Journal of Zhejiang University-SCIENCE C (Computers & Electronics)
ISSN 1869-1951 (Print); ISSN 1869-196X (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Mining item-item and between-set correlated association rules*

Bin SHEN[†1], Min YAO[†‡2], Li-jun XIE[3], Rong ZHU[2], Yun-ting TANG[1]

([1]*Ningbo Institute of Technology, Zhejiang University, Ningbo 315100, China*)

([2]*School of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China*)

([3]*Center for Engineering & Scientific Computation, School of Aeronautics and Astronautics, Zhejiang University, Hangzhou 310027, China*)

[†]E-mail: {tsingbin, myao}@zju.edu.cn

Received Nov. 19, 2009; Revision accepted Mar. 22, 2010; Crosschecked Dec. 6, 2010

Abstract: To overcome the failure in eliminating suspicious patterns or association rules existing in traditional association rules mining, we propose a novel method to mine item-item and between-set correlated association rules. First, we present three measurements: the association, correlation, and item-set correlation measurements. In the association measurement, the all-confidence measure is used to filter suspicious cross-support patterns, while the all-item-confidence measure is applied in the correlation measurement to eliminate spurious association rules that contain negatively correlated items. Then, we define the item-set correlation measurement and show its corresponding properties. By using this measurement, spurious association rules in which the antecedent and consequent item-sets are negatively correlated can be eliminated. Finally, we propose item-item and between-set correlated association rules and two mining algorithms, I&ISCoMine_AP and I&ISCoMine_CT. Experimental results with synthetic and real retail datasets show that the proposed method is effective and valid.

Key words: Item-item and between-set correlated association rules, All-confidence, All-item-confidence, Item-set correlation, Mining algorithms, Pruning effect

doi:10.1631/jzus.C0910717 **Document code:** A **CLC number:** TP311

1 Introduction

Association rules mining, first proposed by Agrawal *et al*. (1993), has many successful applications, especially in the analysis of consumer market-basket data. An association rule can be denoted as '$X \Rightarrow Y[s, c]$' (X, Y are item-sets), if (1) s of transactions in the database D contain $X \cup Y$ (the support), and (2) c of transactions in D that contain X also contain Y (the confidence).

However, association rules that are based on frequent patterns still have several limitations in real-life applications.

Limitation 1 If the minimum support threshold is set low, too many spurious weakly-related cross-support patterns will be extracted (Xiong *et al*., 2006).

The weakly-related cross-support patterns are patterns that involve items with different support levels. These spurious frequent patterns are extremely poor, since they can express only the concurrent relationship between the items, not the implication relationship. For example, {Milk, Caviar} is possibly a spurious cross-support pattern, as the support of the item 'Milk' is much higher than that of the item 'Caviar'. We should not be surprised if such cross-support patterns are found. But for the customers, these patterns are meaningless, as they represent only the items' concurrent relationship rather than their implication relationship.

Limitation 2 If the minimum support threshold is set low, too many spurious patterns that contain negatively correlated items will be extracted.

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (Nos. 10876036 and 70871111) and the Ningbo Natural Science Foundation, China (No. 2010A610113)

These spurious patterns can also express only the concurrent relationship. Customers are inclined to buy the baskets where the items are positively correlated with each other. For example, if we mine a certain transaction database with a low minimum support threshold (e.g., 1%), we may find the following spurious frequent pattern: {Dell Computer, Legend Computer, Office Software} [s=1.5%] (s is the support), which contains two negatively correlated items, i.e., 'Dell Computer' and 'Legend Computer'. Based on the high supports of the item-item positively correlated sub-itemsets {Dell Computer, Office Software} [s=30%] and {Legend Computer, Office Software} [s=39%], we are not surprised that frequent patterns containing negatively correlated items are generated, e.g., {Dell Computer, Legend Computer, Office Software}. But these are misleading to the users. If we apply this rule and establish a bundling selling promotion that contains {Dell Computer, Legend Computer, Office Software}, customers will not pay much attention to it, since the promotion contains two replaceable negatively correlated items: 'Dell Computer' and 'Legend Computer'.

Limitation 3 The selling of the rule consequent item-set may not be promoted by that of the rule antecedent item-set.

If we apply the association rules whose antecedent and consequent item-sets are negatively correlated, the selling of the antecedent item-set may not promote that of the consequent item-set, and may even bring about an unexpected loss in sales of the consequent item-set. For example, consider a certain transaction database in which P(Coffee)=50%, P(Coffee, Tea)=5%, and P(Coffee|Tea)=25%. If we mine this database by setting the minimum support threshold at 2%, and obtain a frequent pattern: {Coffee, Tea} [s=5%], then we obtain the rule 'Tea $\Rightarrow$ Coffee [s=5%, c=25%]'. That is, the prior and posterior probabilities of the selling of 'Coffee' are 50% and 25%, respectively. Therefore, if this rule is applied, an unexpected loss will be made in the selling of 'Coffee' by easily misleading the customers.

Existing methods cannot overcome the above three limitations completely. Xiong *et al*. (2006) proposed the h-confidence-based hyperclique pattern and its mining algorithm. Lee *et al*. (2003) presented the all-confident patterns, and Kim *et al*. (2004) put forward the confidence-closed correlated patterns. These patterns can filter only the spurious patterns in Limitation 1 and cannot overcome Limitations 2 and 3. The corr-confidence-based associated and correlated patterns (Zhou *et al*., 2006a) can ensure that the pattern X contains two positively correlated subsets, but it cannot guarantee that the pattern is positively item-item correlated. Zhou *et al*. (2006b) proposed mutually and positively correlated patterns, which can eliminate the spurious frequent patterns of Limitation 2. But its definition in relation to the correlation is too strong; i.e., it will filter many frequent patterns that are positively item-item correlated, but in which their subsets are not all positively correlated. In our previous work (Shen and Yao, 2009a), we presented a new correlation measurement (i.e., all-item-confidence L), and proposed a method to mine the associated and item-item correlated frequent patterns, which can completely eliminate spurious cross-support patterns and patterns containing negatively correlated items. But if we use the proposed patterns to generate the association rules directly, the problem mentioned in Limitation 3 will still exist.

To overcome all three limitations, we propose a new method to mine the item-item and between-set correlated association rules. The mining process for the item-item and between-set correlated association rules includes two steps. First, to mine the associated and item-item correlated frequent patterns, we choose the all-confidence as the association measurement and the all-item-confidence as the correlation measurement. We then use the item-set correlation measurement to test the correlation between the rule antecedent and rule consequent item-sets, when we apply these patterns to generate the rules.

This paper extends our previous work (Shen and Yao, 2009a) to mining item-item and between-set correlated association rules and makes the following contributions: First, we clearly spell out various spurious frequent patterns or association rules that might be involved in current association rules mining. Also, we provide a solution to eliminating these spurious patterns or association rules. Second, we introduce all-item-confidence and make comparisons between all-item-confidence and other interestingness measures. Third, we introduce the item-set correlation measurement and discuss its corresponding properties. Also, we give the definition of the item-item and between-set correlated association rules and an

illustrative example. Finally, we refine our algorithms that are used to discover item-item and between-set correlated association rules. Experimental results show that the proposed method can complete the mining task efficiently. In addition, we demonstrate the pruning effect of the measurements.

2 Related works

Association rules, originally introduced by Agrawal *et al*. (1993), and the support-confidence framework used to mine them, have been studied extensively (Han *et al*., 2004; Palshikar *et al*., 2007; Shen and Yao, 2009b; Shen *et al*., 2010). They are intended to capture potential and meaningful implications between antecedents and consequents.

A series of studies has been carried out with the aim of mining various patterns or association rules, including cyclic association rules (Ozden *et al*., 1998), inter-transaction rules (Lu *et al*., 1998), episode rules (Qin and Hwang, 2004), dynamic association rules (Shen and Yao, 2009b; Shen *et al*., 2010), and heavy item-sets (Palshikar *et al*., 2007).

Unfortunately, support-confidence framework based association rules mining tends to generate a large number of meaningless, redundant, or misleading patterns or association rules, as discussed in Section 1. To solve this problem, several studies have focused on mining interestingness measures based patterns or association rules. They focus mainly on the following two research directions:

Some studies discussed the interestingness measurements for patterns or association rules mining (Tan *et al*., 2002; Omiecinski, 2003; Xiong *et al*., 2006; Hahsler and Hornik, 2007; Kenett and Salini, 2008a; Lenca *et al*., 2008). For example, Omiecinski (2003) proposed three alternative interest measures: any-confidence, all-confidence, and bond. Xiong *et al*. (2006) independently defined a metric called *h*-confidence, which is equivalent to the all-confidence measure. Hahsler and Hornik (2007) developed two probabilistic framework based interest measures: hyper-lift and hyper-confidence. Kenett and Salini (2008a) put forward relative linkage disequilibrium, which is powerful in its intuitive visual interpretation. These interest measures have their own virtues and are suited to different applications. Although many different measures of interestingness have been proposed for association rules, none has yet been accepted widely. The need to select the right measure for a given application domain has been recognized by some researchers (Tan *et al*., 2002; Lenca *et al*., 2008). Tan *et al*. (2002) discussed how to choose suitable interestingness measures for various applications. Lenca *et al*. (2008) presented eight properties for evaluating the measures, and provided a multiple criteria decision aid approach. This approach can be seen as an alternative to that of Tan *et al*. (2002).

Some studies adopted different interestingness measures to mine various correlated patterns or association rules (Brin *et al*., 1997; Kim *et al*., 2004; Xiong *et al*., 2006; Zhou *et al*., 2006a; 2006b; 2006c). Brin *et al*. (1997) used the measure of interest and the chi-squared test to validate the correlations of patterns, and proposed a corresponding mining algorithm. Xiong *et al*. (2006) adopted a support measure and *h*-confidence measure to mine highly-correlated association patterns called hyperclique patterns. Zhou *et al*. (2006a; 2006b; 2006c) mined both associated and correlated patterns or rules, and applied support measure, corr-confidence, and all-confidence to the mining process. To reduce the number of correlated patterns produced without information loss, Kim *et al*. (2004) proposed confidence-closed correlated patterns mining. Confidence-closed correlated patterns should satisfy both the minimal support and minimal all-confidence thresholds, and have no proper superset with the same support and the same all-confidence. These correlated patterns or association rules have their own features and are suited to different application domains.

In association rules mining, spurious patterns or association rules can be seen as outliers. The users would like to list association rules implying an implication relationship but not a concurrent relationship. Thus, the work can be considered in the general context of outliers, causality, and interactions. In the data mining literature, some studies have considered these issues. Zhang *et al*. (2006; 2009) focused on identifying a kind of outlier called bridging rules between conceptual clusters. This kind of rule represents interactions that look like bridges linking different clusters, and is useful in many domains such as criminal detection and biological grafting. Causality has been considered in the statistical literature using

tools such as Bayesian networks and cause-and-effect diagrams (Ruggeri *et al*., 2007). In the market baskets analysis domain, if we apply Bayesian networks directly, some difficulties still exist. First, there are commonly thousands of items in a transaction database. If each item is assigned a node, a huge Bayesian belief-network is created, and the cost of its construction and maintenance is considerable. Second, some hidden variables may exist in market basket analysis. For example, the correlations among 'baby food', 'diaper', and 'milk' are well known. We do not put arcs between two of these items directly, but we are likely to add a hidden variable 'baby' that is the hidden reason for the purchasing of these items. Cause-and-effect (or fishbone) diagrams provide a structured way to help people think through all possible causes of a problem. Several interest measures, e.g., the chi-squared test (Brin *et al*., 1997) and relative linkage disequilibrium (Kenett and Salini, 2008a; 2008b), have considered the interactions between antecedents and consequents. Unlike these approaches, our approach has been developed for mining item-item and between-set correlated association rules, with all-confidence, all-item-confidence, and item-set correlation measurements as interestingness measures.

3 All-item-confidence measure

3.1 Definition and properties of all-item-confidence

Different interestingness measures have different meanings, properties, and scopes of application. To measure the item-item correlations of patterns effectively, we propose a new correlation measurement, all-item-confidence L (Shen and Yao, 2009a).

Definition 1 (All-item-confidence) Given a certain pattern $X=\{i_1, i_2, \ldots, i_n\}$, $n>1$, the all-item-confidence is defined as

$$L(X)=\min\left\{\frac{P(i_j,i_k)-P(i_j)P(i_k)}{P(i_j,i_k)+P(i_j)P(i_k)} \mid \forall j,k=1,2,...,n,\ j\neq k\right\}. \tag{1}$$

Example 1 Consider an item-set $X=\{i_1, i_2, i_3\}$. Assume $P(i_1)=0.1$, $P(i_2)=0.1$, $P(i_3)=0.06$, $P(i_1, i_2)=0.06$, $P(i_1, i_3)=0.03$, and $P(i_2, i_3)=0.03$. Since

$$\frac{P(i_1,i_2)-P(i_1)P(i_2)}{P(i_1,i_2)+P(i_1)P(i_2)}=0.714,$$

$$\frac{P(i_1,i_3)-P(i_1)P(i_3)}{P(i_1,i_3)+P(i_1)P(i_3)}=0.667,$$

$$\frac{P(i_2,i_3)-P(i_2)P(i_3)}{P(i_2,i_3)+P(i_2)P(i_3)}=0.667,$$

$L(X)=\min\{0.714, 0.667, 0.667\}=0.667$.

This measure can be interpreted as follows: if the correlation degree of any two items in the pattern is not less than the minimal threshold for all-item-confidence, this pattern can be regarded as an interesting pattern.

This measurement has many good properties. First, it has an appropriate boundary, [−1, 1], which is suitable for measuring the degree of correlation and easily adapts and controls the input parameters. Second, if the items in the pattern are independent of each other, the all-item-confidence measurement $L(X)$ will equal 0; if the items in the pattern are positively correlated with each other, $L(X)$ will be greater than 0; if two negatively correlated items exist in the pattern, $L(X)$ will be less than 0. Third, the all-item-confidence has a good anti-monotone property, which can be applied to promote the performance of the mining algorithm. Fourth, if pattern X satisfies the condition $L(X)\geq L$ $(1\geq L>0)$, we can guarantee that the probability $P(B)$ is significantly less than the conditional probability $P(B|A)$, for any two items A and B in the pattern. Thus, we can use it to promote the probability of occurrence of B. Finally, if X satisfies the condition $L(X)\geq L$ $(1\geq L>0)$, the probability of occurrence of the item in X can promote the other item's probability of occurrence by $(1+L)/(1-L)$ times, and this can be used to promote all of the items in X.

In the following, we will describe the properties of the all-item-confidence in detail. The corresponding proofs are given in our previous publication (Shen and Yao, 2009a).

Property 1 For a given pattern $X=\{i_1, i_2, \ldots, i_n\}$, its all-item-confidence $L(X)$ has the upper and lower boundary $-1\leq L(X)\leq 1$.

Property 2 If $X'\subset X$ and $L(X)\geq L$ (L is the minimal all-item-confidence threshold), $L(X')\geq L$ can be obtained; i.e., the all-item-confidence is anti-monotone.

Property 3 For a given pattern $X=\{i_1, i_2, \ldots, i_n\}$ and its all-item-confidence $L(X)$, the following propositions are true:

(1) If i_1, i_2, …, i_n are independent of each other, $L(X)=0$.

(2) If the items in X are positively correlated with each other, $L(X)>0$.

(3) If there are two negatively correlated items in X, $L(X)<0$.

Property 4 Given a certain pattern X, if $L(X)\geq L>0$ (L is the minimal all-item-confidence threshold), for any two items i_1 and i_2 in X, $P(i_1|i_2)\geq P(i_1)\cdot(1+L)/(1-L)>P(i_1)$, $P(i_2|i_1)\geq P(i_2)\cdot(1+L)/(1-L)>P(i_2)$.

Corollary 1 Given a certain pattern X, if $L(X)\geq L$ (L is the minimal all-item-confidence threshold, $1\geq L>0$), the probability of occurrence of the item in X can promote that of the other item by $(1+L)/(1-L)$ times.

3.2 Comparisons between all-item-confidence, all-set-confidence, and the chi-squared (χ^2) test

In this section, we first introduce the relationship between all-item-confidence and all-set-confidence and then discuss the difference between all-item-confidence and the χ^2 test.

3.2.1 Relationship between all-item-confidence and all-set-confidence

Since the constraint of mutually and positively correlated patterns (Zhou *et al*., 2006b) means that any two non-empty subsets of the pattern should be positively correlated, we call it all-set-confidence. It follows that:

$$\rho(X)=\min\left\{\frac{P(A,B)-P(A)P(B)}{P(A,B)+P(A)P(B)}\,\middle|\,\forall A,B\subset X, A\neq\varnothing, B\neq\varnothing\right\}, \tag{2}$$

where X is a pattern and A and B are two non-empty sub-item-sets of X.

Property 5 Given a pattern $X=\{i_1, i_2, \ldots, i_n\}$, its all-item-confidence $L(X)$ and its all-set-confidence $\rho(X)$, the following equation is true:

$$L(X)\geq\rho(X). \tag{3}$$

Proof

$$L(X)=\min\left\{\frac{P(i_j,i_k)-P(i_j)P(i_k)}{P(i_j,i_k)+P(i_j)P(i_k)}\,\middle|\,\forall j,k=1,2,\ldots,n, j\neq k\right\}$$

$$\geq\min\left\{\frac{P(A,B)-P(A)P(B)}{P(A,B)+P(A)P(B)}\,\middle|\,\forall A,B\subset X, A\neq\varnothing, B\neq\varnothing\right\}$$

$$=\rho(X).$$

This property shows that the constraint of all-set-confidence is much stronger than that of all-item-confidence. The reason is that all-item-confidence requires any two items in the pattern be positively correlated, while all-set-confidence requires any two non-empty sub-item-sets be positively correlated. For a given pattern X, if we use the same value v as the minimal all-item-confidence threshold and the minimal all-set-confidence threshold, the pattern set obtained using the all-item-confidence measure will contain the pattern set generated by the all-set-confidence measure.

The relationships between the pattern sets that satisfy the all-item-confidence measure and the all-set-confidence measure are displayed in Fig. 1.

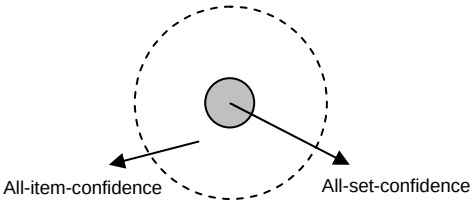

Fig. 1 Relationships between pattern sets produced by all-item-confidence and all-set-confidence

3.2.2 Difference between all-item-confidence and the χ^2 test

The χ^2 test (Brin *et al*., 1997) is a well-known statistical measure used to determine the dependence between items. Computing the χ^2 statistic for a pair of items requires constructing two contingency tables, the observed contingency table and the expected one. The square statistic is defined as follows (Alvarez, 2003):

$$\chi^2=\sum_{0\leq i,j\leq 1}\frac{(\text{observed}_{i,j}-\text{expected}_{i,j})^2}{\text{expected}_{i,j}}, \tag{4}$$

where $\text{observed}_{i,j}$ and $\text{expected}_{i,j}$ ($0\leq i, j\leq 1$) are elements in the observed and the expected contingency tables, respectively.

The difference between all-item-confidence and the χ^2 test includes the following main aspects:

1. χ^2 tests are commonly used to test the dependence between two items, while all-item-confidence is applied to test the correlation of a pattern that includes two or more items.

Although χ^2 tests may sometimes be extended to a pattern containing two or more items, a high χ^2

value can indicate only that two items in a pattern are dependent, but not any two.

2. The χ^2 test can determine the dependence between two items, but it cannot distinguish directly between a positive and a negative correlation for two items. An additional measure is always needed to indicate a positive or negative correlation.

For example, Brin *et al*. (1997) adopted not only the χ^2 test, but also the interest measure for examining all possible pairs of items. Liu *et al*. (1999) first used the χ^2 test to determine whether a rule is correlated or not. Then, for a given correlated association rule $A \Rightarrow B$, to determine the type of correlation, they compared the value of f_0/f with 1, where f_0 is the observed frequency of $\{A, B\}$ and f is the expected frequency. Since $\frac{f_0}{f} = \frac{P(A \cap B)}{P(A)P(B)}$, it is actually the interest measure. Liu *et al*. (1999) also applied the χ^2-interest test for examining the type of correlation.

3. There are conditions for the use of χ^2 tests in typical basket data analysis. The χ^2 test depends on the normal approximation to the binomial distribution, and the approximation may break down if the expected values are small (Brin *et al*., 1997).

4 Item-item and between-set correlated association rules mining

4.1 Choice of the association measurement

Currently, although several association or correlation measurements have been proposed, no effective measurement has been widely accepted. The most frequently used correlation measurements are all-confidence and *h*-confidence, proposed by Omiecinski (2003) and Xiong *et al*. (2006), respectively. The two measurements have the same basis:

Definition 2 (All-confidence) Given a certain pattern $X=\{i_1, i_2, \ldots, i_n\}$, the all-confidence is defined as

$$\text{all\_conf}(X) = \frac{\sup(X)}{\text{max\_item\_sup}(X)}, \quad (5)$$

$$\max\_\text{item}\_\sup(X) = \max\{\sup(i_j) \mid \forall i_j \in X\}, \quad (6)$$

where sup() denotes the support of an item-set.

According to this definition, if the all-confidence of the pattern X is greater than or equal to the minimal threshold of the all-confidence, we can infer that for any two items A and B, the conditional probabilities $P(A|B)$ and $P(B|A)$ are greater than or equal to the minimal threshold of all-confidence. Thus, we can ensure that A and B are correlated with each other with enough strength. Meanwhile, the all-confidence has the properties of anti-monotone and cross-support. The anti-monotone property can be applied to improve the performance of the mining algorithm and the cross-support property can be used to eliminate spurious cross-support patterns. For these reasons, we choose all-confidence as the association measurement of X.

4.2 Choice of the correlation measurement

Different interest measures have different meanings, properties, and scopes of application. Thus, it is necessary to select the right measure for a particular application. Corr-confidence

$$\rho(X) = \frac{P(i_1, i_2, \cdots, i_n) - P(i_1)P(i_2)\cdots P(i_n)}{P(i_1, i_2, \cdots, i_n) + P(i_1)P(i_2)\cdots P(i_n)}, \ n \geq 1,$$

proposed by Zhou *et al*. (2006a), can ensure that there are two positively correlated sub-item-sets in X, but it cannot ensure that any two items in X are positively correlated with each other. All-set-confidence (Zhou *et al*., 2006b), as discussed in Section 3.2.1, is suitable for mining patterns in which any two non-empty subsets are positively correlated. Clearly, the condition of all-set-confidence is too strong. If we use all-set-confidence to generate patterns for a minimal threshold v, then the results will filter out the patterns that satisfy all-item-confidence but do not satisfy all-set-confidence. The χ^2 test is commonly used to test the dependence between two items. Since the lift measure $\frac{P(A \cap B)}{P(A)P(B)}$ has no appropriate upper or lower bound, it is not convenient for users to adjust the minimal threshold value. Thus, these measures are not suitable for mining patterns in which any two items are positively correlated.

We know that all-item-confidence has many good properties (Section 3.1). For example, it has an appropriate boundary, which is suitable for measuring the degree of correlation and easily adapts the input parameters. If pattern X satisfies the minimal

threshold of all-item-confidence L, i.e., $L(X)\geq L$ ($1\geq L>0$), the probability of occurrence of the item in X can promote the other item's probability of occurrence by $(1+L)/(1-L)$ times, and this can be used to promote all of the items in X. In addition, it has a good anti-monotone property, which can be used for pruning uninteresting candidates efficiently.

For these reasons, we choose all-item-confidence as the correlation measurement to solve the problem in Limitation 2.

4.3 Definition of the item-set correlation measurement

Definition 3 (Item-set correlation measurement) If the association rule r is $A\Rightarrow X-A$, the item-set correlation measurement $C(A, X-A)$ (or $C(r)$) can be defined as

$$C(A, X-A)=C(r)=\frac{P(X)-P(A)P(X-A)}{P(X)+P(A)P(X-A)}. \quad (7)$$

This definition is derived from the transformation of the treatment of the correlation between two item-sets by applying the all-item-confidence, and is the same as the correlation-confidence proposed by Zhou *et al*. (2006c). It has the following properties:

Property 6 For a given association rule $A\Rightarrow X-A$, its item-set correlation between the antecedent and consequent item-sets, $C(A, X-A)$, has the boundary $-1\leq C(A, X-A)\leq 1$.

Proof

$$C(A, X-A)=\frac{P(X)-P(A)P(X-A)}{P(X)+P(A)P(X-A)}\leq\frac{P(X)}{P(X)+P(A)P(X-A)}\leq 1,$$

$$\frac{P(X)-P(A)P(X-A)}{P(X)+P(A)P(X-A)}\geq\frac{-P(A)P(X-A)}{P(X)+P(A)P(X-A)}\geq -1. \quad (8)$$

Property 7 For a given association rule $A\Rightarrow X-A$, its item-set correlation between the antecedent and consequent item-sets, $C(A, X-A)$, has the following properties:

(1) If A and $X-A$ are independent of each other, $C(A, X-A)=0$.

(2) If $C(A, X-A)>0$, A and $X-A$ are positively correlated.

(3) If $C(A, X-A)<0$, A and $X-A$ are negatively correlated.

(4) $C(A, X-A)=C(X-A, A)$; i.e., the item-set correlation measurement is symmetrical.

Proof

(1) If A and $X-A$ are independent of each other, that is to say, $P(X)=P(A)P(X-A)$, $C(A, X-A)=0$ can be obtained.

(2) If $C(A, X-A)>0$, $P(X)>P(A)P(X-A)$, meaning that A and $X-A$ are positively correlated.

(3) If $C(A, X-A)<0$, $P(X)<P(A)P(X-A)$, meaning that A and $X-A$ are negatively correlated.

(4) $C(A, X-A)=[P(X)-P(A)P(X-A)]/[P(X)+P(A)P(X-A)]=C(X-A, A)$.

We can adopt Property 7(4) to improve the computing efficiency of the mining process.

Property 8 Given an association rule $A\Rightarrow X-A$ ($A\subset X$, $A\neq\varnothing$), if it satisfies $C(A, X-A)\geq C$ (C is the minimal positive item-set correlation threshold, $1\geq C>0$), $P(X-A|A)\geq P(X-A)\cdot(1+C)/(1-C)$; i.e., the probability of occurrence of A can promote that of $X-A$ by $(1+L)/(1-L)$ times.

Proof

By Definition 3,

$$C(A, X-A)=\frac{P(X)-P(A)P(X-A)}{P(X)+P(A)P(X-A)}\geq C.$$

We obtain

$$\frac{P(X)}{P(A)P(X-A)}\geq\frac{1+C}{1-C}.$$

Then, we can derive $P(X-A|A)\geq P(X-A)\cdot(1+C)/(1-C)$.

This property has good practical value and can be used to promote the subsequent item-set.

4.4 Definition of item-item and between-set correlated association rules

First, we give the definition of the item-item and between-set correlated association rules. Then, an example is given to explain the concepts.

Definition 4 (Between-set positively correlated association rules) If the associations rule $A\Rightarrow X-A$ ($A\subset X$, $A\neq\varnothing$) is generated by the pattern X, which satisfies $C(A, X-A)\geq C$ ($C(A, X-A)$ is the item-set correlation measurement and C is the minimal positive item-set correlation threshold), this rule is called a between-set positively correlated association rule.

According to Property 8, for a certain between-set positively correlated association rule, the selling of the antecedent item-set can promote the selling of the consequent item-set.

Definition 5 (Associated and item-item positively correlated frequent patterns) If the pattern X satisfies $\sup(X)\geq\sup$, $\text{all\_conf}(X)\geq\lambda$, and $L(X)\geq L$ ($\sup$, λ, and L are the minimal support, all-confidence, and all-item-confidence thresholds, respectively), it is called the associated and item-item positively correlated frequent pattern.

Definition 6 (Item-item and between-set correlated association rule) If the association rule $A\Rightarrow X-A$ ($A\subset X$, $A\neq\varnothing$) is generated by the associated and item-item positively correlated frequent pattern X, which satisfies $C(A, X-A)\geq C$ ($C(A, X-A)$ is the item-set correlation measurement and C is the minimal positive item-set correlation threshold), this rule is called the item-item and between-set correlated association rule.

Because the associated and item-item positively correlated frequent pattern can eliminate spurious frequent patterns, it is the basket that the customers are inclined to actually buy. The item-set correlation measurement $C(A, X-A)$ guarantees that the selling of the antecedent item-set can promote that of the consequent item-set. Thus, the item-item and between-set correlated association rule is quite well-suited for applications such as cross-selling.

Property 9 Let the number of elements in the associated and item-item positively correlated frequent pattern X be $|X|$, $\forall A\subset X$, $A\neq\varnothing$, if $|X|$ is set at 2, $C(A, X-A)=L(X)$.

Proof It is easy to show that this proposition is true, according to the definitions of the item-set correlation measurement and the all-item-confidence.

According to Property 9, if $|X|$ is set at 2, we need not compute the value of $C(A, X-A)$, and thus the efficiency of the mining algorithm can be improved.

Example 2 Given a transaction database (Table 1), let the minimal support threshold sup, the minimal all-item-confidence threshold L, minimal all-confidence threshold λ, and the minimal positive item-set correlation threshold C be 0.5, 0.01, 0.6, and 0.12, respectively.

Table 1 Transaction database

Transaction ID	Items	Transaction ID	Items
100	A, C, D, E	300	A, B, C, E
200	B, C, E	400	B, E

The whole process includes the following two steps:

Step 1: Generate the associated and item-item positively correlated frequent patterns.

Compute frequent 1-item-sets first. The result is shown in Table 2. Join frequent 1-item-sets to generate candidate 2-item-sets of associated and item-item positively correlated frequent patterns. Then we compute the corresponding supports, associations, and correlations (Table 3).

For example, let the support of pattern AC be 0.5. The association and correlation of AC can be computed as follows:

$$\text{all\_conf(AC)}=\frac{0.5}{\max(0.5,\ 0.75)}=0.67,$$

$$L(\text{AC})=\min\left(\frac{0.5-0.5\times0.75}{0.5+0.5\times0.75}\right)=0.14.$$

Table 2 Frequent 1-item-sets

Item	Support	Support≥sup
A	0.50	Yes
B	0.75	Yes
C	0.75	Yes
D	0.25	No
E	1.00	Yes

sup=0.50

Table 3 Candidate 2-item-sets of associated and item-item positively correlated frequent patterns*

Candidate	Sup.	Sup.≥sup?	All-conf.	All-conf.≥λ?	All-item-conf.	All-item-conf.≥L?	A&IICFP?
{A, B}	0.25	No	0.33	No	−0.20	No	No
{A, C}	0.50	Yes	0.67	Yes	0.14	Yes	Yes
{A, E}	0.50	Yes	0.50	No	0.00	No	No
{B, C}	0.50	Yes	0.67	Yes	−0.06	No	No
{B, E}	0.75	Yes	1.00	Yes	0.10	Yes	Yes
{C, E}	0.75	Yes	0.75	Yes	0.00	No	No

* sup=0.5, λ=0.6, L=0.01. Sup.=support; conf.=confidence; A&IICFP: associated and item-item correlated frequent pattern

In the following, we check if the candidate 2-item-sets satisfy the minimal support threshold, the minimal all-item-confidence threshold, and the minimal all-confidence threshold simultaneously. We obtain the 2-item-sets of associated and item-item positively correlated frequent patterns: AC and BE. Since these two patterns cannot join together to generate the candidate 3-item-sets of the associated and item-item positively correlated frequent patterns any more, the computing process for generating the associated and item-item positively correlated frequent patterns will finish.

Step 2: Generate the item-item and between-set correlated association rules based on the patterns in Step 1.

We obtain four candidates of the item-item and between-set correlated association rules, i.e., $A \Rightarrow C$, $C \Rightarrow A$, $B \Rightarrow E$, $E \Rightarrow B$. The item-set correlations between the antecedent and consequent item-sets are computed (Table 4). Finally, according to the minimal item-set correlation threshold C, the final item-item and between-set correlated association rules can be obtained: $A \Rightarrow C$ and $C \Rightarrow A$.

Table 4 Candidates of the item-item and between-set correlated association rules

Candidate rule	Item-set correlation degree	Item-set correlation degree≥C?	Item-item & between-set correlated association rule?
$A \Rightarrow C$	0.14	Yes	Yes
$C \Rightarrow A$	0.14	Yes	Yes
$B \Rightarrow E$	0.10	No	No
$E \Rightarrow B$	0.10	No	No

C=0.12

5 Mining algorithms: I&ISCoMine_AP and I&ISCoMine_CT

In the following, we propose two mining algorithms of the item-item and between-set correlated association rules, I&ISCoMine_AP and I&ISCoMine_CT, which are based on ItemCoMine_AP (Shen and Yao, 2009a) and ItemCoMine_CT (Shen and Yao, 2009a), respectively.

5.1 I&ISCoMine_AP algorithm

I&ISCoMine_AP applies the minimal all-item-confidence threshold, minimal all-confidence threshold, and minimal support threshold to obtain the associated and item-item correlated frequent patterns, and then uses the item-set correlation measurement to check the correlations between the antecedent and consequent item-sets.

Algorithm 1 I&ISCoMine_AP

Input: Dataset T, minimal support threshold sup, minimal all-item-confidence threshold L, minimal all-confidence threshold λ, and minimal item-set correlation threshold C.

Output: Item-item and between-set correlated association rules.

Methods

1 L_1={frequent 1-item-sets};
2 For (k=2; $L_{k-1}\neq\varnothing$; k++) {
3 Join L_{k-1} and L_{k-1} together to generate the candidates of the associated and item-item positively correlated frequent patterns and assign these candidates to P_k, and then incrementally test if P_k satisfies the item-item positive correlation threshold L;
4 For each transaction $t\in T$ {
Cumulate the supports of the candidates that are contained in t; }
5 For each $p\in P_k$ {
6 If sup(p)<sup, delete p from P_k;
7 If all_conf(p)<λ, delete p from P_k; }
8 $L_k\leftarrow P_k$;
9 For $p\in P_k$ {
10 Generate the candidate item-item and between-set correlated association rules r: $A \Rightarrow p-A$, $r\in \mathrm{RS}_k$;
if $|A|>|p-A|$, $C(A \Rightarrow p-A)\leftarrow C(p-A \Rightarrow A)$;
if k=2, $C(r)\leftarrow L(p)$;
11 If $C(r)<C$, delete r from RS_k; }
12 }
13 Return$\cup \mathrm{RS}_k$.

5.2 I&ISCoMine_CT algorithm

In I&ISCoMine_CT, after the associated and item-item positively correlated frequent patterns are obtained, we need to deposit the patterns and their supports into a trie-tree structure: TrieTree_with_count. TrieTree_with_count is then transferred to a Rule-Generation() function, where CFP-Construction(T, sup) and CFP-ItemCoMining(Tree, minsup, λ, L) are exactly the same as those in ItemCoMine_CT (Shen and Yao, 2009a).

Algorithm 2 I&ISCoMine_CT

Input: Dataset T, minimal support threshold sup, minimal all-item-confidence threshold L, minimal all-confidence threshold λ, and minimal item-set correlation threshold C.

Output: Item-item and between-set correlated association rules.

Methods

1 Tree←CFP-Construction(T, sup);
2 (P, s)←CFP-ItemCoMining(Tree, minsup, λ, L);
3 TrieTree_with_count←BuildTrieTree(P, s);
4 R=Rule-Generation(TrieTree_with_count);
5 Return R.

Function CFP-ItemCoMining(Tree, minsup, λ, L)

6 For each item $i\in$ItemTable from the least to the most frequent {
7 Set i as the root, call Construct_LocalItemCoTable(i), and set up the new local index; }
8 For each node i in CFP-Tree {
9 Initialize mappedTrans;
10 For each j in the path to the root of CFP-tree {
11 If $j\in$LocalItemCoTable mappedTrans=mappedTrans$\cup$GetIndex(j); }
12 Sort mappedTrans in ascending order of item ids, and call insert_LocalCFPTree(i); }
13 NonRecMine(i);
14 Traverse the LocalFreqAssoCoPatternTree, and check the measurements of $L(X)$ and all_conf(X). Generate associated and item-item positively correlated frequent patterns P and their supports s.

Function BuildTrieTree(P, s)

15 For each $p\in P$, insert p with its s into TrieTree_ with_count.

Function Rule-Generation(TrieTree_with_count)

16 Traverse TrieTree_with_count, for each $p\in$TrieTree_with_count {
17 Generate the candidate item-item and between-set correlated association rules r: $A\Rightarrow p-A$, $r\in \mathrm{RS}_k$; if $|A|>|p-A|$, $C(A\Rightarrow p-A)\leftarrow C(p-A\Rightarrow A)$; if k=2, $C(r)\leftarrow L(p)$;
18 If $C(r)<C$, delete r from R; }

5.3 Experimental evaluation

We applied VC6.0 to implement I&ISCoMine_AP and I&ISCoMine_CT. I&ISCoMine_AP adopts a high-performance trie-tree structure to store and deal with the candidate patterns. We compared the mining algorithms of the item-item and between-set correlated association rules (I&ISCoMine_AP and I&ISCoMine_CT) to the mining algorithm of both the association and correlation rules (ACR_Mining) proposed by Zhou *et al*. (2006c). The experiments were executed on a PC with AMD Sempron 2400+ 1.67 GHz, 512 MB RAM, and WinXP OS. The time taken for disk writing the generated rules was not considered. The first experiment was designed to test the performance of three algorithms using various datasets. The second experiment was used to test the pruning effect of the measurements. Finally, we used a real-life dataset to validate the proposed rules.

Experiment 1 (Performance on various datasets) The T10I4D100K and Connect4 datasets were adopted. T10I4D100K was generated by a generator provided by the IBM Almaden Laboratory Data Mining Research Group (IBM Almaden Research Center, 2009), and is intermediate between a sparse and an intensive dataset. Connect4 was downloaded from the UCI Machine Learning Repository (http://archive. ics.uci.edu/ml/), and is a very dense dataset.

First, while changing the minimal all-confidence threshold (λ), we set the minimal support (sup), the minimal all-item-confidence (L), and the minimal item-set correlation threshold (C) of I&ISCoMine_AP and I&ISCoMine_CT to be fixed, and also kept the minimal corr-confidence threshold (ω) of ACR_Mining fixed. Figs. 2 and 3 show that there were clear differences in performance between these three algorithms, with I&ISCoMine_CT giving the best performance, followed by I&ISCoMine_AP and then ACR_Mining. Because of its high performance, I&ISCoMine_CT gave a good result in a reasonable computation range. For the Connect4 dataset, when sup=0.85, L=0.5, C=0.5, ω=0, λ=0.92, I&ISCoMine_CT and I&ISCoMine_AP needed 3 s and 61 s respectively, while the execution time of ACR_Mining was 267 s.

In the following, sup, λ, C, and ω were fixed, while changing L. The ranking of the algorithms was also I&ISCoMine_CT>I&ISCoMine_AP>ACR_Mining (Figs. 4 and 5). For the T10I4D100K dataset, when sup=0.0003, λ=0.03, C=0.5, ω=0, and L=0.9, I&ISCoMine_CT and I&ISCoMine_AP needed 9 s and 26 s respectively, while the execution time of ACR_Mining was 65 s.

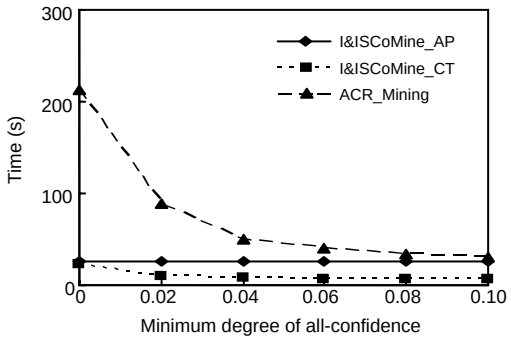

Fig. 2 Effect of all-confidence in the T10I4D100K dataset (sup=0.0003, L=0.9, C=0.5, ω=0)

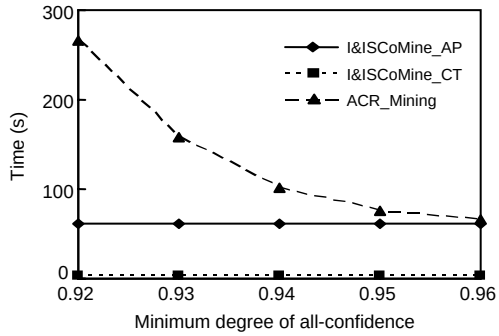

Fig. 3 Effect of all-confidence in the Connect4 dataset (sup=0.85, *L*=0.5, *C*=0.5, *ω*=0)

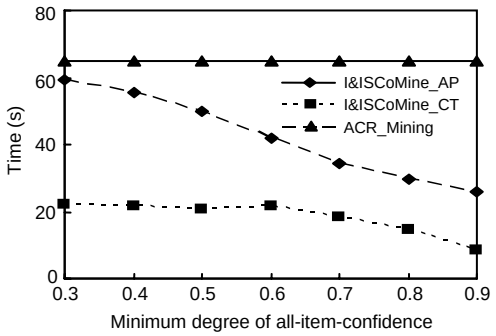

Fig. 4 Effect of all-item-confidence in the T10I4D100K dataset (sup=0.0003, *λ*=0.03, *C*=0.5, *ω*=0)

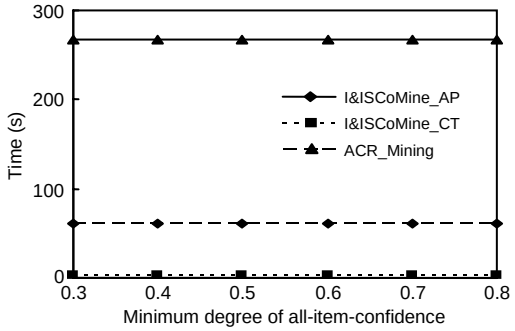

Fig. 5 Effect of all-item-confidence in the Connect4 dataset (sup=0.85, *λ*=0.92, *C*=0.5, *ω*=0)

Experiment 2 (Effect of varying parameters) We varied λ and kept the other parameters fixed. The performance of I&ISCoMine_AP and I&ISCoMine_CT did not change obviously with increasing λ, while the execution time of ACR_Mining decreased greatly with the variation in the all-confidence measurement (Figs. 2 and 3). This can be explained as follows. I&ISCoMine_AP and I&ISCoMine_CT have several measurements for pruning uninteresting patterns or rules. When L is 0.9, there is an apparent pruning effect. Thus, the varying of λ does not affect the performance of I&ISCoMine_AP and I&ISCoMine_CT too much. In contrast, ACR_Mining greatly depends on the all-confidence measurement for pruning.

We then varied L and fixed the other parameters. Comparing Figs. 4 and 5, we find that the effect differed between the datasets. As the all-item-confidence measurement increased, the execution time of I&ISCoMine_AP and I&ISCoMine_CT clearly dropped off in the T10I4D100K dataset, while the effect was insignificant in the Connect4 dataset. This is caused by the setting of λ. In Fig. 4, λ is not very large, and L shows its pruning effect. Thus, the execution time of I&ISCoMine_AP and I&ISCoMine_CT decreased with the increase in L. In Fig. 5, when λ was 0.92, it produced a significant pruning effect. Thus, the varying of L had little effect on the performance of I&ISCoMine_AP and I&ISCoMine_CT. For ACR_Mining, since there is no all-item-confidence measurement, the execution time did not change.

In the following, we tested the performance of the three algorithms with variation in C. The relative performance of the algorithms did not change (Fig. 6). This is because the item-set correlation measurement is a posterior test. It takes action when the candidate rules are produced and thus cannot reduce the search space.

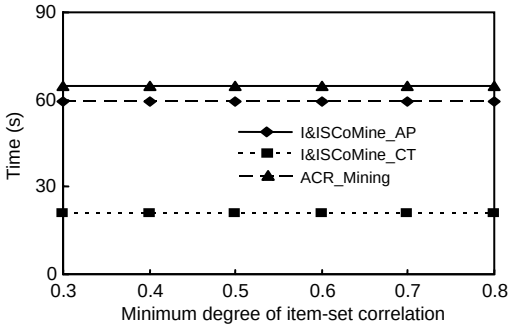

Fig. 6 Effect of item-set correlation in the T10I4D100K dataset (sup=0.0003, *λ*=0.03, *L*=0.3, *ω*=0)

Finally, we varied the minimal support threshold (sup), while the other measures were fixed. The execution time of the algorithms decreased with the increase in the support value (Fig. 7). Since ACR_Mining greatly depends on support measure to reduce the searching space, changing the support can significantly affect its performance. Although the performance of I&ISCoMine_AP and I&ISCoMine_CT

is also affected by support measure, the effect is not as significant as for ACR_Mining. The reason is that I&ISCoMine_AP and I&ISCoMine_CT have multiple measures which can prune uninteresting patterns or rules effectively.

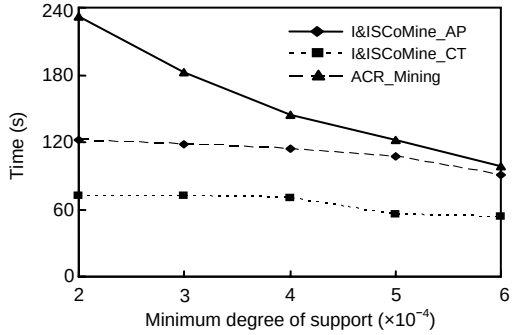

Fig. 7 Effect of support in the T10I4D100K dataset (λ=0.01, L=0, C=0, ω=0)

Experiment 3 (Pruning effect of measurements) For convenience, the association rules generated by the associated and item-item positively correlated frequent patterns are called item-item positively correlated association rules. The pruning effects of λ (sup=0.00025, L=0.3, C=0.4) and L (sup=0.00035, λ=0.03, C=0.7) in the T10I4D100K dataset are listed in Tables 5 and 6, respectively. Both λ and L performed well in pruning uninteresting patterns or association rules. With an increase in λ (L), the numbers of item-item positively correlated association rules and of item-item and between-set correlated association rules declined. The degree of decline depends on the distribution curve of λ (L) within the range. The number of association rules decreased dramatically after adding λ (L). If we add the measure C, the number decreased too, but only slightly.

Table 7 lists the pruning effect of the item-set correlation measurement (C) in the T10I4D100K dataset with sup=0.0003, λ=0.03, and L=0.3. Under these thresholds, the number of normal association rules was 1 351 728. After including the pruning of the item-set correlation measurement, the number of rules reduced to about 430 000, which is about 31.8% of the original number. After using the minimal all-item-confidence threshold L and the minimal all-confidence threshold λ, the rule number reduced further to around 360 000, about 26.6% of the original size. The changing of the numbers of the between-set correlated association rules and the item-item and between-set correlated association rules, with the variation of the item-set correlation threshold, shows that the pruning effect of this measurement is good.

Table 5 Pruning effect of all-confidence (λ) on the T10I4D100K dataset*

λ	N_1	N_2	N_3
0.02	1 456 602	472 100	470 664
0.03	1 456 602	361 218	360 256
0.04	1 456 602	231 766	231 206
0.05	1 456 602	165 620	165 300
0.06	1 456 602	129 134	128 996
0.07	1 456 602	78 826	78 764

* sup=0.00025, L=0.3, C=0.4. N_1: number of normal association rules; N_2: number of item-item positively correlated association rules; N_3: number of item-item and between-set correlated association rules

Table 6 Pruning effect of all-item-confidence (L) on the T10I4D100K dataset*

L	N_1	N_2	N_3
0.2	1 143 818	421 866	414 638
0.3	1 143 818	360 696	354 198
0.4	1 143 818	306 816	301 288
0.5	1 143 818	244 368	240 198
0.6	1 143 818	148 948	146 574
0.7	1 143 818	67 378	67 378

* sup=0.00035, λ=0.03, C=0.7. N_1: number of normal association rules; N_2: number of item-item positively correlated association rules; N_3: number of item-item and between-set correlated association rules

Table 7 Pruning effect of item-set correlation measurement (C) on the T10I4D100K dataset*

C	N_1	N_4	N_3
0.3	1 351 728	431 586	360 926
0.4	1 351 728	430 624	359 964
0.5	1 351 728	429 264	358 604
0.6	1 351 728	427 466	356 806
0.7	1 351 728	425 054	354 414
0.8	1 351 728	421 902	351 340
0.9	1 351 728	409 702	342 920

* sup=0.0003, λ=0.03, L=0.3. N_1: number of normal association rules; N_3: number of item-item and between-set correlated association rules; N_4: number of between-set correlated association rules

After eliminating the association rules in which the antecedent and consequent item-sets are negatively correlated, the quality of the obtained item-item and between-set correlated association rules is better than that of the original association rules.

Experiment 4 (Test on a real-life retailing dataset) We applied the item-item and between-set correlated association rules to a real-life retailing dataset called Retail (Xiong *et al*., 2006). Retail is a market-basket dataset obtained from a large mail-order company and was provided by Dr. Xiong with Rutgers University, USA.

We set sup=0.0001, λ=0.5, L=0.9, and C=0.9 to mine the Retail dataset, and obtained a group of item-item and between-set correlated association rules. Take one of the rules as an example:

$$\{\text{Adidas tights}\} \Rightarrow \{\text{Adidas fleece Jacket}\}$$
$$\{\text{sup}=0.0001, \lambda=0.75, L=0.999, C=0.999\}.$$

Although the support of this rule was low, it had a high all-confidence degree, all-item-confidence degree, and item-set correlation degree. Therefore, this rule can be applied not only to the symmetrical applications of the antecedent and consequent item-sets, but also to unsymmetrical applications, such as a cross-selling strategy; this will achieve a better effect than a normal association rule.

6 Conclusions

Because the association rules mining methods based on frequent patterns or correlated frequent patterns cannot completely eliminate suspicious cross-support patterns, spurious association rules containing negatively correlated items, or spurious association rules in which the antecedent and consequent item-sets are negatively correlated, in this paper, we propose a new method to mine the item-item and between-set correlated association rules.

The all-item-confidence measure means that if the correlation degree of any two items in the pattern is not less than the minimal threshold for all-item-confidence, this pattern can be regarded as an interesting pattern. It has many good properties such as appropriate boundary and anti-monotone property. We choose all-item-confidence as the correlation measurement of the pattern to eliminate spurious association rules that contain negatively correlated items. The all-confidence measure is chosen as the association measurement to filter suspicious cross-support patterns, and the item-set correlation measurement is used to eliminate spurious association rules in which the antecedent and consequent item-sets are negatively correlated. Thus, the item-item and between-set correlated association rules can be obtained. We propose two mining algorithms, I&ISCoMine_AP and I&ISCoMine_CT, and tested their performance and the pruning effect of the item-set correlation measurement. Experimental results showed that the proposed method is effective and valid.

The application of item-item and between-set correlated association rules focuses mainly on marketing strategies, such as shelf and inventory arrangement and cross-selling. The applicability of this kind of association rule should be studied in other application domains.

Acknowledgements

We would like to thank the anonymous reviewers for valuable comments and useful suggestions, and Dr. Hui XIONG for providing us with the Retail dataset.

References

Agrawal, R., Imielinski, T., Swami, A., 1993. Mining association rules between sets of items in large databases. *ACM SIGMOD Rec.*, **22**(2):207-216. [doi:10.1145/170035.170072]

Alvarez, S.A., 2003. Chi-Squared Computation for Association Rules: Preliminary Results. Technical Report No. BC-CS-2003-01, Computer Science Department, Boston College, MA.

Brin, S., Motwani, R., Silverstein, C., 1997. Beyond market baskets: generalizing association rules to correlations. *ACM SIGMOD Rec.*, **26**(2):256-276. [doi:10.1145/253260.253327]

Hahsler, M., Hornik, K., 2007. New probabilistic interest measures for association rules. *Intell. Data Anal.*, **11**(5):437-455.

Han, J., Pei, J., Yin, Y., Mao, R., 2004. Mining frequent patterns without candidate generation: a frequent-pattern tree approach. *Data Min. Knowl. Disc.*, **8**(1):53-87. [doi:10.1023/B:DAMI.0000005258.31418.83]

IBM Almaden Research Center, 2009. Quest Synthetic Data Generation Code. Available from http://www.almaden.ibm.com/cs/projects/iis/hdb/Projects/data_mining/datasets/syndata.html [Accessed on Jan. 21, 2009].

Kenett, R.S., Salini, S., 2008a. Relative linkage disequilibrium: a new measure for association rules. *LNCS*, **5077**:189-199. [doi:10.1007/978-3-540-70720-2_15]

Kenett, R.S., Salini, S., 2008b. Relative linkage disequilibrium applications to aircraft accidents and operational risks. *IEEE Trans. Mach. Learn. Data Min.*, **1**(2):83-96.

Kim, W.Y., Lee, Y.K., Han, J., 2004. CCMine: efficient mining of confidence-closed correlated patterns. *LNAI*, **3056**: 569-579. [doi:10.1007/b97861]

Lee, Y.K., Kim, W.Y., Cai, Y.D., Han, J., 2003. CoMine: Efficient Mining of Correlated Patterns. Proc. 3rd IEEE Int. Conf. on Data Mining, p.581-584. [doi:10.1109/ICDM.2003.1250982]

Lenca, P., Meyer, P., Vaillant, B., Lallich, S., 2008. On selecting interestingness measures for association rules: user oriented description and multiple criteria decision aid. *Eur. J. Oper. Res.*, **184**(2):610-626. [doi:10.1016/j.ejor.2006.10.059]

Liu, B., Hsu, W., Ma, Y., 1999. Pruning and Summarizing the Discovered Associations. ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, p.125-134. [doi:10.1145/312129.312216]

Lu, H., Han, J., Feng, L., 1998. Stock Movement Prediction and *N*-dimensional Inter-transaction Association Rules. Proc. 3rd ACM-SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, p.1-7.

Omiecinski, E.R., 2003. Alternative interesting measures for mining associations in databases. *IEEE Trans. Knowl. Data Eng.*, **15**(1):57-69. [doi:10.1109/TKDE.2003.1161582]

Ozden, B., Ramaswamy, S., Silberschatz, A., 1998. Cyclic Association Rules. Proc. 14th Int. Conf. on Data Engineering, p.412-421. [doi:10.1109/ICDE.1998.655804]

Palshikar, G.K., Kale, M.S., Apte, M.M., 2007. Association rules mining using heavy itemsets. *Data Knowl. Eng.*, **61**(1):93-113. [doi:10.1016/j.datak.2006.04.009]

Qin, M., Hwang, K., 2004. Frequent Episode Rules for Internet Anomaly Detection. Proc. 3rd IEEE Int. Symp. on Network Computing and Applications, p.161-168. [doi:10.1109/NCA.2004.1347773]

Ruggeri, F., Kenett, R.S., Faltin, F.W., 2007. Encyclopedia of Statistics in Quality and Reliability. Wiley, Chichester, England.

Shen, B., Yao, M., 2009a. Mining associated and item-item correlated frequent patterns. *J. Zhejiang Univ. (Eng. Sci.)*, **43**(12):2171-2177 (in Chinese). [doi:10.3785/j.issn.1008-973X.2009.12.008]

Shen, B., Yao, M., 2009b. A new kind of dynamic association rule and its mining algorithms. *Control Dec.*, **24**(9):1310-1315 (in Chinese).

Shen, B., Yao, M., Wu, Z.H., Gao, Y.J., 2010. Mining dynamic association rules with comments. *Knowl. Inform. Syst.*, **23**(1):73-98. [doi:10.1007/s10115-009-0207-1]

Tan, P.N., Kumar, V., Srivastava, J., 2002. Selecting the Right Interestingness Measure for Association Patterns. Proc. 8th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, p.32-41. [doi:10.1145/775047.775053]

Xiong, H., Tan, P.N., Kumar, V., 2006. Hyperclique pattern discovery. *Data Min. Knowl. Disc.*, **13**(2):219-242. [doi:10.1007/s10618-006-0043-9]

Zhang, S., Chen, F., Wu, X., Zhang, C., Wang, R., 2006. Identifying Bridging Rules Between Conceptual Clusters. Proc. 12th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, p.815-820. [doi:10.1145/1150402.1150509]

Zhang, S., Chen, F., Jin, Z., Wang, R., 2009. Mining class-bridge rules based on rough sets. *Exp. Syst. Appl.*, **36**(3):6453-6460. [doi:10.1016/j.eswa.2008.07.044]

Zhou, Z.M., Wu, Z.H., Wang, C.S., Feng, Y., 2006a. Mining both associated and correlated patterns. *LNCS*, **3994**:468-475. [doi:10.1007/11758549]

Zhou, Z.M., Wu, Z.H., Wang, C.S., Feng, Y., 2006b. Efficiently mining mutually and positively correlated patterns. *LNCS*, **4093**:118-125. [doi:10.1007/11811305]

Zhou, Z.M., Wu, Z.H., Wang, C.S., Feng, Y., 2006c. Efficiently mining both association and correlation rules. *LNCS*, **4223**:369-372. [doi:10.1007/11881599]