International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-2, Issue-4, July- 2014

Mining Positive and Negative Fuzzy Association Rules with Item Cost

Ruchi Bhargava, Prof. Shrikant Lade, Prof. Daya Shnakar Panday

***Abstract*— While ancient algorithms concern positive associations between binary or quantitative attributes of databases, this paper focuses on mining each positive and negative fuzzy association rules. This work tends to show however, by a deliberate selection of formal logic connectives considerably hyperbolic expressivity is on the market at very little additional value. Ancient algorithms for mining association rules area unit engineered on the binary attributes databases, that as few limitations. Firstly, it cannot concern quantitative attributes; second, solely the positive association rules area unit discovered; third, it treats every item with a similar frequency though completely different item might have different frequency. during this paper, argue a discovery algorithmic rule for mining positive and negative fuzzy association rules to resolve these 3 limitations. Novel approach is given for effectively mining weighted fuzzy association rules (ARs). This paper solve the matter of mining weighted association rules, exploitation associate degree improved model of weighted support and confidence framework for classical and fuzzy positive and negative association rule mining.**

***Index Terms*— Association rules, fuzzy, weighted support, weighted confidence, Fuzzy association rules. Positive negative rules.**

I. Introduction

Association rules (ARs) are wide wont to confirm client shopping for patterns from market basket information. The task of mining association rules is especially to get association rules (with robust support and high confidence) in largedatabases. Traditional Association Rule Mining (ARM) algorithms treat through the relationships among the items exists in transactional databases item records.

Association rules mining is a crucial analysis topic in data processing and information discovery. associate association rule is created as A→B, wherever A and B are disjoint itemsets, and its support is not any but a user-specified minimum support. Since this sort of correlation is positive, we have a tendency to decision it positive association rule.

Manuscript received July 08, 2014.

Ruchi Bhargava, Information Technology RKDF Institute, Bhopal, India. (e-mail: ruchibhargava17@gmail.com).

Prof. Shrikant Lade, Information Technology RKDF Institute, Bhopal, India. (e-mail: shrikantlade@gmail.com).

Prof. Daya Shnakar Panday, Information Technology RKDF Institute, Bhopal, India (e-mail: er_daya@rediffmail.com).

Contrasted to positive association rules, mining negative association rules is planned in literature. The negative association rule is shaped as A→¬B, which suggests that if A is in an exceedingly transaction, then B wouldn't within the same transaction with high chance. There area unit different forms negative association rules like ¬A→B and ¬A→¬B.

Association rules (ARs) [11] are wide wont to confirm client shopping for patterns from market basket knowledge. The task of mining association rules is especially to find association rules (with robust support and high confidence) in massive databases. Classical Association Rule Mining (ARM) deals with the relationships among the items present in transactional databases [9, 10]. The standard approach is to initial generate all massive (frequent) itemsets (attribute sets) from that the set of ARs comes. an oversized itemset is defined jointly mutually additional oftentimes within the given information set than a user provided support threshold. To limit the amount of ARs generated a confidence threshold is employed. the amount of ARs generated will thus be influence by careful choice of the support and confidence thresholds, but care should be taken to make sure that itemsets with low support, however from that high confidence rules could also be generated, aren't omitted.

Given a group of items I= and a database of transactions D = wherever ti = I with K = |X|⊆I, if X∊p≤m and Iij is named a k -itemset or just associate itemset. Consider a transaction record D be a multi-set of subsets of I as shown. Every D∊T supports associate itemset if I⊆X and T⊆X holds. an association rule is an expression X => Y, where X, Y are Y=ø holds.∩item sets and X variety of transactions T supporting an item X w.r.t D is named support of Sup(X) = |D∊T|/|D|. The strength or confidence (c) for an association rule X => Y is that the magnitude relation of the quantity of transactions that contain X U Y to the quantity of transactions that contain X, Conf (X →Y) = Supp (X U Y)/ Supp (X). For non-boolean items fuzzy association rule mining was planned exploitation fuzzy sets specified quantitative and categorical attributes are often handled [12]. A fuzzy quantitative rule represents every item as (item, value) combine. Fuzzy association rules ar expressed within the subsequent type:

If X might be A assure Y is B.

For instance, if (experience is fresher) => (income is less)

Given a database T, attributes I with itemsets X⊆I Y⊆I, and X = { x1, x2, ….., xn} and Y = {y1, y2, …., yn} and X∩Y=ø, we can define fuzzy sets A = {fx1, fx2, ….

Fxn} and B = {fx1, fx2, …….,fxn) associated to X and Y respectively. For example (X, Y) could be (experience, fresher), (experience, more), (income, high) etc. The semantics of the law is to facilitate when the ancestor "X is A" is fulfilled, we be able to mean that "Y is B" is also fulfilled, which clear there are enough records that give their take part in an election to the feature fuzzy set pairs and the computation of these votes is larger than the user specific threshold.

However, the classical ARM framework assumes that every one thing have identical significance or importance. That during which case their item value inside a dealings or record is that the same (item cost=1) which isn't continually the case. as an example, from Table one, the rule [printer →computer, 50%] is also additional necessary than [scanner →computer, 75%] despite the fact that the previous holds a lower support as a result of those items within the 1st rule typically go along with additional profit per unit sale. the most challenge in item value ARM is corroborative that is crucial for the economical repetitive method of generating and pruning frequent item sets from subsets. The holding thought of each frequent item set means their subsets also are frequent. This section, tend to address the problem of ARM in Item value ARM.

Table 1. Weighted Items Database

ID	Item	Profit	Weight	…
1	Scanner	10	0.1	…
2	Printer	30	0.3	…
3	Monitor	60	0.6	…
4	Computer	90	0.9	…

Table 2.Transactions

TID	Items
1	1,2,4
2	2,3
3	1,2,3,4
4	2,3,4

Weighted ARM deals with the importance of individual items in an exceedingly info [2, 3, 4]. For instance, some products are a lot of profitable or is also below promotion, thus a lot of attention-grabbing as compared to others, and therefore rules regarding them are of larger price. Items are allotted cost (C) per their significance as shown in table one. These weights are also set per Associate in nursing item's margin of profit. This generalized version of ARM is named Weighted Association Rule Mining (WARM) . From table one, we are able to see that the rule computer→Printer is a lot of fascinating than Computer → Scanner as a result of the profit of a printer is larger than that of a scanner. the most challenge in weighted ARM is that "downward closure property" that is crucial for economical repetitious method of generating and pruning frequent item sets from subsets. during this paper we have a tendency to address the problem of downward closure property in heat. We have a tendency to generalize and solve the matter of downward closure property and propose a weighted support as well as confidence structure for each Boolean in addition to quantitative items for standard and fuzzy rules mining (FWARM).

II. Problem definition

Let I = {i1, i2, … ,im} be a group of literals known as items. Let the database D = be a group of transactions, wherever every group action may be a set of I. A non-empty set of I is named item set. AN item set containing k items is named k-item set. The support of an item set X denoted as sup(X) is outlined because the range of transactions containing X in D. an item set is frequent if its support is larger than a user-specified threshold minimum support minsup. an association rule is an expression of the from A⇒B, wherever X and Y square measure sets of items, A∩B=ø.

The support of the corresponding to a positive association rule like A→B, there are three attainable negative association rules, A→¬B, ¬A→B and ¬A→¬B. For a negative association rule A→¬B and a particular transaction T, if A →T and ¬B→T, we are saying that the transaction T supports A→¬B. Assume there's a negative association rule like ({i1}, ¬{ i2, i3}), which suggests that if i1 is in an exceedingly transaction T, i2 and i3would not appeared within the transaction T at same time, however there's a clear stage that one amongst the i2 and i3is in transaction T. to get negative association rule, we'd like to think about all of the attainable item sets in transaction databases. If A→¬B may be a negative association rule, it'll hold that sup(A→¬B)>= min sup. a better price for min sup presumably suggests that sup(A→B) rare sequence. However, there are too several infrequent sequences in database. If A may be a frequent item set whereas B may be a infrequent item set with support one, we'll have: sup(A) > = min sup, sup(B)≈0, sup(A→¬B) ≈ sup(A) > = minsup. Therefore, it appears that A→¬B may be a negative association rule. In fact, this sort of sequences is very prevailing in real database, as an example, a group of the products seldom bought by customers in market is an infrequent item set. In observe, since the task of knowledge mining is to search out all types of valuable correlations, we tend to sometimes additional specialize in the correlations between the well-sold product, that are supported the frequent sequence. In different word, if A→¬B, ¬A→B and ¬A→¬B are negative association rules, A and B would be frequent sequence. In usually speaking, we tend to solely specialize in the frequent sequence whether or not the association rules are positive or negative.

III. Related Work

Classical ARM information items are viewed as having equal importance however recently some approaches generalize this wherever items are given weights to replicate their significance to the user [17]. The weights could correspond to special promotions on some merchandise or the gain of various items etc. Currently, two approaches exist: pre-and post-processing. Post process solves initial the non-weighted downside (weights=1 per item) and so prune s the principles later. Pre-processing prunes the non-frequent item sets earlier using weights once each iteration. The problem post-processed weighted ARM is that first; items are scanned while not considering their weights.

Finally, the rule base is checked for frequent weighted ARs. this offers us a awfully restricted item set pool to examine weighted A Rs and will miss several potential item sets. In pre-processing, classical ARM prunes item sets by checking frequent ones against weighted support once each scan. In pre -processing, less rules are obtained as compared to post process as a result of several potential frequent super sets are lost.

In [16] a post-processing model is projected. two algorithms were planned to mine item sets with normalized and un-normalized weights. The K-support certain metric was wont to guarantee validity of the downward closure property. Even that didn't guarantee each set of a frequent set being frequent unless the k-support certain worth of (K- 1) set was higher than (K).

Paper [14] addressed three limitations of ancient association rules mining: crisp item sets, solely positive association rules and every item with constant frequency. The fuzzy extension of crisp item sets ends up in approaches of mining fuzzy association rules, whereas negative association rules will be discovered as positive association rules are mined , and item with its support extension conduces association rules mining with multiple minimum supports. This paper suggests an method for mining each positive and negative fuzzy association rules by combining these three extensions.

Author in paper [15] proposes a brand new scheme for with efficiency mining positive and negative association rules in a very transaction database. The rule is termed PNAR_IMLMS and is suitable for mining positive association rules from frequent item sets and negative association rules from each frequent and rare item sets discovered by the IMLMS model. The IMLMS model adopted an efficient pruning technique to prune uninteresting item sets

IV. ALGORITHM FOR MINING POSITIVE AND NEGATIVE FUZZY ASSOCIATION RULES CITH MULTIPLE MINIMUM SUPPORTS

A. Positive and Negative Fuzzy Association rules

In this section, we present the definitions of both positive and negative item cost fuzzy association rules with the assumption that transaction data are fuzzy. Assume μx is the membership function of x for all $x \in I$. For each transaction $t \in D$, $\mu x(t)$ represents the degree that t contains the item x. The positive and negative item cost fuzzy Association rules formed as $A \Rightarrow B$, $A \Rightarrow \neg B$, $\neg A \Rightarrow B$ and $\neg A \Rightarrow \neg B$ are described as the follows

1) Positive Fuzzy Association rules

We take the support of item set X as the number of transaction in D that contains X, which is denoted as sup(X). The minimum support is denoted as min sup. Let A and B be two item sets.

Definition 1: The $A \Rightarrow B$ is positive fuzzy association rule, if the following conditions hold:

(1) $A \cap B = \emptyset$;
(2) $sup(A \cup B) - sup(A) \times sup(B) >=$ mininterest
(3) $sup(A \cup B) = \sum_{t \in SD} \prod_{x \in A} \mu_x(t) \prod_{y \in B} \mu_y(t) \geqq$ minsup;
(4) $Conf(A \Rightarrow B) \geqq$ minconf.

2) Negative Fuzzy Association rules

Let A and B be two itemsets, if $A \Rightarrow \neg B$ is a negative association rule, both A and B must be frequent, which means that their support should be not less than the support threshold, while $A \cup B$ should be infrequent. The three types of negative fuzzy association rules can be defined as follows.

Definition 2: $A \Rightarrow \neg B$ is a negative fuzzy association rule, if the following conditions hold:
(1) $A \cap B = \emptyset$;
(2) $sup(A) >= minsup$, $sup(B) >= minsup$;
(3) $sup(A \cup \neg B) - sup(A) \times sup(\neg B) >=$ mininterest
(4) $sup(A \cup \neg B) = \sum_{t \in D} \prod_{x \in A} \mu_x(t) \prod_{y \in B} (1 - \mu_y(t)) >=$ wminsup;
(5) $Conf(A \Rightarrow \neg B) = sup(A \cup \neg B) / sup(A) = sup(A) - sup(A \cup B) / sup(A) \geqq$ minconf.

Definition 6: $\neg A \Rightarrow B$ is a negative fuzzy association rule, if the following conditions hold:
(1) $A \cap B = \emptyset$;
(2) $sup(A) >= minsup$, $sup(B) >= minsup$;
(3) $sup(\neg A \cup B) - sup(\neg A) \times sup(B) >=$ mininterest
(4) $sup(\neg A \cup B) = \sum_{t \in D} \prod_{x \in A} (1 - \mu_x(t)) \prod_{y \in B} \mu_y(t) >=$ minsup;
(5) $Conf(\neg A \Rightarrow B) = sup(\neg A \cup B) / sup(A) = sup(B) - sup(A \cup B) / (1 - sup(A)) \geqq$ minconf.

Definition 3: $\neg A \Rightarrow \neg B$ is a negative fuzzy association rule, if the following conditions hold:
(1) $A \cap B = \emptyset$;
(2) $sup(A) >= minsup$, $sup(B) >= minsup$;
(3) $sup(\neg A \cup \neg B) - sup(\neg A) \times sup(\neg B) >=$ mininterest
(4) $wup(\neg A \cup \neg B) = \sum_{t \in D} \prod_{x \in A} (1 - \mu_x(t)) \prod_{y \in B} (1 - \mu_y(t)) >=$ mins
(5) $Conf(\neg A \Rightarrow B) = sup(\neg A \cup \neg B) / sup(\neg A) = 1 - sup(A) - sup(B) + sup(A \cup B) / (1 - sup(A)) \geqq$ minconf.

Our mining scheme first transforms every quantitative worth into a fuzzy set with linguistic terms applying membership functions. It then calculates the scalar cardinality of every linguistic term on all the transaction information. every item uses solely the linguistic term with the most cardinality in later mining processes, so creating range of fuzzy regions to be processed a similar because the number of original method. The algorithmic program thus focuses on method necessary linguistic terms that reduce its time complexity. The mining method supported fuzzy counts is then performed to search out fuzzy association rules from these giant item sets.

B. Algorithm: MPNFARMIC (Mining Positive and Negative Fuzzy Association rules using Item Cost)

Input: A body of n transaction data, each consists of customer ID, the purchased items with their quantities, a set of membership functions and a predefine least minimum support threshold LS, minimum fuzzy confidence threshold mini conf and C cost of each item in dataset;

Output: A set of positive item cost fuzzy association rules: FPAR;

A set of negative item cost fuzzy association rules: FNAR;

Initialize FPAR=ø, FNAR=ø

Calculate support of each item present in dataset as Sup(X) for more variable Sup(XY) with item cost as Eq.

$$Sup\ (X) = \sum_{t \epsilon D} Sup(X,t,c) = \sum_{t \epsilon D} \prod_{i=1}^{k} \mu_x(t) \times C_i$$

(3) C1= {candidate 1-itemset};

(4) For each item c in C1 do

$$Sup\ (X) = \sum_{t \epsilon D} Sup(X,t,c) = \sum_{t \epsilon D} \prod_{i=1}^{k} \mu_x(t) \times C_i$$

(5) If A frequent itemset T: support(I) >= minsupp(x)
else An infrequent itemset NT: support(J) < minsupp(x)

(6) Positive association rules
For every element in T calculate
If (Supp(X U Y) ≥ minsupp)
Conf(X,Y)= Supp(X U Y) / supp(X)
If(Conf(X,Y)≥ minconf)
FPAR= {FPAR ∪ X→Y}

(7) For every element in NT calculate
Supp(AU~B)=supp(A)-supp(AUB)
If (Supp(A) >= minsupp & supp(B) > minsupp & supp(A U ~B) >= minsupp)
Conf(AU~B)=supp(AU~B)/supp(B)
if(Conf(A,B)> minconf)
FNAR= {FPNR ∪ A→~B}

(8) Output FPAR and FNAR

(9) Return

1) Support calculation

For single items sets the support is the sum of the product calculation for each item cost/fuzzy member ship pair (c*f). For 2-itemsets and larger the support is the sum of the products of all the costs and fuzzy membership calculations. Thus given the data set below:

<c,1.0>
<a,0.25> <b,0.5>
<a,0.5> <c,0.75>
<a,0.5> <b,0.25>

and the associated Item cost file:
0.2
1.0
0.3

The support calculations will be as follows:

{a}= ((0.25*0.2)+(0.5*0.2)+(0.5*0.2))/4= *(0.05 + 0.1 + 0.1)/4 = 0.0625

{b} = ((0.5*1.0)+(0.25*1.0))/4 = (0.5+0.25)/4 = 0.1875

{c} = ((1.0*0.3)+(0.75*0.3))/4 = (0.3+0.225)/5 = 0.13125

{a, b} = ((0.25*0.2*0.5*1.0) + (0.5*0.2*0.25*1.0))/4 = (0.025+0.025)/4 = 0.0125

{a, c} = ((0.5*0.2*0.75*0.3))/4 = 0.0225/4 = 0.005625

V. EXPERIMENT AND RESULTS

To test the performance of our planned scheme we've got done some experiments. The method is implemented with C++. The artificial experiment information set is generated by Assocgen [18] program of IBM Almaden research facility. projected work is implemented in matlab 2010a on platform intel I3 processor with 2GB RAM using window 7 operation system. The meanings of used parameters ar showed in Table 3.

Table III. PARAMETERS

Symbol	Meaning
D	Number of customers(=size of database)
C	Average number of transactions per Customer
T	Average number of items per Transaction
NI	Number of maximal potentially large Item sets
N	Number of items

We place factors C=10, T=5, NI=2500, N =10000, total number of consumers D=100000, and the produced dataset is known as C10T5I25. The MPNFARMIC method is used in proposed experiment are shown as figure 1.

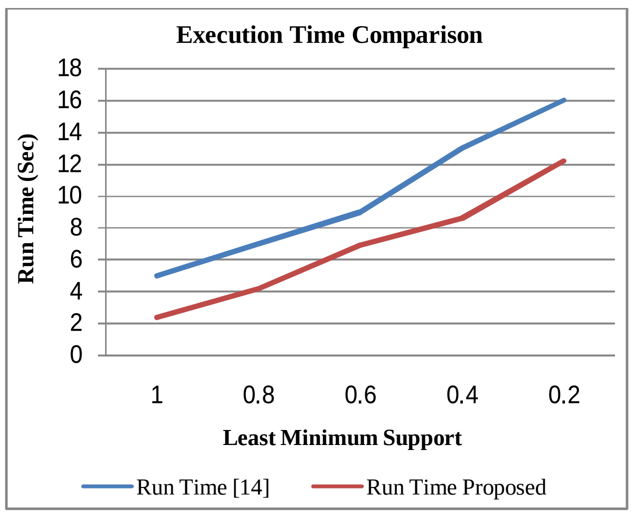

Fig. 1. Execution time

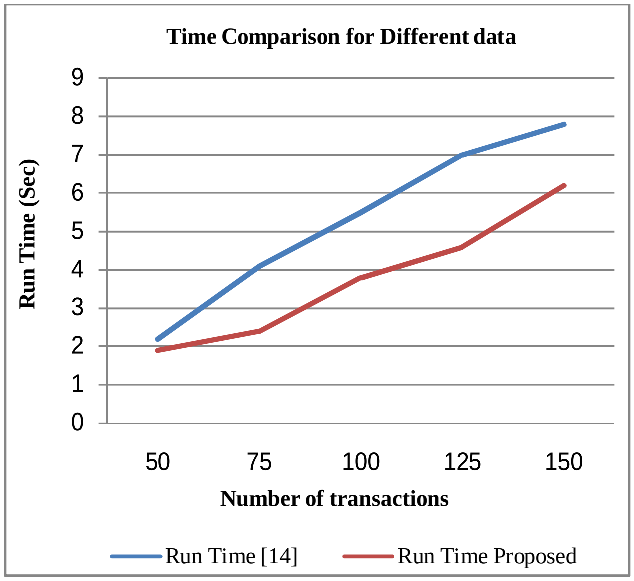

Fig. 2. Scale-up: Number of transactions

Figure 1 explains the algorithm accomplish time variation among smallest minimum support decreasing from 1% to 0.2%. It make obvious that the algorithm increases with the declining of LS it's in compare with existing method given in paper [14].

To observe the scalability of algorithm we increased the quantity of transactions from 50,000 to 150000, through LS=1%. The time evaluation for different dataset with [14] results are revealed in Figure 2. The executing time is increased almost linearly with the increasing of dataset size and gives better result in comparison. It can be concluded that our algorithm has a good scalable performance.

VI. Conclusion

We have given a weighted support and confidence framework for mining weighted association rules. we tend to used fuzzy positive and negative association rules mining with item cost to unravel the difficulty of in weighted ARM. we tend to generalized the fuzzy association rule mining and proposed a fuzzy weighted PANARM framework. The matter mentioned is resolved applying improved model of weighted support and confidence framework for fuzzy association rule mining. However, the data provided by alternative rules the information lost though the opposite rules represent the record in an exceedingly less degree of matching. it might be attention-grabbing to switch the proposed approach by as well as multiple rules for every record to search out the fuzzy PANARM applying item cost. Thus proposed method able to drive infrequent rules as well which not get by other existing methods. The time result comparison gives effectiveness of proposed method.

References

[1] Tao, F., Murtagh, F., Farid, M.: Weighted Association Rule Mining Using Weighted Support and Significance Framework. In: Proceedings of 9th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp.661–666, Washington DC(2003).

[2] Cai, C.H., Fu, A.W-C., Cheng , C. H ., Kwong, W.W .:Mining Association Rules with Weighted Items. In: Proceedings of 1998 Intl. Database Engineering and Applications Symposium (IDEAS'98), pages 68--77, Cardiff, Wales, UK, July 1998

[3] Wang, W., Yang, J., Yu, P. S.: Efficient Mining of Weighted Association Rules (WAR). In: Proceedings of the KDD, Boston, MA, August 2000, pp. 270-274

[4] Lu,S., Hu,H., Li, F.: Mining Weighted Association Rules, Intelligent data Analysis Journal, 5(3), 211--255(2001)

[5] Wang , B -Y., Zhang, S-M .: A Mining Algorithm for Fuzzy Weighted Association Rules. In: IEEE Conference on Machine Learning and Cybernetics, 4, pp.2495--2499(2003)

[6] Gyenesei, A.:Mining Weighted Association Rules for Fuzzy Quantitative Items , Proceedings of PKDD Conference pp. 416--423(2000).

[7] Shu, Y. J.,Tsang, E., Yeung, Daming, S.:Mining Fuzzy Association Rules with Weighted Items, IEEE International Conference on Systems, Man, and Cybernetics, (2000).

[8] Lu, J-J.: Mining Boolean and General Fuzzy Weighted Association Rules in Database s, Systems Engineering -Theory & Practice, 2, 28--32 (2002)

[9] Agrawal, R., Srikant, R.:Fast Algorithms for Mining A ssociation Rules.In: 20th VLDB Conference, pp. 487--499 (1994)

[10] Bodon, F.: A Fast Apriori implementation. In:ICDM Workshop on Frequent Itemset Mining Implementations, vol. 90, Melbourne, Florida, USA(2003)

[11] Agrawal, R., Imielinski, T., Swami , A.: Mining Association Rules Between Sets of Items in Large Databases . In: 12th ACM SIGMOD on Management of Data, pp. 207--216 (1993)

[12] Kuok, C.M., Fu, A., Wong, M.H.: Mining Fuzzy Association Rules in Databases. SIGMOD Record, 27, (1), 41--46 (1998)

[13] Coenen, F., Leng , P., Goulbourne,G.:Tree Structures for Mining Association Rules. Data Mining and Knowledge Discovery, 8(1) 25--51(2004)

[14] Weimin Ouyang, "Mining Positive and Negative Fuzzy Association Rules with Multiple Minimum Supports", International Conference on Systems and Informatics (ICSAI 2012), Pages: 2242-2246, IEEE, 2012.

[15] Idheba Mohamad Ali O. Swesi, Azuraliza Abu Bakar, Anis Suhailis Abdul Kadir, "Mining Positive and Negative Association Rules from Interesting Frequent and Infrequent Itemsets", 9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2012), Pages: 650-655, IEEE, 2012.

[16] Dong, X., Niu, Z., Zhu, D., Zheng, Z., Jia,Q.: Mining Interesting Infrequent and Frequent Itemsets Based on MLMS Model. International Conference on Advanced Data Mining and Applications, ADMA, IEEE, 2008.

[17] PENG Hui-ling, SHU Yun-xing, "A New FP-tree-based Algorithm MMFI for Mining the Maximal Frequent Itemsets", Page 61-64, 2012 IEEE.

[18] [18] Agrawal R,Srikant R. Mining association rules. In the Proc.1995 International Conference on Data Engineering, Taibei, Taiwan, March 1995.