International Journal of Hybrid Information Technology
Vol. 9, No.11 (2016), pp. 35-46
http://dx.doi.org/10.14257/ijhit.2016.9.11.04

The Research of a Novel Weighted Association Rules Algorithm Based On Unidirectional FP Tree

Fengxia Yang[1] * and Kehao Cao[2]

** Department of Information Engineering, Henan College of Finance & Taxation Zhengzhou, Henan, 451464, China 1 82848960 @qq.com*
[2] The Career Center, Henan College of Finance & Taxation Zhengzhou, Henan, 451464, China

Abstract

Weighted association rules are to research on association rule mining expansion. It is aimed at users of different project importance of different database mining, to help users find interesting association rules, but also does not produce too many useless rules. The most important process of weighted association rules mining is weighted to generate frequent item sets. Unidirectional FP- tree node entry's serial number marking and recording the support count. Item number by the support from big is to small order determination. The paper presents the research of a novel weighted association rules algorithm based on unidirectional FP tree. Implementation results show that the algorithm unidirectional FP running efficiency is relatively high.

***Keywords**: Weighted association rules, Unidirectional FP tree, data mining*

1. Introduction

The study of association rules is an important branch of data mining, has attracted a lot of scholars. This paper introduces the related knowledge of data mining technology, especially the mining of association rules, and introduces the basic concept and the classical algorithm. Past research often think database importance of each item is the same. In fact, importance of the items to the user is different. Therefore the existing algorithm of excavated is not necessarily our interest rules. In view of this situation, put forward weighted association rules.

Weighted association rules research background in traditional algorithms of mining association rules in database items that the important degree of the same, however in reality the importance of each item is different. For example, decision makers tend to give priority to high profit item, and ignore the less profitable project [1]. In addition, the passage of time and changes in consumption habits will also affect the association rules, time interval shorter affairs despite the support degree is not very high, but is likely to reflect the new consumption trend. In view of the limitation of association rules, weighted association rules is proposed.

This paper proposes a new weighted association rule model, and based on the Arr algorithm proposed a based on the small support weighted association rules mining algorithm by, for each data item through setting is not the same weight and minimum support, and solve the transaction number from the database data item to each are not identical, the actual rate of occurrence frequency and the problem of uneven distribution. The support can be avoided by a single multiple support degree caused by the occurrence frequency of difficult to the lower association rules.

Based on the FP- tree to generate frequent item set FP-growth algorithm, the algorithm first database all frequent information compression and storage on the FP- tree, so the database mining problem into FP- tree mining problem; then through frequent pattern mining frequent item set. Frequent pattern growth through suffixation itemsets and

ISSN: 1738-9968 IJHIT

conditional pattern tree generated frequent itemsets connection. The method reduces the search cost, faster than the Apriori algorithm about one order of magnitude.

The algorithm presented in this paper will be a FP tree is divided into multiple small FP tree, so that the network of the machine running in parallel, in the process of decomposition of pruning and merge, simplified sub FP tree, thus improving the mining efficiency. Because each computer generated FP tree exists part of intersection, so if it were taken and set, can in a single FP tree, resulting in various computer on the mining cost sum may be greater than in a large type machine for single FP tree digging. But due to the parallel mining, parallel mining of real time cost depends on the Tmax, rather than the host for temporal summation.

Apparently the commodity as importance are the same, it will produce the support threshold value. If the threshold value is too high, the discovered association rules may be involved not to appear in low frequency but important project; if the threshold is too low, too many meaningless, or even misleading association rules, may also lead to combinatorial explosion, thereby reducing the efficiency of the algorithm. In order to solve practical database for each item in the importance of difference and uneven distribution of weight, we introduce concept, thus expanding the problem model, put forward the so-called weighted association rules. The paper presents the research of a novel weighted association rules algorithm based on unidirectional FP tree.

2. The Research of Weighted Association Rules Mining

In recent years, with the popularization and application of association rules in depth, association rule mining has become a practical significance of the mining technology of mining association rules is derived from known data all meet the minimum support and minimum of association rules. Usually this problem can be decomposed into two sub problems, namely: calculated database satisfying the minimum support degree all the frequent itemsets and use of frequent itemsets generated to meet the minimum confidence all association rules [2]. In fact, mining association rules in the process of the implementation of the first sub-problem is the core question. When finding all frequent itemsets, the corresponding the association rules can easily generate.

This paper proposes a new weighted association rules in weighted support measure and weighted reliability method, through the analysis of Boolean data examples, with Apriori algorithm and horizontal weighted association rules mining algorithm, prove that the method can keep the Apriori algorithm to frequent sets closed under properties, and can quickly and effectively mining important association rules.

Weighted association rules in database transaction set D m project, I={i1, i2, ... , im} items, each item has a weight corresponding to, their weights are {w1, w2, ... , wm}, wi ∈[0,1]. Specify the minimum weighted support threshold wminsup and minimum confidence threshold wmincof. Weighted association rules support degree and confidence degree of association rules in a X support for supp (X). MINWAL (O) algorithm definition weighted support suppw (A→ B) = (∑ WI) supp (A ∪B).

First put forward the suitable weighted association rules in MINWAL (O) algorithm. This algorithm is based on the Apriori algorithm, the principle and steps are as follows: the total number of transactions in the N database D in any q- set X, its support for the transaction database contains the X transaction number, denoted by SC (X). If X is weighted frequent itemsets, it must have the following formula1.

$$SC(X) \geq \frac{wminsup \times n}{\sum_{i_j \in X} w_j} \tag{1}$$

Weighted association rules mining method improvement, Boolean association rule mining. I= {I1, i2... i|I|} is a collection of items, D is a database transaction set. Each of the D service T is a collection of items, and satisfy the T I. The corresponding I weight vector for the W={w1, w2, ... , w|I|}, I transaction is a subset of I, such that every item in the Ti J project (named Ti [i J]) has a weight W. Head with an W value corresponding to, for example, (I, w) is I ∈I weighted terms, on the I transaction of the first j project weighted value denoted as Ti [i j [w]].

Therefore, a conclusion: non frequent superset is a frequent, this allows the algorithm to add new item prior to the cutting off of frequent items, so as to improve the efficiency of mining association rules. Through the attribute weights, a general algorithm of weighted association rules mining frequent items often do not meet the downward closure character. The improved method in mining Boolean and numeric data can maintain frequent downward closure character.

Algorithm to use Scan function scanning database, access to various project attribute value, the function Generate scan of the database, according to the definition of the weighted support formula for solving weighted support, and in accordance with user defined minimum support degree is 1- frequent set. The Apriori-gen function (k-1) - frequent sets are combined and the pruning, get k- a candidate. Because this algorithm maintains the "non frequent superset is also a frequent" this property, so the Apriori-gen function and Apriori algorithm. The Weight function is used to calculate the weights of candidate itemsets affairs, algorithm of weighted support is the option Ck traditional support and services right value product, the efficiency of the whole algorithm similar to that of Apriori, but added a number of database scanning process, in order to obtain the project attribute weights.

In order to solve this problem, can use fuzzy set smoothness to soften the boundary, and the attribute probability to calculate weights, so, almost all boundary nearby elements can neither be excluded, and also will not be unduly stressed; while at the same time, can do possible to excavate the small probability events in association rules [3]. Let I = {I1, I2, in }, is the project attribute set, while using I to construct a weighted data base T = {T1, T2,, TN}, weighted data base on fuzzy sets for the database properties, and is defined as a fuzzy attributes. Definition 1 item attribute the weight of ij w (ij) = P (ij). Among them, P (ij) ij in transaction database probability, as is shown by Figure 1.

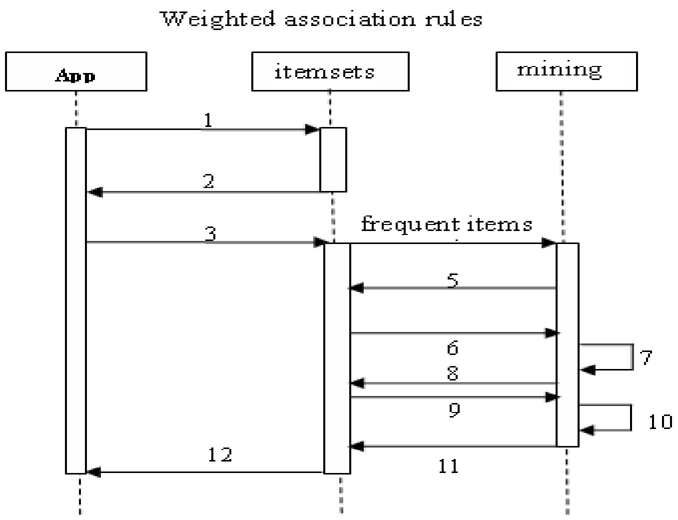

Figure 1. The Basic Model of Weighted Association Rules Mining

The association rule mining is only for the positive association rules, but the negative association rules of the effect are also very important. When making any decision often faced with many factors and it are to turn negative factors to a minimum. When the desired beneficial factors occur, while those adverse factors appear, to through the use of previous data master, which are the negative association rules? In view of the above problems, the paper will be weighted association rules with negative association rules of combination, put forward weighted positive and negative association rules mining.

Mining association rules from large amounts of data can be found in the itemsets asked potential and interesting interconnected. According to each item of interest to varying degrees, some scholars put forward horizontal weighted association rules. However, each generation of new candidate for the entire database. After scanning has become a bottleneck of the efficiency of the algorithm. In order to further improve the efficiency of mining weighted association rules, in the original horizontal weighted association rules algorithm based on depth-first strategy, taken out, which is based on the B C horizontal weighted association rules mining algorithm to U. The improved algorithm can greatly reduce the database scan range. The experimental results show that, the improved algorithm has better performance.

Weighted association rules can be divided into vertical and horizontal weighted association rules association rules. Vertical weighted association rules item weight changes over time, horizontal weighted association rules is based on the project of decision-making important degree distribution of different weight.

Association rule mining is a Web data mining field is an important aspect of data mining. Firstly, Web data mining and processing of Web data and other related knowledge are described; then studies the basic theory of association rules and the classical association rules algorithm: the last to solve reality database for each item in the distribution unevenness and the importance of differences, focus on the mining of weighted association rules algorithm. In depth analysis of the well-known weighted association rules mining algorithm to delete the ew.Apriori algorithm, the algorithm found the problems existing in the, as is shown by equation2 [4].

$$p(s(k) \mid a(k)) = \frac{1}{\sqrt{2\pi M^{-1}}} \exp\left[-\frac{(s(k) - \mu_{s|a})^2}{2M^{-1}} \right] \tag{2}$$

The weighted support will tend to support the right major attribute; weighted trust will tend to trust right now, then the weighted support rate and weighted trust analysis table 3 let W = {0. 5, 0.2, 0.4, 1}, the given threshold 0.1; called R on the normalized fuzzy set if and only if 0 x∈ R, A (x 0) = 1; then the 0 called the A x regular; 3) if theΠ K (0, 1], A K bounded set, then A is called bounded fuzzy set; - 4) R regular convex fuzzy set A known as a fuzzy number; regular closed convex fuzzy set is called a closed fuzzy number; regular bounded closed convex fuzzy set called 1, x = 0; bounded closed fuzzy number; 0 zero fuzzy number, 0 (x) = 0, X ≠ 05) let A be a fuzzy number, if suppA contains number is a positive real number, then A is called a normal fuzzy number, where suppA = {x∈ R A (x) > 0}.

Weighted association rules are to research on association rule mining expansion. It is aimed at users of different project importance of different database mining, to help users find interesting association rules, but also does not produce too many useless rules [5]. This paper mainly introduces a method of generating positive and negative association rules algorithm, in the traditional support confidence framework based on joining the correlation parameters to determine the set of correlation, thereby eliminating contradictory rules, as is shown by equation3.

$$\begin{aligned} D(x,y,\sigma) &= L(x,y,k_i\sigma) - L(x,y,\sigma) \\ &= G(x,y,k_i\sigma) * I(x,y) - G(x,y,\sigma) * I(x,y) \\ &= (G(x,y,k_i\sigma) - G(x,y,\sigma)) * I(x,y) \end{aligned} \tag{3}$$

The candidate 1 sets, 2 sets of candidate itemsets and candidate K. (k>2) generation are discussed respectively, first scans transaction database, find all items to the project appears set I, and the I project according to the weight sequence from high to low, then generates a 1 candidate set CI, in scanning each transaction, in addition to each item count, but also record that contains the transaction identifier (TID). It scanned over the database; get the candidate item set C. for each item set includes a corresponding transaction identifier list data structure of.Cl: (item items, weighted support wsup, transaction identifier list.

Input: The database D; minimum support threshold min_sup; X u Y is weighted frequent itemsets and wconf≥ wminconf (X=> VI;

Output: Weighted association rules;

Step 1: For a set of X weights, apparently 0 ≤ w (X)≤ L. project set I arbitrary subset of the X weighted support measure is defined as the wsup (X) =w (X);

Step 2: for each transaction in D do.

Step 3: Ct=subset(Ck,t);//get the subsets of t that are candidates, the rule is still useless, go to Step 7; Otherwise, go to Step 5.

Step 4: The user input the minimum weighted support expressed as wminsup, if wsup (X) >7 wminsup, then X is called weighted frequent itemsets. All weighted frequent items sets is denoted as L. when weighted frequent itemsets X contains k project, then known as the weighted frequent itemsets frequent K., K. item all weighted a collection of sets is denoted as Lk;

Step 5: If the project is set X is weighted frequent, V YsX, sup (Y) >7wminsup. X is weighted frequent itemsets, go to Step 4;

Step 6: if l1[1]=l2[1]∧l1[2]=l2[2]∧…∧l1[k-2]=l2[k-2]∧l1[k-1]<l2[k-1] then {;

Step 7: (Ck,Lk)=Checking(Ck,D)};

Step 8: Lk={C∈Ckl wsup(c)>一 wminsup};

Step 9: When k=l, Ci={{x}ixeI}: when k=2, C2={{x, y}l{x} ∈L1, Ys'∈CI, x<y); when k>2, let Ck={XoYIX∈ Lk.1, YELt.I, and X, Y can be connected;

Step 10: Rules(SC,L);

Thus can calculate each item set weighted support, and with the minimum weighted support, to produce weighted frequent 2 set L2. (3) In the candidate item sets Ck (k>2) before, on the Lk.i pruning, to reduce in the connection (k.1). Item sets, thus reducing the connection of the calculation, and can reduce the number of candidate Ck. (4) Lk.1 pruning is generated after the Lk-l' from Ck. to Ck connected, generated by pruning, in addition to impossible to generate weighted frequent itemsets candidate itemsets [6]. One item in the Ck set transaction identifier list equal to generate two of its set transaction identifier list item in the Lk. I am looking forward intersection.

For transaction databases D={TI, T2, ..., T.), N for the transaction record number |DI. to explore the weighted frequent itemsets, according to the following algorithm. WARDM: algorithm for mining weighted frequent itemsets. Input: transaction database D, minimum weighted support wminsup. Output: the set of all weighted frequent itemsets L. C1=Generate-Ci (D).

The weighted association rules describe: D, scanning, weighted candidate itemsets support too: WSup ({B,A, C and E}) = 0.3. Therefore, L3, = {{B, C and E}}. End of the loop, the algorithm ends. Generated from the L = L ∪ of Lk weighted association rules, Apriori algorithm..

New.Apriori algorithm require repeated scanning database to count, each generate a candidate itemsets when, to the database to conduct a comprehensive search. If the database store a large amount of data, so in a limited memory capacity, system I/O load is quite large, each scan the database could take a long time while WARDM algorithm only needs scan the database once, you can quickly find all weighted frequent itemsets, and stores the auxiliary information needed for space less. Only in the first step in the generation of candidate 1 set CI when the need to scan database, accessible to each item set transaction list, in computing any other candidate itemsets Ck support count, only the corresponding transaction list in Ck TID number can be equation4.

$$\mu_i(k) = P\{m_i(k) \mid Z(k)\} = \frac{f_i(k)\sum_{j=1}^{n}\pi_{ji}\mu_j(k-1)}{\sum_{i=1}^{n} f_i(k)\sum_{j=1}^{n}\pi_{ji}\mu_j(k-1)} \tag{4}$$

In this paper we respect science at the same time, the weighted association rules technology and negative association rules data mining technology development and the origin of international and domestic research status and outcomes were very fruitful research, but also for the weighting of the positive and negative association rules mining algorithm and construct the theoretic necessity premise. This paper is mainly on the weighted negative association rules shellfish interested, would be of interest measure into weighted mining positive and negative association rules, and put forward a kind of effective mining algorithm, and the user can according to their own requirements for setting interest to dig out more in line with the requirements of users of the rules.

The traditional association rule mining algorithm that database items in the important degree of the same, however in reality the importance of each item is different [7]. For example, decision makers tend to give priority to high profit item, and ignore the less profitable project. In addition, the passage of time and changes in consumption habits will also affect the association rules, time interval shorter transaction generated association rules although support is not too high, it can reflect the new consumption trend, therefore, in the actual data analysis, using the weighted association rules is meaningful. Proposed weighted Boolean association rules concept, and gives 2 kinds of weighted association rules mining algorithm: MINWAL (O) algorithm and MINWAL (W) algorithm, but the weighted support may be greater than 1, the weighted support not support containing multiple weighted attribute Association rules.

The method can highlight the effect of weight. Because of the above algorithm is introduced for each weight, thus damaging the Apriori algorithm in generating candidate item pruning. In this paper, considering the number of attributes and attributes weight on association rules influence, proposes a new weighted association rules support degree and confidence calculation method.

3. The Development of Unidirectional FP Tree

This paper analyses the FP- tree and one-way FP- tree two tree structure, summarize the problem existing in FP-growth algorithm, as well as Ming Fan puts forward one-way FP- tree and its algorithm. From the overall analysis, the proposed algorithm has better performance than the classical FP-growth algorithm. It is based on the research and conclusion, the author put forward the following maximal frequent itemset mining algorithm and the algorithm for mining frequent closed.

FP2 grow algorithm is more efficient algorithm of mining frequent patterns in one, but it will be used for mining maximal frequent patterns can be unable to obtain high efficiency 1 in-depth analysis of the causes of inefficient, proposes the use of FP2 sorted tree mining maximum frequent pattern algorithm SFP2Max1 algorithm as follows:①FP2 tree based on the sort; the use of the properties of maximal frequent patterns, reduce generate candidate max size;③setting intermediate result sets, reduce the test range, and thereby reduce the test candidate max time 1 experiments that show, SFP2 Max is an efficient algorithm for mining maximal frequent patterns, for the test data set, the performance of SFP2Max most cases are better than the MAFA algorithm, as is shown by equation 5.

$$\begin{cases} w_{j,\min}^{\xi}(m,n) = \frac{1}{2} - \frac{1}{2}\left[\frac{1 - M_{j,AB}^{\xi}(m,n)}{1-T}\right] \\ w_{j,\max}^{\xi}(m,n) = 1 - w_{j,\min}^{\xi}(m,n) \end{cases} \tag{5}$$

A frequent item set is {I1, I2, I3, ... , In} the possible frequent continued on surface analysis, which can be I1I5: 2, I1I3: 2, and I2: a set number of 2n 1. Proof: by the elements of I1, I2, I3, ... , In constitute a set of numbers are 2n, remove a null set, get 2n 1. Theorem 2 the method of separating the various FP tree by digging out to frequent item sets, not repeat [8].

Different FP2growt h algorithm is the frequent pattern directly into frequent itemsets; and FP2Max algorithm require frequent pattern as a candidate pattern 1 from the experimental results, the FP2Max low efficiency, even need to spend a lot of time;② to dense database, recursive layers too much, so much time and memory, but also produced a lot of redundancy mode;③for every candidate maximum mode to verify existing results to centralize all mode.

FP2 tree (FP2 tree or conditions) the item header table in A1, a2, ... , a n ' if we will FP2 tree all the same ', the father node of the child node according to their item header table sequence of row (column, then if A1, a2, ... , a I ' I≤ n) is in the FP2 tree ' a model, then it must be a FP2 tree on the left branch of the 1 1 proved by assuming that A1, a2, ... , AI ' FP2 ' is a type of tree in the mold, and FP2 trees have according to their item header table in the order sorting, so A1 must be the left most children, the A2 must be A1's left child, ... , AI a I - 1 of the children left, so the ' A1, a2, ... AI ', FP2 must be the left branch of the tree, as is shown by Figure 2.

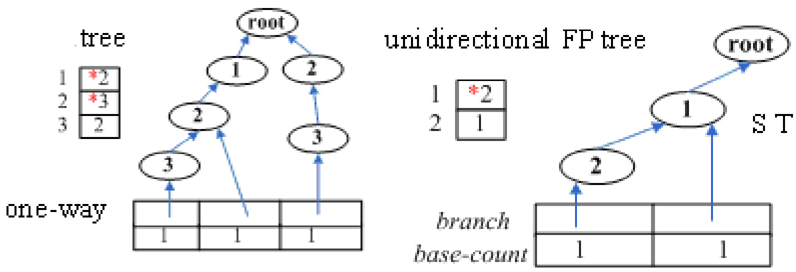

Figure 2. The Development of Unidirectional FP Tree Chart

In the one-way FP- mining frequent item sets of main process is simple. It uses a full array of FP storage has been a frequent itemsets (project its corresponding serial number states), length is the frequent item sets of length [9]. For each serial number I, the main process begins by item (I) will be converted into a number of I into FP, and output the frequent item set { 1- Ii } and its support of count[I / N (n is the number of transactions in the transaction database, similarly hereinafter). Then, constructed with one-way FP-tree bound subtree S T (I), and will be more frequent itemsets mining to process mine, as is shown by equatin6.

$$\mu_{ik} = \frac{\exp(-\frac{d_{ik}^2}{2\sigma^2})}{\sum_{j=1}^{c}\exp(-\frac{d_{jk}^2}{2\sigma^2})} \qquad i = 1, 2, \cdots c; k = 1, 2, \cdots n; \sum_{i=1}^{c}\mu_{ik} = 1 \tag{6}$$

In the initial FP2 tree, with a head table in a x suffix of the resulting models, not to x front for the suffix the patterns generated by the pattern of 1 proved 1 from the FP2 tree to generate frequent patterns known to item header table method, the X suffix produced by mode, the suffix is x 1 y item header table row in front of the X of any one item, due to X in the Y, so in a Y suffix to generate frequent patterns must not contain x, so these models can not be generated by the X model Super Collection 1 QED 1 theorem 21 let FP2 tree (FP2 tree or conditions) the item header table in Xiang Wei A1, a2, ... , a n ' if the FP 2 tree exists from the roots, to a node of the path formed by the model of A1, a2, ... , a I ' ' (I≤ n) and the mode of support degree greater than or equal to the minimum support threshold, while A1, a2, ... , AI - 1 in any of the patterns generated by a suffix is not the biggest mode.

The character of data sets and algorithms about 1 D strategy 1 kinds of characteristics of the data set is a set and frequent 12 item set is larger, the frequent itemsets is scattered, is sparse data sets; and D 2 data set and 12 set of frequent itemsets is smaller, but the data is very dense 1 for sparse data set MAF IA efficiency low reason, it to the transaction data is used in vertical bitmap representation, such as frequent 12 item set is larger, need to test the data amount is large, so the efficiency greatly, and the SFP2 tree representation of this do not exist issue 1 for dense data sets, due to the frequent 12 set of small size, MAF IA can give full play to the superiority of the bitmap representations and operations; and when the support threshold gain is very small, and the maximum frequent pattern scale is very large, due to the SFP2Max candidate set size [10].

FP 2Grow algorithm based on frequent pattern growth method for all frequent itemsets mining. It uses a divide and rule of the idea: the implicit frequent pattern database is compressed into a frequent pattern tree, but still retains a set of related information; and then generates compressed database into a group of conditions database, each of which is associated with a frequent item, and the structure of the conditional pattern tree mining.

4. A Novel Weighted Association Rules Algorithm Based On Unidirectional FP Tree

This paper proposes a new weighted association rule mining support degree and confidence degree method, the algorithm by weighting the practical importance in projects, and can maintain the Apriori to frequent the downward closure property. In practical use, there are relatively good results, but in order to meet the processing needs of large database must continue to optimize and improve the algorithm.

Based on FP-Growth algorithm multilayer associated rule mining algorithm, it is by introducing a repair, repair of the concept of cross layer to improve the performance of the original algorithm [11]. The algorithm only scans the database for 3 times, can produce various layers of frequent item sets and 1- repair items, each layer of the cross level frequent item sets and 1- cross level (lowest level) FP-tree tree. Then from the m layer of FP-tree repair, and set up the M tree layer, layer by layer traversal FP-tree tree, using path generating a layer of FP-tree tree, to achieve the same level association rules from low to high to establish in each layer of the FP-tree tree, mining and cross level association rules mining. So, the FAMML_FPT algorithm for any number of levels of the multilevel association rules mining, up to 3 times of scanning database. For Apriori based multilevel association rules mining algorithm, each layer in mining frequent itemsets and cross level of frequent item sets, needs to scan the database with multiple times, as is shown by equation7.

$$\frac{\lambda_1(k)}{\lambda_2(k)} = \frac{\pi_{21}}{\pi_{12}} \cdot \frac{\pi_{12}\mu_1(k) + \pi_{22}\mu_2(k)}{\pi_{11}\mu_1(k) + \pi_{21}\mu_2(k)} \tag{7}$$

Unidirectional FP- tree is one-way, from the roots to the leaves do not exist in the path, relative to the FP- tree with less pointers, each node needs at most two pointers, a pointer to the parent node, a pointer to the next node name [12]. Thus the occupied storage space is less than FP- tree.

This article proposes a new weighted association rule model, and based on the one-way FP tree is presented based on multiple minimum supports weighted association rules mining algorithm by it, for each data item set different weights and the minimum support degree, thus solving the transaction database data items vary in importance, and the actual frequency of the problem of uneven distribution of. Multiple minimum supports can be avoided by the single support leads to discovery showed a lower frequency of association rules or redundant rules, while the weight set so that the algorithm tends to mining and important data items related to the association rules [13]. In this paper gives the corresponding theorem, proved. In synthetic data on the experimental verification of the algorithm is effective.

Input: The database D; minimum weighted support wminsup; Unidirectional FP- tree Unid_FP-tree and minimum support count min_count;

Output: All weighted frequent items set L, All the maximum frequent set MFI;

Step1: Scans transaction database D, get each item support count;

Step2: If XcI, Y ∈I, and Xn Y=9. Xj Y weighted reliability defined as wconf (X=: =sup (XUY) /sup, VI (X);

Step3: Procedure in SERT tree (tran _ s, set< Item T > 3) {fo r tran s[I] array for each item in the {P = T find (tran s[I]);

Step4: for each transaction in D do (SC,C1)=Counting(D,W);

Step5: Because w (X) =w (Y) wsup (Y) t>wsup (X) >/wminsup;

Step6: build S T (i) based on Unid_FP-tree;

Step7: for each transaction t∈D{//scan D for counts ; Ct=subset(Ck,t);//get the subsets of t that are candidates;

Step8: build S T (k1 , ··· , km, i) based on S T (k1 , ··· , km);

Step9: if there is a non-root node in S T (i) and S T (i).count [ ]

Step10: Procedure has_frequent_subset(c:candidate k-itemset;Lk-1:frequent (k-1)-itemset);

Step11: (Ck,Lk)=Checking(Ck,D); L=L∪Lk;

Step12: if there is an non-root node in S T (X∪i) then call Unid_FP-FCI(S T (X∪i) , FCI);

Step13: Rules (SC,L);

In order to display the MWFI algorithm relative to MINWAL (O) algorithm advantage in performance, this section of MWFI algorithm and MINWAL (O) algorithm performance testing, and the two test results were compared and evaluated. Experiment hardware environment: CPU P4 2.8G, memory, hard disk 160G. experiment software environment: Windows XPSP2 operating system using the SQL Server system, the2005.NET integrated development environment, as a database programming environment respectively realizes the WARDM algorithm and New-Apriofi algorithm.

Comparative experiments on a Intel Pentium 4 3.80GHz CUP, 1GB memory, Windows XP operating system PC machine. A real data set using the UCI machine is learning database of mushroom data set (7672 records, 320 projects, and the average transaction length 23). Source code is using Visual C++ 6.0. Implementation results show that, compared with FP-Max algorithm, Unid_FP-Max algorithm mining speed is improved by over 1 time.

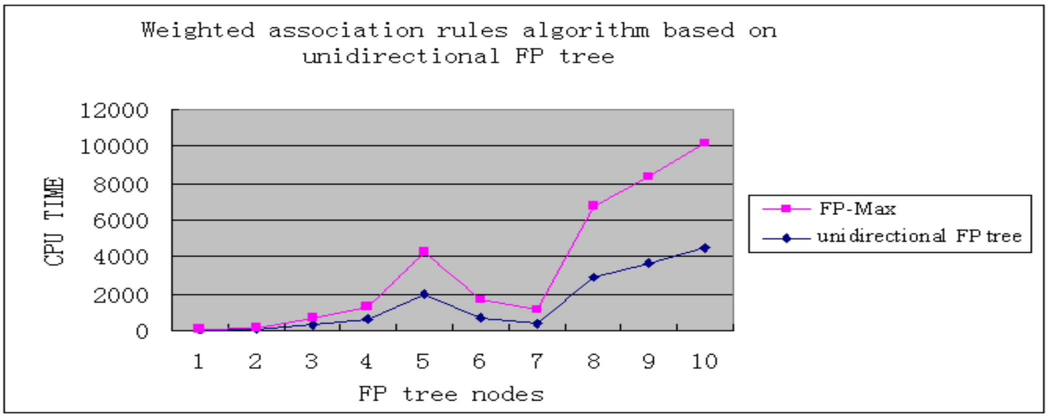

Figure 3. Comparison Results of Novel Weighted Association Rules Algorithm Based on Unidirectional FP Tree and FP-Max

The paper presents the research of a novel weighted association rules algorithm based on unidirectional FP tree. In order to display the WARDM algorithm compared with the New.Apriori algorithm advantage in performance, in this section will be on WARDM algorithm and New.Apriori algorithm performance testing, and the two test results were compared and evaluated, which reflected the two differences in performance on tests from the following three aspects of the WARDM algorithm and New.Apriori algorithm performance comparison: (1) in different minimum weighted support, the number of candidate itemsets. (2) In different minimum weighted support, generating a weighted frequent itemsets using time. On this basis, based on the design of the one-way FP- tree Unid_FP-Max algorithm, Unid_FP-Max algorithm theoretical analysis shows that the space overhead than FP-Max algorithm. The last of the FP-Max algorithm and Unid_FP-Max algorithm experiment results show that the algorithm comparison and analysis, the operating efficiency of Unid_FP-Max is relatively high.

5. Conclusions

The mining of weighted association rules in data mining technology is an important research field, and the mining of weighted association rules is the most important process is weighted frequent itemsets generation. This paper focuses on the analysis and research of weighted association rules of the basic theory and the classical algorithm. In order to solve the existing weighted association rule model of weighted support definition and weighted frequent itemsets mining algorithms, this paper proposes a new weighted association rule model, defines the weighted support, and gives the mining weighted frequent itemsets algorithm ---MWFI algorithm. This one-way FP- tree and its algorithm of frequent itemsets mining are based on one-way FP tree algorithm for mining maximal frequent itemsets Unid_FP-Max.

Acknowledgements

This paper is supported by Science andTechnology Department of Henan Province (No. 142400410263， No. 142400410267).

References

[1] H.-C. Chen, J.-L. Zhang and Y.-Q. Deng, "Application of Mixed-Weighted-Association-Rules-Based Data Mining Technology in College Examination grades Analysis", JDCTA, vol. 6, no. 10, **(2012)**, pp. 336 - 344.

[2] W. Yan, G. Li, Y. Zhao, J. A. Tian and W. Peng, "A Novel Weighted Frequent Pattern Mining Algorithm and Using in the Maintenance of Engine CBR System", IJACT, vol. 4, no. 20, **(2012)**, pp. 600 - 608.

[3] G. Rezvan and M.B. Behrouz, "Detecting Data Errors with Employing Negative Association Rules", JDCTA, vol. 3, no. 3, **(2009)**, pp. 91 - 95

[4] Y. Ge, X. Wang and Q. Li, "A improved Harmony Search Algorithm based on Association Rules", IJACT, Vol. 3, No. 9, **(2011)**, pp. 122 - 128

[5] Y. Hai and X.-.L Li, "A General Temporal Association Rule Frequent Itemsets Mining Algorithm", IJACT, vol. 3, no. 11, pp. 63 - 71, **(2011)**.

[6] M. Farhan, M. Mohsin, M. Helmy, A. Wahab, M. Fairuz Zaiyadi and C. F. Hibadullah, "An Investigation into Influence Factor of Student Programming Grade Using Association Rule Mining", AISS, vol. 2, no. 2, pp. 19 - 27, **(2010)**

[7] C. Zhang and S. Wang, "Application of Data Mining in Urban Traffic Accidents Governance Based on Association Rules", AISS, vol. 4, no. 19, **(2012)**, pp. 169 – 176.

[8] J. Cui, Q. Liu, K. Zhang, Z. Zhang and Y. Yang, "A Trust Cloud Model Based on the Association Rule Table of Multidimensional Trust-attribute", AISS, vol. 4, no. 16, **(2012)**, pp. 304 - 313

[9] S. Yu and X. Xu, "An Input and Output Monitoring System for FPGA-based Hardware PLC", IJEI, vol. 3, no. 1, **(2012)**, pp. 34 - 44

[10] Z. Liu, W.-X. Xiao, X.-L. Zhao and Q. Lu, "The Application of Collaborative Filtering Recommendation and FP-Tree Algorithms in Building Web Mining System", JCIT, vol. 7, no. 17, **(2012)**, pp. 215 – 222.

[11] X. Wu and X. Li, "Mining Maximal Frequent Subtrees based on Fusion Compression and FP-tree", JDCTA, vol. 5, no. 7, **(2011)**, pp. 191 – 200.

[12] D. Liu, "Research on the Multi-Relational Association Rule Mining Algorithm Based on User Guidance", IJACT, vol. 4, no. 22, **(2012)**, pp. 779 - 787

[13] Y. He and P. Du, "The Research of Landslide Monitring and Pre-warning Based on Association Rules Mining", JCIT, vol. 6, no. 9, **(2011)**, pp. 89 -95

Authors

Fengxia Yang, she was born on 1979, graduate educational background. Sheobtain a master's degree in Wuhan University, 2008 in Wu'han of China.

Her research area is Digital Resource Environment.She Engaged in computer teaching now in Henan College ofFinance & Taxationand has been interested in Intelligent Computingall along.

Kehao Cao, he was born on 1981, graduate educational background. He obtains a master's degree in Zhengzhou University, 2010 in Zhengzhou of China.

His research area is computer application.He Engaged in computer teaching now in Henan College of Finance & Taxationand has been i nterested in Intelligent Computing all along.

