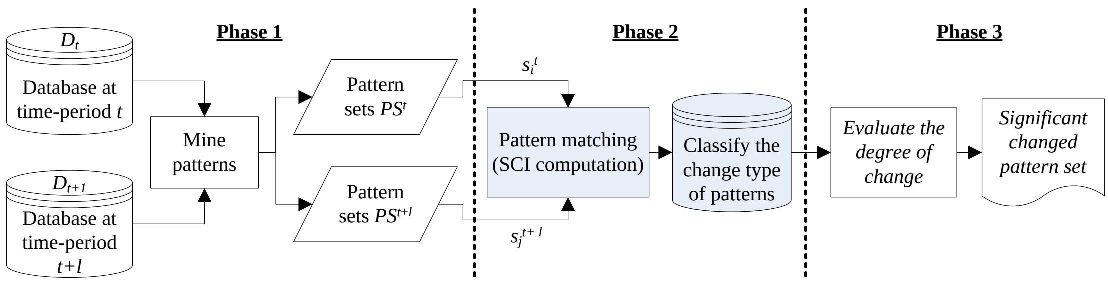


Fig 3. The proposed model, *FuzzChgMining*.

To distinguish the difference among these three types of changed pattern, we need to define them one after another by computing their similarities with formula (1), (2), (3) and (4). We define a threshold, *Pattern Matching Threshold (PMT)* to provide each user or manager a subjective method to judge the types of changed pattern. *PMT* is determined by each user or manager, and then can be different from time to time. The idea of *PMT* is shown in Fig.7, and reveals the way how the customer behavioral changes be differentiated by *PMT*.

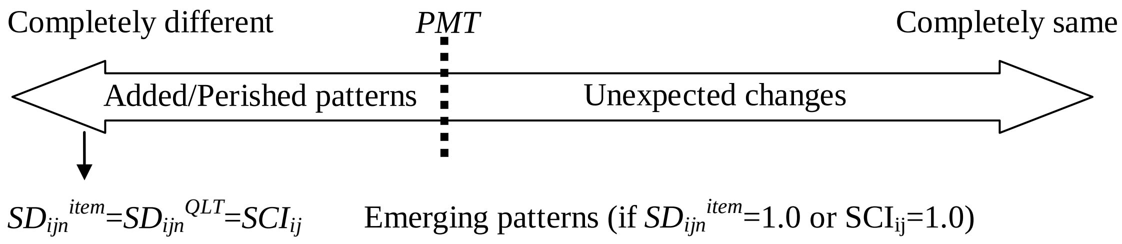


Fig 4. The relationships of three different changed types.

In retail shops, some customer behavioral patterns found in the past database are not significant, but obvious in present one. In this case, we would call them as emerging patterns in this study, which was defined by Dong and Li (1999). It also implies that the patterns' supports are increasing significantly from one database to another, indicting some emerging trends now exist in customer behavior. We define emerging patterns as follows.

**Definition 1** (Emerging fuzzy quantitative sequential patterns) The fuzzy quantitative sequential pattern $s_j^{t+l}$ is defined as an emerging pattern with regard to pattern $s_i^t$ if $SD_{ijn}^{item}$=1.0 or $SCI_{ij}$=1.0, which is derived from formula (1) and formula (4) in Fig.2, and the supports of the two patterns from different pattern sets at time point $t$ and $t+l$ are significantly different.

**Definition 1.1** (Quantity change)Based on Definition 1, three quantity-changed emerging patterns can be recognized by judging the changing degree of quantity. Here, we define $com\_qpt(s_j^{t+l})$ as the comparative quantity point of a pattern at time $t+l$ compared with patterns at time $t$, *and the comparative quantity table is presented in Table 3*. From example, the quantities of pattern $s_i^{\,t}$ are *Low*, *Low*, and *Middle*; however, the quantities of $s_j^{t+l}$ are *High*, *Low*, and *Low*. Thus, the comparative quantity points accumulated from time $t$ to time $t+1$ are (+2+0-1)/3=0.33. Based on the comparative quantity points computing rules, we can tell apart those patterns quickly and correctly.

When the quantities of pattern $s_j^{t+l}$ are relatively on the rise compared with those of pattern $s_i^t$, we call it a *Qty-increased emerging pattern*, where *Qty* refers to quantity. The phrase "on the rise" indicates that the comparative quantity points accumulated in

$s_j^{t+l}$, $com\_qpt(s_j^{t+l})$, is larger than 0. If the fuzzy quantities of pattern $s_j^{t+l}$ are equal to those of pattern $s_i^t$, we call it a *Qty-stable emerging pattern*. The term "equal" implies that the comparative quantity points got in $s_j^{t+l}$, $com\_qpt(s_j^{t+l})$, is equal to 0. Finally, if the fuzzy quantities of pattern $s_j^{t+l}$ are relatively fewer compared to those of pattern $s_i^t$, we call it a *Qty-decreased emerging pattern*. The term "fewer" means that the comparative quantity points turned out in $s_j^{t+l}$, $com\_qpt(s_j^{t+l})$, is smaller than 0.

**Example 2** According the examples in Table 1, we scrutinize a pattern $s_1^{t+l}$ to see whether it is an emerging pattern to any patterns in $PS^t$. We say that $s_1^{t+l}$*:<{(f,M)(k,H)}{(c,H)(h,L)(m,L)}{(b,M)(d,M)(e,L)}{(a,L)(g,M)}>* with support of 6% is an emerging pattern with respect to pattern $s_1^t$*:<{(f, M)(k, H)}{(c, M)(h, L)(m, L)}{(b, L)(d, M)(e, M)}{(a, L)(g, H)}>* because $SD_{111}^{item}$, $SD_{112}^{item}$ , $SD_{113}^{item}$, $SD_{114}^{item}$ are all equal to 1.0, and the supports of $s_1^{t+l}$ and $s_1^t$ are different.

**Example 2.1** Following Example 2, $s_1^{t+l}$ is a Qty-increased emerging pattern with support of 6% because:

(1) Given a comparative quantity table in Table 5;

(2) For *fq*-itemset 1 in $s_1^{t+l}$, *Middle*($t+l$) and *High*($t+l$), to be compared with *fq*-itemset 1 in $s_1^t$, *Middle*($t$) and *High*($t$), we get a score of 0 (+0+0);

(3) For *fq*-itemset 2 in $s_1^{t+l}$, *High*($t+l$), *Low(t+l)* and *Low*($t+l$), to be compared with *fq*-itemset 2 in $s_1^t$, *Middle*($t$), *Low(t)* and *Low*($t$), we get a score of 1 (+1+0+0);

(4) For *fq*-itemset 3 in $s_1^{t+l}$, *Middle* ($t+l$), *Middle (t+l)* and *Low*($t+l$), to be compared with *fq*-itemset 3 in $s_1^t$, *Low* ($t$), *Middle (t)* and *Middle* ($t$), we get a score of 0 (+1+0-1);

(5) For *fq*-itemset 4 in $s_1^{t+l}$, *Low*($t+l$) and *Middle (t+l)*, to be compared with *fq*-itemset 4 in $s_1^t$, *Low* ($t$), and *High* ($t$), we get a score of -1 (+0-1);

(6) As a result, we have $com\_qpt(s_1^{t+l})$=(+0+1+0-1)/10=0, which is equal to 0 so that we call it an Q-stable emerging pattern.

**Example 3.2** $S_3^{t+l}$= *<{(f, H)(k, H)}{(c, M)(h, M)(m, H)}{(b, L)(d, H)(e, H)}{(a, L)(g, M)}>* with support of 8% is a Qty-increased emerging pattern of $s_1^t$ because its com_qpt($s_1^{t+l}$)=(+1+3+2-1)/10=0.5 is larger than 0.

**Example 3.3** $S_5^{t+l}$= <{(f, M)(k, L)}{(c, M)(h, L)(m, L)}{(b, L)(d, M)(e, L)}{(a, M)(g, L)}> with support of 10% is a Qty-decreased emerging pattern of $s_1^t$ because its com_qpt($s_1^{t+l}$)=(-2+0-1-1)/10=-0.25 is smaller than 0.

Table 3 The comparative quantity table

| *t+l* \ *t* | *Low* | *Middle* | *High* |
|---|---|---|---|
| *Low* | +0 | -1 | -2 |
| *Middle* | +1 | +0 | -1 |
| *High* | +2 | +1 | +0 |

In fact, consumer behavior is sometimes prone to be fickle and changeful, which baffles managers all the time. It also brings about misinterpretation of consumer cognition and false decipherment of consumer affection. Therefore, we want to introduce another type of changed pattern, *unexpected changes*, originating from the study by Padmanabhan and Tuzhilin (1999). *Unexpected change* literally means that the patterns found in the present database are beyond people's expectation. Manager, in this case, cannot deal with these unexpected changes timely and appropriately. To avert the situation like this, we have to find unexpected changes in patterns from different databases. *Unexpected changes* are defined as followings.

**Definition 2** (Unexpected changes) Fuzzy quantitative sequential pattern $s_j^{t+l}$ is defined as an unexpected pattern change with respect to $s_i^t$ if $MinSCI_j^{t+l}>PMT$ and $SCI_{ji}>0$, where $MinSCI_j^{t+l}=\min(SCI_{j1}, SCI_{j2}, \ldots, SCI_{j|S^t|})$.

**Example 4.** Following the example in Table 3, we set *PMT* as 0.345, and then $s_4^{t+l}$: *<{(a,M)(f,M)(k,M)}{(b,H)(h,M)(m,H)}{(b,L)(d,H)(e,M)}{(a,M)(g,M)}>* with support of **15%** is an unexpected pattern with respect to $s_1^t$ : *<{(f, M)(k, H)}{(c, M)(h, L)(m, L)}{(b, L)(d, M)(e, M)}{(a, L)(g, H)}>*in $PS^t$ because $MinSCI_4^{t+l}= Min(SCI_{41,} SCI_{42})$= *Min*(0.69, 0.54)=0.5395 is larger than *PMT* (0.345) and $SCI_{41}^{t+l}$=0.69 is larger than 0. By the same token, $s_4^{t+l}$ is an unexpected pattern with respect to $s_2^t$ :*<{(a,L)(f,H)(g,H)}{(b,M)(m,H)}{(b,M)(e,M)}{(g,L)}{(e,M)(f,M)(g,M)}>* because $SCI_{42}^{t+l}$=0.54 is larger than 0. The pattern, $s_4^{t+l}$, is a new trend that goes beyond the original knowledge of managers.

Because the change of customer behavior in online shopping varies dramatically, we should first measure the variation of patterns and then detect the differences between two databases. The variation happens when the patterns exist at present time but cannot to be found in the past, or the patterns that can be found in the past but disappear at present. Those special changed patterns are defined as *added patterns* and *perished patterns* respectively by Lanquillon (1999). Their pattern structures are entirely novel and unique compared with any patterns from pattern sets at any different time. We define the added/perished patterns as follows.

**Definition 3.** (Added/Perished fuzzy quantitative sequential patterns) The fuzzy quantitative sequential pattern $s_j^{t+l}$ is defined as an added pattern with respect to all patterns discovered in $PS^t$ if $MaxSCI_j^{t+l} \leq PMT$ and $SCI_{ji}>0$. The fuzzy quantitative sequential pattern $s_i^t$ is defined as a perished pattern with respect to all patterns discovered in $S^{t+l}$ if $MaxSCI_i^t \leq PMT$ and $SCI_{ij}>0$.

**Example 5.** Following the example in Table 3, we say that $s_2^{t+l}$: *<{(b,M)(f,H)(h,H)}{(a,M)(m,H)(k,L)}{(e,M)}{(d,L)(g,H)}{(g,M)}{(j,M)}{(p,M)}>* with support of **13%** is an added pattern with respect to patterns in $PS^t$ because $MaxSCI_2^{t+l}= Max(SCI_{21}, SCI_{22})= Max$ (0.26, 0.33)=0.33 is smaller than $PMT$ (0.345) and $SCI_{21}^{t+l}$=0.26 and $SCI_{22}^{t+l}$=0.33 are both larger than 0. The pattern is necessary for managers to append new beliefs in their knowledge base.

**Example 5.1.** Assume we have a new sequential pattern $s_3^t$: *<{(k,H)(j,H)}{(b,L)(m,M)(r,H)}>* with support value **10%** from $PS^t$. We say that $s_3^t$ is a perished pattern with respect to patterns in $PS^{t+l}$ because $MaxSCI_3^t= Max(SCI_{31}^t, SCI_{32}^t, SCI_{33}^t, SCI_{34}^t, SCI_{35}^t)= Max$(0.29, 0.06, 0.29, 0.27, 0.23)=0.29 is smaller than $PMT$ (0.345) and $SCI_{31}^t$, $SCI_{32}^t$, $SCI_{33}^t$, $SCI_{34}^t$, and $SCI_{35}^t$ are all larger than 0. Managers can remove out-of-date beliefs from their original knowledge base by using these types of patterns.

According to the foregoing rules, we can easily categorize all customer behavior into three types of change patterns, including emerging pattern, unexpected change and added/perished pattern. However, not all of change patterns are worth our attention. To help manager make decision more effectively, we filter out trivial change patterns by setting an user-specified minimum threshold $\psi_{min}$ and only keep significant ones. In this case, managers can judge and predict the market trends correctly with organized mining information.

$\theta_{ij}$ ($\theta_{ji}$) is an index set to measure the significance between $s_i^t$ and $s_j^{t+l}$. To calculate the significance of an emerging pattern, we define support changing ratio$\theta_{ij}$ as $\frac{\text{support}\left(s_j^{t+l}\right) - \text{support}\left(s_i^t\right)}{\text{support}\left(s_i^t\right)}$, where sup ($s_i^t$) and sup ($s_j^{t+l}$) are the supports of $s_i^t$ in $S^t$ and $s_j^{t+l}$ in $S^{t+l}$, respectively. To compute the significance of an unexpected change, we have the support changing ratio multiplied by a weight ($SCI_{ji}$); thus $\theta_{ji}$ is defined as $SCI_{ji} \times \frac{\text{support}\left(s_j^{t+l}\right)}{\text{support}\left(s_i^t\right)}$, where $SCI_{ji}$ is the similarity between $s_i^t$ and $s_j^{t+l}$. To obtain the significance of an added pattern $\theta_{ji}$, we use $MinSCI_j^{t+l}$ as a weight to multiply its support, sup($s_j^{t+l}$), i.e. $(MinSCI_j^{t+l} \times 100) \times \text{support}\left(s_j^{t+l}\right)$. In the end, the significance of an added pattern$\theta_{ij}$ is denoted as follows, we use $MinSCI_i^t$, a weight, to multiply its support, sup($s_i^t$), i.e. $MinSCI_i^t \times \text{support}\left(s_i^t\right)$. All the above computations are summarized as follows:

$$\theta_{ij}(\theta_{ji}) = \begin{cases} \frac{\text{support}\left(s_j^{t+l}\right) - \text{support}\left(s_i^t\right)}{\text{support}\left(s_i^t\right)} & \text{, emerging pattern} \\ SCI_{ji} \times \frac{\text{support}\left(s_j^{t+l}\right)}{\text{support}\left(s_i^t\right)} & \text{, unexpected change} \\ (MinSCI_j^{t+l} \times 100) \times \text{support}\left(s_j^{t+l}\right) & \text{, added pattern} \\ (MinSCI_i^t \times 100) \times \text{support}\left(s_i^t\right) & \text{, perished pattern} \end{cases}$$

$\theta_{ij}$ for emerging and perished pattern, $\theta_{ji}$ for unexpected and added pattern.

Fig 5. The significance of four different changed types

**Example 6.** Following Example 6.1, 6.1 and 6.3, $s_1^{t+l}$, $s_3^{t+l}$ and $s_5^{t+l}$ are emerging patterns, and the significance are $\theta_{11}$ is equal to 1=(6-3)/3, $\theta_{13}$ is equal to 1.67=(8-3)/3 and$\theta_{15}$ is equal to 2.33=(10-3)/3. Following Example 7, $s_4^{t+l}$ is an unexpected pattern, and its $\theta_{41}$ is equal to 3.45=0.69×(15/3), and$\theta_{42}$ is equal to 1.62=0.54×(15/5). Following Example 8, $s_2^{t+l}$ is an added pattern, and its $\theta_{21}$ is equal to 3.38=(0.26×100)×13%. Following Example 8.1, $s_3^{t}$ is a perished pattern, and its $\theta_{31}$ is equal to 0.6=(0.06×100)×10%. All change patterns are significant when $\psi_{min}$ is specified as 0.5.

# 5. Experimental results

In this section, we use a real dataset to study the effectiveness of the *FuzzChgMining* model. We present the results of the experiments and show some interesting patterns to discuss their business implications.

The secondary data is collected from five branches of the *Songchine* supermarket in Taiwan, which sold groceries for daily use from 2001/12/28 to 2002/12/28. After the data pre-processing, there were 10,798 members considered, and 100,848 customer data-sequences, with an average of 3.0 items per data-sequence. There were a total of 18,162 items (products). Since those items had the potential to cause excessively trivial patterns, affecting decision making, we refer to the idea of fuzzy taxonomies (Chen and Huang 2008) to find frequent patterns at their upper level categories. For example, the upper level of the items including skim milk, whole milk, and apple milk, is milk. Finally, we had a total of 603 categories, which were used as items in the experiment.

According to the time-period in *Songchine*, we divided the dataset into two sub-datasets, generating two disjoined time-periods. The time-period of the first sub-dataset is from December 28, 2001 to June 28, 2002, referred to as time-period 1 (TP1). The second is from June 29, 2002 to December 28, 2002, referred to as time-period 2 (TP2). As a result, TP1 contains 56,338 data-sequences and TP2 contains 44,510. In this experiment, we used five linguistic terms to represent the quantities, including: *Very Low*, *Low*, *Middle*, *High*, and *Very High*. The detailed construction of linguistic terms can be referred to in Chen and Huang (2006). Next, we conducted four tests with different parameters to show the experimental results. Table 4lists the parameters used in the experiment and Table 5shows the variations in the four tests.

Table 4 Parameters

| | |
|---|---|
| *min_sup* | A user-specified minimum support threshold |
| $\beta$ | A weighting argument in *SSCI* for items |
| $(1-\beta)$ | A weighting argument in *SSCI* for linguistic terms |
| $\psi_{min}$ | A user-specified minimum threshold for identifying the significant change patterns |

Table 5 Four tests

| Name | $\beta$ | $(1-\beta)$ | $\psi_{min}$ |
|---|---|---|---|
| B0.65-P0.3 | 0.65 | 0.35 | 0.3 |
| B0.55-P0.3 | 0.55 | 0.45 | 0.3 |
| B0.65-P0.6 | 0.65 | 0.35 | 0.6 |
| B0.55-P0.6 | 0.55 | 0.45 | 0.6 |

First of all, we employed the DFSM algorithm (Chen and Huang 2006) to generate the complete sets of fuzzy quantitative sequential patterns, where the minimum support threshold, *min_sup*, varied from 1.5% to 0.75%. Table 6summarizes the number of fuzzy quantitative sequential patterns for two time-periods, TP1 and TP2. The results indicate that as we decreased the *min_sup*, the number of patterns increased. The reason is that the lower minimum support threshold generated a huge set of candidate patterns so that more frequent patterns could pass it easily to be generated. Our result also matches the experimental results in Chen and Huang (2006).

Table 6The number of fuzzy quantitative sequential patterns for two time-periods

| *min_sup* | December 28, 2001 ~ June 28, 2002 (TP1) | June 29, 2002 ~ December 28, 2002 (TP2) |
|---|---|---|

| | | |
|---|---|---|
| 1.5% | 416 | 381 |
| 1.25% | 726 | 674 |
| 1% | 1,139 | 1,023 |
| 0.75% | 1,693 | 1,550 |

Second, we present the results for the number of three types of change patterns (Emerging, Unexpected, and Added/Perished) and the number of three types of significant change patterns (S_Emerging, S_Unexpected, and S_Added/S_Perished). The results of B0.65-P0.3, B0.55-P0.3, B0.65-P0.6, and B0.55-P0.6 are shown in Fig. 9, which uses a logarithmic scale for the Y-axis. Their minimum support thresholds vary from 1.5% to 0.75%. Besides, Fig. 13(a) and Fig. 13(b) show the number of emerging patterns for (significant) quantity trend change in B0.65-P0.3 (or B0.55-P0.3) and B0.65-P0.6 (or B0.55-P0.6).

As observed, we have four findings about the experimental results and discuss them as follows. (1) We found that the number of change patterns for unexpected and added/perished patterns increase when *min_sup* is decreased, whereas emerging patterns do not. The reason is that when *min_sup* is decreased, more candidate patterns exceeding the minimum support threshold can be generated; therefore, the combination of frequent patterns to generate unexpected patterns and added/perished patterns increases as well. Also, the frequent patterns, with more specific and longer sequences, are generated so that complete matching at all items between any pair of patterns becomes more difficult. This explains why the number of emerging patterns decreases as *min_sup* is decreased.

(2) When $\beta$ is changed from 0.65 to 0.55, the number of change patterns for unexpected and added/perished patterns increases. This means that increasing $\beta$ leads to more rigorous matching in the similarity computation between two patterns. Therefore, if $\beta$ is higher, the matching precision can be also improved. This finding suggests that we can set some higher $\beta$ values to control the quality of change detection.

(3) We also found that the number of the significant change patterns for emerging, unexpected, and added/perished patterns decrease when $\psi_{min}$ is changed from 0.3 to 0.6. This is because the higher threshold to identify the significant change patterns causes less significant patterns to exceed the threshold. This finding suggests that if managers do not want to know too much redundant information about change patterns, $\psi_{min}$ could be specified higher to gain more outstanding change knowledge for management.

(4) As shown in Figs. 13(a) and 13(b), we found that the number of the significant quantity trend changes in QTY-Increased emerging patterns was the most, followed by those in QTY-Stable emerging patterns, with those in QTY-Decreased emerging patterns being the least. The results imply that the purchased quantity dimension in customer buying behavior becomes more between TP1 and TP2. Consequently, managers should prepare more stock for some specified products and restock them frequently. In summary, the detection for the three types of patterns can help managers to investigate the quantity trend change both in products and in customer behavior.

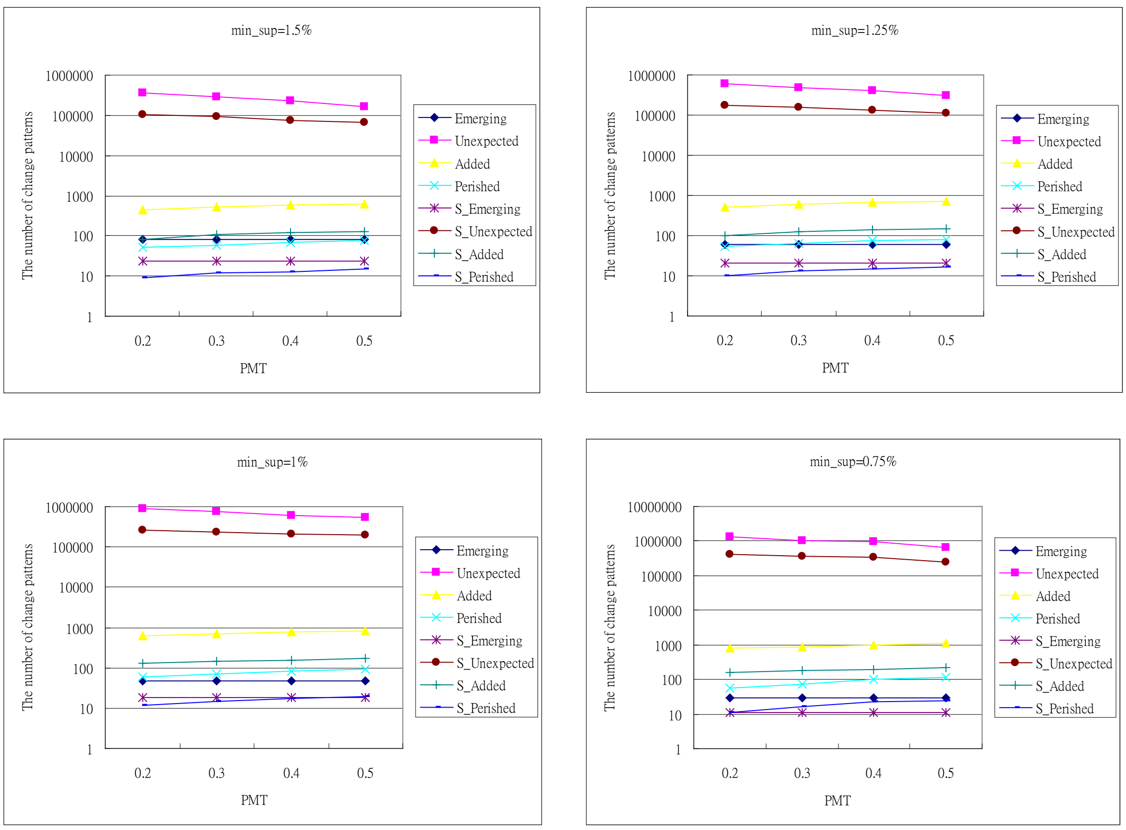


Fig 6. The number of change patterns for three (significant) change types (B0.65-P0.3).

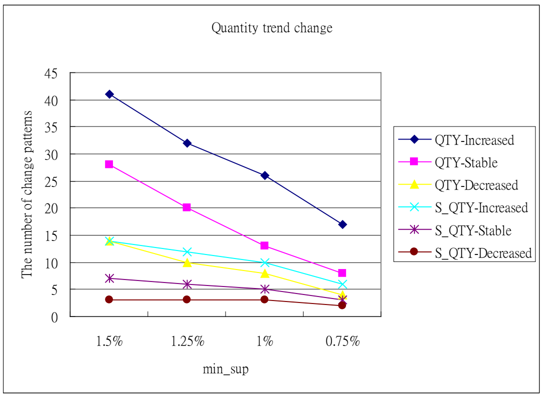


Fig 7(a). The number of emerging patterns for (significant) quantity trend changes in B0.65-P0.3 (or B0.55-P0.3).

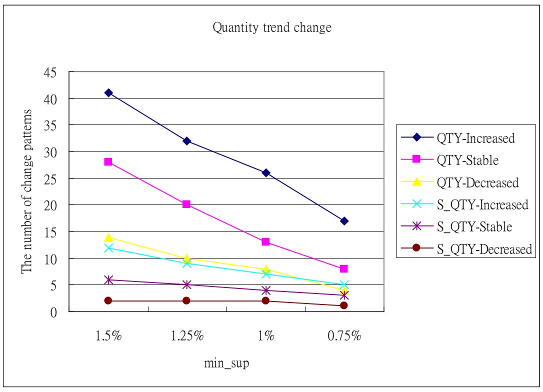


Fig 13(b). The number of emerging patterns for the (significant) quantity trend changes in B0.65-P0.6 (or B0.55-P0.6).

Finally, we discuss some interesting change patterns for the three change types of their business implications. They are all based on the first test, B0.65-P0.3, *min_sup*=1.5%, and PMT=0.3.

In Table 7 we show four significant emerging patterns, which appear both at TP1 and TP2. Observing the patterns, we found that their items are foods and daily products, which are necessary for human beings every day. The most significant emerging pattern is No. 1, of which support grows from 1.54 to 3.68 and its significant values is 1.34. Other patterns are also having the significant emerging characteristic at TP2. These results reveal that the products become more welcome to customers in the market. Managers can adjust their inventory plans to stock the products more to meet the customers' demand.

In addition, their quantity trend changes are as follows. No. 1, No. 2, and No. 3 are QTY-Increased emerging patterns, and only No. 4 is a QTY-Stable emerging pattern. According to the results, managers have to prepare more products, including Cola, Edible Oils, and Toilet Paper, to stock to cater to customers' demand since their purchased quantities are increased at TP2. Other products, such as Fruit Juice, Soy Sauce, House Cleaner, Pork, and Fresh Shrimp, can keep their regular inventory plans since their purchased quantities are stable. Gaining knowledge of emerging patterns can help managers to bear their existing beliefs and continue to execute their original marketing plans. Also, they can adjust the inventory plans for products by the detection of the quantity trend changes.

Table 7The significant emerging patterns from the Songchine real dataset

| No. | Pattern | Support (%) | $\Gamma_{ij}$ |
|---|---|---|---|
| 1 | TP1: < {(Cola, *Middle*)(Fruit Juice, *Low*)} (Cola, *Middle*) > | 1.54 | 1.34 |
| | TP2: < {(Cola, *High*)(Fruit Juice, *Low*)} (Cola, *Middle*) > | 3.68 | |
| 2 | TP1: < {(Edible Oils, *Very Low*)(Soy Sauce, *Very Low*)} > | 1.84 | 0.72 |
| | TP2: < {(Edible Oils, *Low*)(Soy Sauce, *Very Low*)} > | 3.17 | |
| 3 | TP1: < (Toilet Paper, *Middle*) (House Cleaner, *Low*) > | 2.11 | 0.46 |
| | TP2: < (Toilet Paper, *High*) (House Cleaner, *Low*) > | 3.09 | |
| 4 | TP1: < {(Pork, *Middle*)(Fresh Shrimp, *Low*)} > | 1.52 | 0.36 |
| | TP2: < {(Pork, *Middle*)(Fresh Shrimp, *Low*)} > | 2.06 | |

In Table 8 we show two significant unexpected patterns. The pattern structures between TP1 and TP2 of these patterns are partially similar. As observed in No. 1, we found that the characteristic of Leafy Vegetable is seasonal, in that it is sold only at a certain season of the year. Therefore, Leafy Vegetable takes the place of Organic Vegetable at TP2 in which the former season is expired. For No. 2, we found that customers alter their buying behavior, in that they are no longer buying Instant Noodles with Cola. A new purchased behavior is to buy Instant Noodles with Mineral Water, and the former quantity is constant and the latter is *Middle*. As discussed, we think that customers might start to concern about a health-care issue; therefore, the frequency of buying Instant Noodles with Cola is decreased but that of buying Instant Noodles with Mineral Water is increased. Gaining knowledge of unexpected patterns can remind managers that their existing beliefs should be updated partially and that their original marketing plans should also be modified immediately.

Table 8 The significant unexpected patterns from the Songchine real dataset

| No. | Pattern | Support (%) | $\Gamma_{ij}$ |
|---|---|---|---|
| 1 | TP1: < (Milk, *Middle*) {(Egg, *High*)(Leafy Vegetable, *Low*)} > | 1.62 | 1.07 |
| | TP2: < (Milk, *Middle*) {(Egg, *Middle*)(Organic Vegetable, *Middle*)} > | 2.08 | |
| 2 | TP1: < {(Instant Noodles, *High*)(Cola, *Low*)} > | 1.88 | 0.87 |
| | TP2: < {(Instant Noodles, *High*)(Mineral Water, *Middle*)} > | 2.56 | |

As shown in Table 9 there are two added patterns which appear at TP2 but not at TP1. The pattern, No. 1, raises a seasonal issue since Taiwanese Apples and Vegetables cannot be produced at TP2, so these products that are replaced by Imported Apples and Organic Vegetables. Observing the pattern, No. 2, we found that Essence of Chicken and Tea Bag are welcome at TP2. The reason might be cold weather in Taiwan is approaching at this time-period, and the both products could be healthful. Gaining knowledge of added patterns can help managers to append new beliefs to their knowledge base and draw up other marketing plans for the market.

Table 9.The significant added patterns from the Songchine real dataset

| No. | Pattern | Support (%) | $MinSSCI_i^t$ | $\Gamma_{ij}$ |
|---|---|---|---|---|
| 1 | < (Imported Apple, *Low*) (Organic Vegetable, *Middle*) > | 2.68 | 0.35 | 0.94 |
| 2 | < {(Essence of Chicken, *Low*) (Tea Bag, *Middle*)} > | 1.63 | 0.35 | 0.57 |

As shown in Table 10there are two perished patterns which appear at TP1 but not at TP2. As observed in two patterns, we found that they are also affected by the seasonal factor. When the weather becomes cold in Taiwan, the sales volumes of Ice Cream and Tea Drink decrease simultaneously. This explains why these patterns disappear at TP2. Gaining knowledge of perished patterns can help managers to delete some beliefs from their knowledge base and update their marketing plans to address the market.

Table 10.The significant perished patterns from the Songchine real dataset

| No. | Pattern | Support (%) | $MinSSCI_j^{t+l}$ | $\Gamma_{ij}$ |
|---|---|---|---|---|
| 1 | < {(Ice Cream, *Middle*) (Tea Drink, *High*)} > | 2.17 | 0.32 | 0.69 |
| 2 | < (Ice Cream, *Middle*) (Ice Cream, *Low*) > | 2.03 | 0.32 | 0.65 |

Based on the above results, we can thoroughly comprehend the change of customer behavior between TP1 and TP2. Although the time-period to mine change is from 2001/12/28 to 2002/12/28, we still obtain many useful findings which can provide referrals in the future. The seasonal factor, for example, is a regular issue in Taiwan, so managers should pay attention to this change in the future. Understanding what daily products and their quantities have to be stocked is another issue that allows assisting managers to control their inventories efficiently.

# 6. Conclusions

In this study, we introduced the matching concept of fuzzy quantitative sequential patterns. We also developed matching method by constructing new formulas, and defined the method of judging three different types of change pattern (emerging, unexpected and added/perished patterns) with empirical examples. The whole process of fuzzy quantitative sequential patterns from defining, finding and similarity computing could be integrated into a proposed model—FuzzChgMining.

Some interesting results are found as follows. First, three parameters including min_sup, β and ψmin have a great impact on the number of patterns. For example, the number of change patterns for unexpected and added/perished patterns increase when min_sup is decreased, whereas emerging patterns do not. The reason is that lowering the bar of minimum support threshold will make it easy for more candidate patterns to pass. Second, the purchased quantity between TP1 and TP2 is getting more and more. Managers should prepare more stocks for specific products. Third, some QTY-Increased emerging patterns happened in TP2, and managers have to prepare more products such as Cola, Edible Oils and Toilet Paper to live up to customers' demand. Fourth, seasonal and health-care issue were the main factors affect significant unexpected patterns between TP1 and TP2. Fifth, seasonal issue plays a big part in the significant added/perished patterns especially in Taiwan retail industry.

The change model for mining fuzzy quantitative sequential patterns stands for a novel and promising research issue in data mining. In business field, it provides proprietors with useful and valuable information about consuming changes, which helps them to keep track of the latest trends in customer behavior. In academic area, it offers some possible extensions and potential implications for future research. There are still other pattern detections to discover changes, such as multi-level sequential patterns, time-interval sequential patterns, RFM sequential patterns, and sequential patterns with different minimum supports. In fuzzy cases, we can also devise new models to mine some fuzzy patterns such as time-interval sequential patterns and fuzzy multi-level sequential patterns, and try to set up similarity computing formulas. Still, all proposed model's performance should be verified by using B2B, B2C or brick-and-mortar dataset.

# References


1. Chen, M. C., Chiu, A. L., and Chang, H. H., 2005 Mining changes in customer behavior in retail marketing. *Expert Systems with Applications*, 28, 4, 773-781.

2. Chen, Y. L., and Huang, C. K., 2006 A new approach for discovering fuzzy quantitative sequential patterns in sequence databases. *Fuzzy Sets and Systems*, 157, 12, 1641-1661.

3. Chen, Y. L., and Huang, C. K. , 2008 A novel knowledge discovering model for mining fuzzy multi-level sequential patterns in sequence databases. *Data & Knowledge Engineering*, 66, 3, 349-367.

4. Chen, Y. L., Kuo, M. H., Wu, S. Y., and Tang, K. , 2009 Discovering recency, frequency, and monetary (RFM) sequential patterns from customers' purchasing data. *Electronic Commerce Research and Applications*, 8, 5, 241-251.

5. Cho, Y. B., Cho, Y. H., and Kim, S. H. , 2005 Mining changes in customer buying behavior for collaborative recommendations. *Expert Systems with Applications*, 28, 2, 359-369.

6. Dong, G., and Li, J. , 1999 Efficient mining of emerging patterns: discovering trends and differences. In *Proceedings of the Fifth International Conference on Knowledge Discovery and Data Mining*, 43-52.

7. Gusfield, D. , 1997 *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology*. Cambridge University Press, Cambridge.

8. Hong, T. P., Kuo, C. S., and Chi, S. C. , 1999 Mining fuzzy sequential patterns from quantitative data. In *The 1999 IEEE International Conference on Systems, Man, and Cybernetics*, 962-966.

9. Huang, C. L., and Huang, W. L. , 2009 Handling sequential pattern decay: developing a two-stage collaborative recommender system. *Electronic Commerce Research and Applications*, 8, 3, 117-129.

10. Kim, J. K., Cho, Y. H., Kim, W. J., Kim, J. R., and Suh, J. H. , 2002 A personalized recommendation procedure for internet shopping support. *Electronic Commerce Research and Applications*, 1, 3, 301-313.

11. Kim, C., Lim, J. H., Ng, R., and Shim, K. , 2004 SQUIRE: Sequential pattern mining with quantities. In *Proceedings of the 20th International Conference on Data Engineering*, 827-827.

12. Lanquillon, C. , 1999 Information filtering in changing domains," In *Proceedings of the International Joint Conference on Artificial Intelligence*, 41-48.

13. Liu, B., Hsu, W., Han, H. S., and Xia, Y. , 2000 Mining changes for real-life applications. In *Second International Conference on Data Warehousing and Knowledge Discovery*, 337-346.

14. Lin, Q. Y., Chen, Y. L., Chen, J. S., and Chen, Y. C. , 2003 Mining inter-organizational retailing knowledge for an alliance formed by competitive firms. *Information and Management*, 40, 5, 431-442.

15. Padmanabhan, B., and Tuzhilin, A. , 1999 Unexpectedness as a measure of interestingness in knowledge discovery. *Decision Support Systems*, 27, 3, 303-318.

16. Song, H. S., Kim, J. K., and Kim, S. H. , 2001 Mining the change of customer behavior in an Internet shopping mall. *Expert Systems with Applications*, 21, 3, 157-168.

17. Tsai, C. Y., and Shieh, Y. C. , 2009 A change detection method for sequential patterns. *Decision Support Systems*, 46, 2, 501-511.

18. Watson, H. J., and Frolick, M. N. , 1993 Determining information requirement for an EIS. *MIS Quarterly*, 17, 3, 255-269.

19. Yang, T. C., and Lai, H. , 2006 Comparison of product bundling strategies on different online shopping behaviors. *Electronic Commerce Research and Applications*, 5, 4, 295-304.