Multiple Fuzzy Correlated Pattern Tree Mining with Minimum Item All-Confidence Thresholds

Radhakrishnan Anuradha, Nallaiah Rajkumar and Venkatachalam Sowmyaa

Abstract Rare item problem in association rule mining was solved by assigning multiple minimum supports for each item. In the same way rare item problem in correlated pattern mining with all-confidence as interesting measure was solved by assigning multiple minimum all-confidences for each items. In this paper multiple fuzzy correlated pattern tree (MFCP tree) for correlated pattern mining using quantitative transactions is proposed by assigning multiple item all-confidence (MIAC) value for each fuzzy items. As multiple fuzzy regions of a single item are considered, time taken for generating correlated patterns also increases. Difference in Scalar cardinality count for each fuzzy region is considered in calculating MIAC for fuzzy regions. The proposed approach first constructs a multiple frequent correlated pattern tree (MFCP) using MIAC values and generates correlated patterns using MFCP mining algorithm. Each node in MFCP tree serves as a linked list that stores fuzzy items membership value and the super—itemsets membership values of the same path. The outcome of experiments shows that the MFCP mining algorithm efficiently identifies rare patterns that are hidden in multiple fuzzy frequent pattern (MFFP) tree mining technique.

Keywords Fuzzy data mining · MIAC · MFCP tree · Fuzzy correlated patterns

R. Anuradha (✉) · N. Rajkumar · V. Sowmyaa
Department of Computer Science and Engineering, Sri Ramakrishna Engineering College, Coimbatore 641022, TN, India
e-mail: anuradha.r@srec.ac.in

N. Rajkumar
e-mail: nrk29@rediffmail.com

V. Sowmyaa
e-mail: sowmvenk@gmail.com

V. Ravi et al. (eds.), *Proceedings of the Fifth International Conference on Fuzzy and Neuro Computing (FANCCO - 2015)*, Advances in Intelligent Systems and Computing 415, DOI 10.1007/978-3-319-27212-2_2

1 Introduction

Data mining techniques have played an imperative role in deriving interesting patterns from a wide range of data [1]. Different mining approaches are divided based on the required necessity of knowledge like association rules [2, 3], classification rules [4] and clustering [5]. Most of the algorithms in association rule mining use level by level approach to generate and test candidate itemsets. One such algorithm is Apriori, which requires high computation cost for rescanning the whole database iteratively. To overcome the above drawback Han et al. [6] proposed the frequent pattern tree (FP Tree) approach which requires only two scans for processing the entire database and mining frequent patterns. The FP tree represents tree structure of a database which contains only the frequent items.

Fuzzy set theory [7] has been progressively used in intelligent systems for its easiness. Fuzzy learning algorithms for generating membership functions and inducing rules for a given dataset are proposed [8, 9, 10] and used in specific domains. In real world the transaction data mostly consists of quantifiable value. Hong et al. [6, 9] proposed a fuzzy mining algorithm for mining fuzzy association rules from quantitative data. Papadimitriou et al. [11] proposed an approach to mine fuzzy association rules based on FP trees. To mine frequent items from quantitative values, fuzzy mining has been proposed to derive the fuzzy rules using fuzzy FP-trees [12, 13, 14].

A correlation between two items becomes the key factor in detecting associations among items in market basket database. Correlated pattern mining was first introduced by Brin et al. [15] using contingency table that evaluates the relationship of two items. Many interesting measures like All-Confidence, Coherence and Cosine has been used in correlated pattern mining [16]. All confidence [17, 18] measure is used for mining interestingness of a pattern, which satisfies both anti-monotonic and null invariance properties. The above measure at higher threshold invokes rare item problem. Rare item problem in association rule mining was solved using multiple minimum supports by Liu et al. [19]. In the same way rare item problem in correlated pattern mining with all-confidence as interesting measure was solved using multiple minimum all-confidences by Rage et al. [20]. Fuzzy Correlated rule mining was first proposed by Lin et al. [21] using fuzzy correlation coefficient as the interesting measure.

Taking support and confidence as the only interesting measure would not produce rare correlated patterns. As fuzzy frequent and rare itemsets are mined by assigning multiple minimum supports to each fuzzy region, fuzzy correlated frequent and rare patterns are mined by assigning multiple item All-confidence values to each fuzzy regions. In this paper, a correlated pattern tree structure called multiple fuzzy correlated pattern tree (MFCP) is designed to mine frequent correlated patterns using Multiple Item all-confidence (MIAC). The proposed approach first constructs a MFCP tree using MIAC values and generates correlated patterns using MFCP mining algorithm. The amount of patterns generated is compared with the amount of patterns mined in multiple fuzzy frequent pattern rule mining algorithm.

1.1 Fuzzy FP-Tree Algorithm

Frequent pattern mining is one of the most significant research issues in data mining. Mining association rules was first given by Agarwal et al. [3] in the form of Apriori algorithm. The main drawback of the algorithm was repeated scanning of the database which proved to be too costly. To overcome the above problem Han et al. [22] introduced frequent pattern (FP) growth algorithm. FP growth algorithm uses a compressed data structure, called frequent pattern tree (FP-tree) which retains the quantitative information about frequent items. Frequent 1 itemsets will have nodes in the tree. The order of arrangement of tree nodes is that the more frequent nodes have better likelihood of sharing nodes when compared to less frequent nodes. FP-tree based pattern growth mining moves from frequent 1 itemset, scans the conditional pattern base, creates the conditional FP-tree and executes recursive mining till all its frequent itemset are scanned. Fuzzy FP Growth works in a same manner of FP-tree, with each node corresponding to a 1-itemset has a fuzzy membership function. The membership function for each 1-itemset is retrieved from the fuzzy dataset and the sum of all membership function values for the 1-itemset is its support. The support for a k-itemset (where k = 2) is calculated from the nodes corresponding to the itemset by using a suitable t-norm (Table 1).

Hong et al. [23] proposed mining fuzzy association rules using Aprioritid algorithm for quantitative transactions. Papadimitriou and Mavroudi [11] proposed an algorithm for mining fuzzy association rules based on FP trees. In their approach only local frequent 1-itemsets in each transaction were used for rule mining which was very primitive. Lin et al. [13] introduced a new fuzzy FP tree structure for mining quantitative data. In that approach the FP tree structure was huge as two transactions with identical fuzzy regions but different orders were put into two diverse paths. To overcome the above drawback Lin et al. [14] devised a compressed fuzzy frequent pattern tree. Here the items in the transactions, used for constructing the compressed fuzzy FP (CFFP) tree were sorted based on their occurrence frequencies. The above methodologies used only linguistic terms with maximum cardinality for mining process, making the number of fuzzy regions equal to the number of given items thus reducing the processing time.

1.2 Multiple Fuzzy FP-Tree Algorithm

Hong et al. [12] introduced a multiple fuzzy frequent pattern (MFFP) tree algorithm for deriving complete fuzzy frequent itemsets. Here a single itemset would have

Table 1 t-norms in fuzzy sets

$t-norm$
$T_M(x,y) = \min(x,y)$
$T_P(x,y) = xy$
$T_W(x,y) = \max(x+y-1,0)$

more than one fuzzy region which enables it to derive more fuzzy association rules than the previous models. A compressed multiple fuzzy frequent pattern tree (CMFFP) algorithm was proposed by Jerry Chun et al. [24]. This approach was an extension work of CFFP tree. CMFFP algorithm constructs tree structure similar to that of CFFP tree to store multiple fuzzy frequent itemsets for the subsequent mining process and each node contains additional array to keep track of the membership values of its prefix path. Hence the algorithm is efficiently designed for completely mine all the fuzzy frequent item sets.

Notation

n	Number of transactions in D
m	Number of items in D
D^i	The ith transaction datum, $i = 1$to n.
I_j	The jth item, $j = 1$ to m.
V_j^i	The quantity of an item I_j in. t ith transaction
R_{jk}	The K_th fuzzy region of item I_j.
${f_{jl}}^{(i)}$	${V_j^i}$'s fuzzy membership value in the region R_{jl}
Sum_{jl}	The count of fuzzy region R_{jl} in D
max_Sum_{jl}	The maximum count value of the fuzzy region R_{jl} in D

2 The Multiple Fuzzy Correlated Pattern Tree Construction Algorithm

Input: A body of n quantitative transaction data, a set of membership functions, a minimum all confidence threshold minAllconf and MIAC (Minimum item all confidence) values for all frequent fuzzy regions.

Output: A multiple fuzzy correlated pattern tree (MFCP tree).

Step 1 Transform the quantitative value V_j^i of each transaction datum $D^i, i = 1$to n for each item $I_j, j = 1$to m, into a fuzzy set $f_j^{(i)}$ represented as $\left(\frac{f_{j1}^{(i)}}{R_{j1}} + \frac{f_{j2}^{(i)}}{R_{j2}} + \cdots + \frac{f_{jp}^{(i)}}{R_{jp}}\right)$ using the membership functions, where R_{jk} is the Kth fuzzy region of item I_j, $V_j^{(i)}$ is the quantity of item I_j in ith transaction, $f_{jl}^{(i)}$ is $V_j^{(i)}$'s fuzzy membership value in region R_{jl} and p is the number of fuzzy regions for I_j.

Step 2 Calculate the sum of membership (scalar cardinality) values of each fuzzy region R_{jl} in the transaction data as

$$Sum_{jl} = \sum_{i=1}^{n} f_{ijl} \tag{1}$$

Step 3 The sum of fuzzy regions for each item is checked against the predefined minimum support count (α). If the sum is equal to (or) greater than minimum support count, the corresponding fuzzy region is put in the set of frequent fuzzy region (F_1).

$$F_1 = \{R_{jl}\, Sum_{jl} > \alpha, 1 \leq j \leq m\} \quad (2)$$

Step 4 Build the Header table by arranging the frequent R_{jl} in F_1 in descending order of their MIAC values.

Step 5 Remove the item of R_{jl}'s not present in F_1 from the transactions and place the remaining R_{jl}'s in new transaction database (D_1).

Step 6 Sort the remaining R_{jl}'s in each transaction of (D_1) in descending order of their membership values.

Step 7 Set the root node of the MFCP tree as {null} and add the transactions of D_1 into the MFCP tree tuple by tuple.

While inserting, two cases can exist.

Sub step 7-1 If a fuzzy region R_{jl} in a transaction is at the corresponding branch of the MFCP tree for the transaction, add the membership value f_{ijl} of R_{jl} in the transaction to the node of R_{jl} in the branch.

Sub step 7-2 Else insert a node of R_{jl} at the end of the identical branch, fix the sum of the node as the membership value f_{ijl} of R_{jl}, and inset a link from the node of R_{jl} in the last branch to the current node. If there is no such branch with the node of R_{jl}, insert a link from the entry of R_{jl} in the header table of the added node.

2.1 An Example

This section illustrates the proposed multiple fuzzy correlated pattern mining algorithm in quantitative database using an example. Assume six transactions with five items $(A-E)$ as shown in Table 2. The amount given in the database for each items are denoted using three fuzzy regions *Low*, *Middle* and *High*. Thus, each item in a transaction gets three fuzzy membership values based on the predefined membership functions. The minimum support is set as 25 %. The membership functions in Fig. 1 are used for all the items in the database.

Table 2 Database with six transactions

TID	Items
1	(A:5) (C:10) (D:2) (E:9)
2	(A:8) (B:2) (C:3)
3	(B:3) (C:9)
4	(A:7) (C:9) (D:3)
5	(A:5) (B:2) (C:5)
6	(A:3) (C:10) (D:2) (E:2)

Fig. 1 The membership functions used in the example

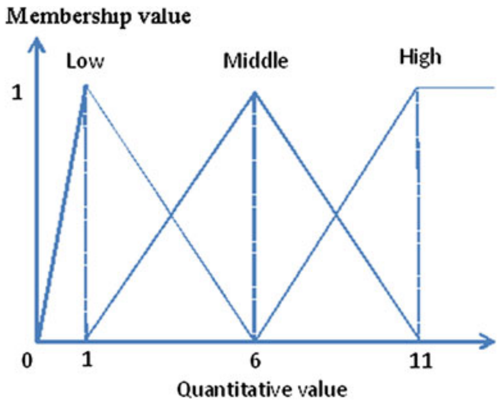

Table 3 shows the transformation of quantitative values of the item to fuzzy sets. For example the amount "5" of item A in Transaction one is transformed into $\left(\frac{0.2}{A.low} + \frac{0.8}{A.middle}\right)$. The sum of membership values of each fuzzy region is then calculated. For example the sum of the membership value of $A.middle$ is calculated as (0.8 + 0.6 + 0.8 + 0.8 + 0.4) which is 3.4. This step is repeated for all the regions and the results are displayed in Table 4.

Let minimum All-Confidence (minAllconf) threshold for the entire database is set to 70 % and the minimum support threshold is set as 24 %. The size of the dataset is 6 so the minimum support count is calculated as 6 * 24 % (=1.4). The sum of fuzzy regions that equals or larger than the minimum support count is kept in the set of frequent fuzzy regions (F_1) are used in successive mining. The frequent fuzzy regions (F_1) are displayed in Table 5.

Table 3 Fuzzy sets for Table 2 transactions

TID	Items
1	$\left(\frac{0.2}{A.low} + \frac{0.8}{A.middle}\right)\left(\frac{0.2}{C.middle} + \frac{0.8}{C.high}\right)$ $\left(\frac{0.8}{D.low} + \frac{0.2}{B.middle}\right)\left(\frac{0.4}{E.middle} + \frac{0.6}{E.High}\right)$
2	$\left(\frac{0.6}{A.middle} + \frac{0.4}{A.high}\right)\left(\frac{0.8}{B.low} + \frac{0.2}{B.middle}\right)\left(\frac{0.6}{C.low} + \frac{0.4}{C.middle}\right)$
3	$\left(\frac{0.6}{B.low} + \frac{0.4}{B.middle}\right)\left(\frac{0.4}{C.middle} + \frac{0.6}{C.high}\right)$
4	$\left(\frac{0.8}{A.middle} + \frac{0.2}{A.high}\right)\left(\frac{0.6}{D.low} + \frac{0.4}{D.middle}\right)$
5	$\left(\frac{0.2}{A.low} + \frac{0.8}{A.middle}\right)\left(\frac{0.8}{B.low} + \frac{0.2}{B.middle}\right)\left(\frac{0.2}{C.low} + \frac{0.8}{C.middle}\right)$
6	$\left(\frac{0.6}{A.low} + \frac{0.4}{A.middle}\right)\left(\frac{0.2}{C.middle} + \frac{0.8}{C.high}\right)$ $\left(\frac{0.8}{D.low} + \frac{0.2}{B.middle}\right)\left(\frac{0.8}{E.low} + \frac{0.2}{E.middle}\right)$

Table 4 Sum of membership values for the fuzzy regions given in Table 3

Item	Sum	Item	Sum
B.low	2.2	*A.low*	1.0
B.middle	0.8	*A.middle*	3.4
B.high	0.0	*A.high*	0.6
C.low	0.8	*D.low*	2.2
C.middle	2.4	*D.middle*	0.8
C.high	2.8	*D.high*	0.0
E.low	0.8	*E.high*	0.6
E.middle	0.6		

Table 5 The frequent fuzzy regions of set F_1

Frequent fuzzy regions	Sum
B.low	2.2
A.middle	3.4
C.middle	2.4
C.high	2.8
D.low	2.2

Mining frequent patterns and rare patterns requires Minimum item all confidence values (MIAC) values for all the fuzzy regions. To generate MIAC values for all fuzzy regions, it is necessary to calculate the difference in sum of the entire fuzzy region existing in the database.

Consider $\max\_Sum_{jl}$ as the maximum sum value in the database. (In this case it is 3.4). The difference in sum of entire fuzzy region (D^{sp}) is calculated using the formula.

$$\text{Sum Difference } (D^{sp}) = max\_Sum_{jl} - max\_Sum_{jl} * \text{ minAllconf} \tag{3}$$

$$= max\_Sum_{jl} * (1 - \text{ minAllconf}) \tag{4}$$

The MIAC value for each fuzzy region (R_{jl}) in the database is given by the formula.

$$\text{MIAC}(R_{jl}) = 1 - \left(\frac{D^{sp}}{Sum_{jl}}\right) \tag{5}$$

The MIAC value of any fuzzy region (R_{jl}) with maximum sum is the minAllconf value itself. For example the fuzzy region *A.middle* contains the maximum sum value of 3.4, the difference in sum value for the entire database is calculated as

$$\begin{aligned} D^{sp} &= \max\_Sum_{jl} * (1 - \text{ minAllconf}) \\ &= 3.4(1 - 0.70) \\ &= 1.02 \end{aligned}$$

Table 6 MIAC values for the frequent fuzzy items

Item	MIAC values (%)
B.low	54
A.middle	70
C.middle	57
C.high	63
D.low	54

Table 7 Header table

Fuzzy region	Sum
A.middle	3.4
C.high	2.8
C.middle	2.4
B.low	2.2
D.low	2.2

.

Substituting the above value in calculating MIAC (A.middle) we get

$$\mathrm{MIAC}\ (\ \mathrm{A.middle}) = 1 - \left(\frac{1.02}{3.4}\right) = 0.70$$

From this we come to a conclusion that the fuzzy regions of maximum sum would get minAllconf value as its MIAC value. In the same way MIAC values of other fuzzy regions are calculated and displayed in Table 6.

The MIAC values generated should be lesser or equal compared to the minAllconf of the entire database. So that rare fuzzy patterns are discovered.

The frequent fuzzy regions in F_1 are arranged in descending order of their MIAC value and are kept in the header table shown in Table 7.

The frequent fuzzy regions from each transaction are picked up from database (D) and placed into modified database (D_1). Sort the fuzzy regions in each transactions of (D_1) in descending order of their membership values. The results are displayed in Table 8. Build the MFCP tree by inserting fuzzy regions of each transactions tuple by tuple using the pattern tree. The resultant tree is given in Fig. 2.

Table 8 Ordered fuzzy regions for each transaction

TID	Items
1	$\left(\frac{0.8}{A.middle}\right)\left(\frac{0.8}{C.high}\right)\left(\frac{0.8}{D.low}\right)\left(\frac{0.2}{C.middle}\right)$
2	$\left(\frac{0.8}{B.low}\right)\left(\frac{0.6}{A.middle}\right)\left(\frac{0.4}{C.middle}\right)$
3	$\left(\frac{0.6}{B.low}\right)\left(\frac{0.6}{C.high}\right)\left(\frac{0.4}{C.middle}\right)$
4	$\left(\frac{0.8}{A.middle}\right)\left(\frac{0.6}{C.high}\right)\left(\frac{0.6}{D.low}\right)\left(\frac{0.4}{C.middle}\right)$
5	$\left(\frac{0.8}{A.middle}\right)\left(\frac{0.8}{B.low}\right)\left(\frac{0.8}{C.middle}\right)$
6	$\left(\frac{0.8}{C.high}\right))\left(\frac{0.8}{D.low}\right)\left(\frac{0.4}{A.middle}\right)\left(\frac{0.2}{C.middle}\right)$

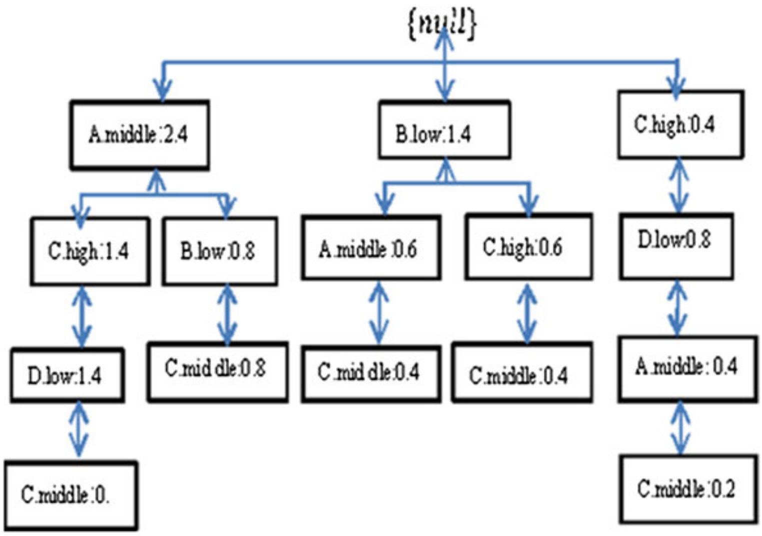

Fig. 2 MFCP tree after scanning all the transactions till tid = 6

3 Proposed Multiple Fuzzy Correlated Pattern Growth Mining Algorithm

The two measures for identifying frequent item set are support and confidence. Low level in these measures leads to the generation of too many patterns which are of least significance. To solve the above problem in quantitative databases All-Confidence measure is used which satisfies both anti-monotonic and null variance properties.

A fuzzy pattern using All-confidence measure for any two fuzzy regions F_x and F_y is defined as

$$\mathrm{fAll-conf}(\{F_x, F_y\}) = \frac{\sum_{i=1}^{n} \min(f_j(t_i) | f_j \in \{F_x, F_y\})}{\max(fsupp\{F_x, F_y\})} \tag{6}$$

where $f_j(t_i)$ is the degree that a fuzzy region appears in transaction t_i and *fsupp* is the fuzzy support value of two fuzzy regions F_x, F_y. In both the cases the minimum operator is used for intersection. Using fuzzy All-confidence measure and generating MIAC values for each fuzzy region the proposed algorithm generates rare correlated patterns.

Input: The MFCP tree, header table and predefined MIAC values for frequent items.

Output: The fuzzy correlated patterns.

Step 1 The fuzzy regions in the header table, arranged in descending order of MIAC values are processed one by one from lowermost to the uppermost. Let the current fuzzy region is taken as R_{jl}.

Step 2 The items in the path of the fuzzy region R_{jl} $(\alpha.term)$ are extorted to form the conditional fuzzy patterns. The minimum operator is used for intersection. The fuzzy values are obtained by grouping of fuzzy itemsets related to R_{jl} using minimum operation, which excludes the items associated with the same I_j. (For example when considering the fuzzy region *C.middle*, all other frequent fuzzy region are associated with it except *C.High* as both regions are from same fuzzy item).

Step 3 Add the fuzzy values of the resultant fuzzy itemsets that are same and put as conditional pattern base of R_{jl}. Let $\beta.term$ be any fuzzy item in conditional pattern base, R_{jl}.

Step 4 The fuzzy correlated patterns are generated by checking it against the following conditions.

$$fsupp(\alpha.term, \beta.term) \geq minsup \tag{7}$$

and

$$\frac{fsupp(\beta.term)withR_{jl}}{sum(\beta.term)} \geq MIAC(\beta.term) \tag{8}$$

As $\beta.term$ would be region with maximum value compared with $\alpha.term$ the Eq. (6) can be rewritten in the form of Eq. (8) and checked against MIAC value of the fuzzy region.

Step 5 Repeat the steps 2–4 for subsequent fuzzy region until all the fuzzy regions in the header table are processed.

3.1 An Example

For the above generated MFCP tree, the items are processed from bottom to top in the order given by the header table $(D.low, B.low, C.middle, C.high, A.middle)$. To generate any correlated pattern of (α, β) where α is fuzzy region of smallest MIAC value compared to β and β is $\alpha' s$ conditional pattern base.

From the given MFCP tree, the conditional pattern base of $(D.low)$ consists of $(B.low)(C.middle)\ (C.high)\ and\ (A.middle)$. If $\beta.term$ is taken as $(C.high)$.

$$fsupp(D.low, C.high) = 1.4 \geq minsup$$

and

Table 9 Fuzzy correlated patterns

S.No	Fuzzy correlated Itemsets
1	$(C.middle, A.middle)$
2	$(D.low, C.high)$
3	$(B.low, C.middle)$

$$\frac{fsupp(D.low, C.high)}{sum(C.high)} = 79\% \geq MIAC(C.high)$$

Hence, $(D.low, C.high)$ *is a correlated pattern*. The final lists of fuzzy correlated patterns are shown in the Table 9. For the above example database only 2-fuzzy correlated pattern are generated. There are no possibilities of mining 3-fuzzy correlated patterns from the given example.

4 Experimental Results

The testing environment consists of a 2.13 GHZ Intel® Core™ i3 processor with 2.00 GB of memory running a windows xp operating system. We took mushroom dataset and Retail dataset publicly available at the FIMI repository [25] is used for evaluation. The details of the datasets are shown in Table 10. Random values from 1 to 11 were assigned to the items in the dataset.

Minimum support value is set to 0.01. Experiments are conducted for a range of minimum All-Confidence values 100–30 %. MIAC values are calculated for the frequent-1 items (items in Header Table) with reference to the highest support item in the database. Figures 3 and 4 show the execution times taken for various minimum All-Confidence values for mushroom and retail datasets respectively.

Table 10 Details of databases

Database	No. of transactions	No. of items	Avg. length of transactions
Mushroom	8124	119	23
Retail	88,162	16,470	76

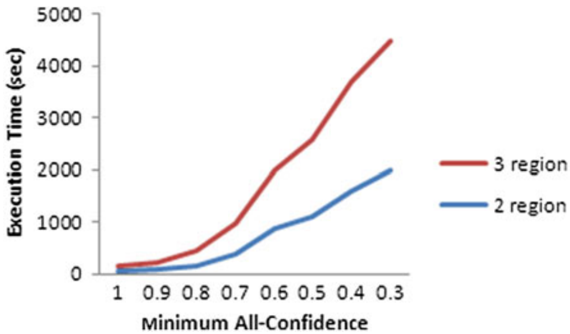

Fig. 3 Execution time at different MIAC values for mushroom dataset

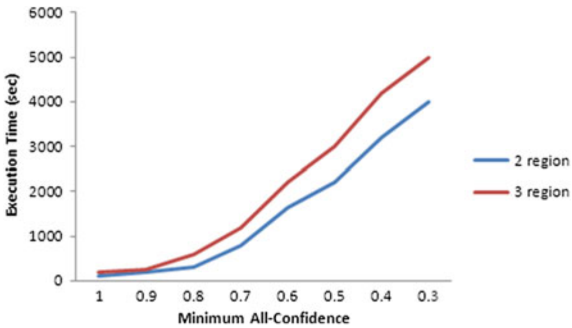

Fig. 4 Execution time at different MIAC values for retail dataset

Figures 3 and 4 show that the execution time is prolonged for three regions when compared to two regions for all minimum All-Confidence thresholds. This is due to the fact that three regions would produce more tree nodes than two fuzzy regions. The processing time also depends upon transformation of fuzzy regions using predefined membership functions and the quantitative values of items.

Figure 5 illustrates the number of correlated patterns generated for various minimum All-Confidence values for mushroom and retail dataset. The number of patterns obtained in the retail dataset is more when compared to the patterns generated from the mushroom dataset. This is because; increase in dataset size causes more patterns to be generated. The number of correlated patterns generated using the MFCP approach is more when compared with multiple fuzzy frequent pattern (MFFP) growth mining algorithm. The MFCP mining approach uses different MIAC values for each item based on its support difference, so at even high confidence levels rare items are generated. As support and confidence are the only interesting measures considered for MFFP growth mining algorithm, the time taken is slightly less when compared to the MFCP tree mining approach. Figure 7 displays the time taken for execution using the above methods using retail dataset. Fuzzy rare patterns are left out in MFFP growth mining algorithm as it works on single support and confidence measure for the entire database. The MFCP approach generates rare patterns along with frequent patterns. So Comparatively MFCP generates more number of patterns than MFFP approach. Using Retail dataset for

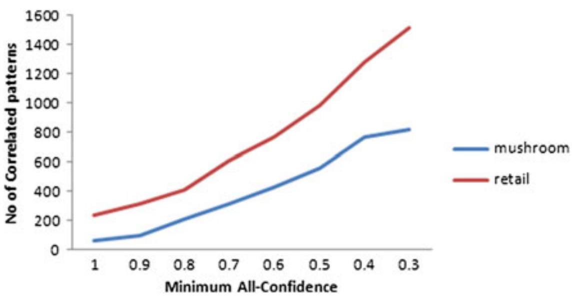

Fig. 5 Number of correlated patterns generated at different min-all-confidence

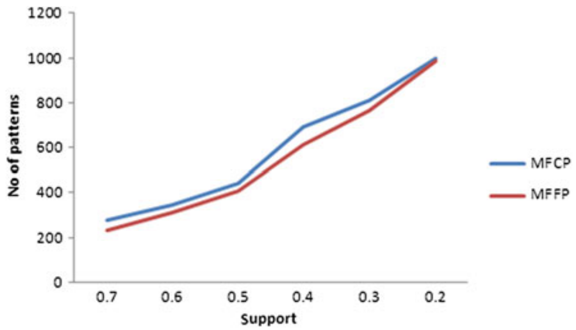

Fig. 6 Comparison on the number of patterns generated using MFFP and MFCP

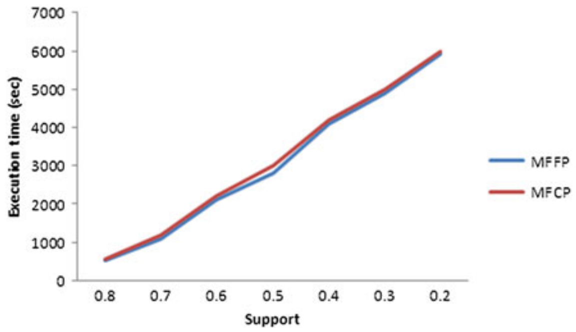

Fig. 7 Execution time of MFFP and MFCP

the given support measure, patterns are generated using the above mentioned approaches. The Fig. 6 displays the number of patterns generated using both MFCP and MFFP growth mining algorithms.

5 Conclusions and Future Work

In Traditional data mining approaches, maximum work is on binary databases to find significant information. The discovered knowledge is represented in a statistical or numerical way. In real world applications, the scope of using binary databases is less. Fuzzy set theory reveals human thinking and aids to take significant decisions. In this paper, the idea of Multiple Fuzzy Frequent Pattern tree [12] and Correlated pattern tree [20] are integrated to find fuzzy correlated patterns in a given quantitative data set. In the MFCP tree mining algorithm MIAC values are set for each item to find the fuzzy correlated patterns. The rational for using two confidence thresholds has been justified. An example is given to demonstrate that the MFCP tree mining algorithm can derive multiple frequent correlated patterns under multiple minimum item confidence in a simple and effective way.

In the future, the proposed approach is extended for dynamic databases and also investigates the extension of proposed work for generating fuzzy closed and maximal correlated patterns.

Acknowledgments The authors like to thank the anonymous referees for their notable comments and Sri Ramakrishna Engineering College for providing resources for our experiments.

References

1. Tan, P.N., Steinbach, M., Kumar, V.: Introduction to Data Mining. Pearson Addison Wesley, Boston (2006)
2. Agarwal, R., Imielinski, T., Swami, A.: Mining association rules between sets of items in large databases. The International Conference on Management of Data, pp. 207–216 (1993)
3. Agarwal, R., Srikant, R.: Fast algorithms for mining association rules in large databases. The International Conference on Very Large Databases, pp. 487–499 (1994)
4. Kim, E., Kim, W., Lee, Y.: Combination of multiple classifiers for the customer's purchase behavior prediction. Decis. Support Syst. (2003)
5. Berkhin, P.: A Survey of clustering data mining techniques. Technical Report, Accrue Software, San Jose, CA (2002)
6. Han, J., Pei, J., Yin, Y., Mao, R.: Mining frequent patterns without candidate generation: a frequent-pattern tree approach. Data Mining Knowl. Discov. (2004)
7. Zadeh, L.A.: Fuzzy set. Inf. Control **8**(3), 338–353 (1965)
8. de Campos, L.M., Moral, S.: Learning rules for a fuzzy inference model. Fuzzy Sets Syst. **59**, 247–257 (1993)
9. Hong, T.P., Chen, J.B.: Finding relevant attributes and membership functions. Fuzzy Sets Syst. **103**, 389–404 (1999)
10. Hong T.P., Lee, C.Y.: Induction of fuzzy rules and membership functions from training examples. Fuzzy Sets Syst. 33–47 (1996)
11. Papadimitriou, S., Mavroudi, S.: The fuzzy frequent pattern tree. In: The WSEAS International Conference on Computers (2005)
12. Hong, T.P., Lin C.W., Lin, T.C.: The MFFP-Tree fuzzy mining algorithm to discover complete linguistic frequent itemsets. Comput. Intell. **30**(1) (2014)
13. Lin, C.W., Hong, T.P., Lu, W.H.: Mining fuzzy association rules based on fuzzy fp-trees. In: The 16th National Conference on Fuzzy Theory and Its Applications, pp. 11–16 (2008)
14. Lin, C.W., Hong, T.P., Lu, W.H.: An efficient tree-based fuzzy data mining approach. Int. J. Fuzzy Syst. **12**(2) (2010)
15. Brin, S., Motwani, R., Silverstein, C.: Beyond market baskets: generalizing association rules to correlations. SIGMOD Rec. **26**, 265–276 (1997)
16. Kim, S., Barsky, M., Han, J.: Efficient mining of top correlated patterns based on null invariant measures. In: ECML PKDD. pp. 172–192 (2011)
17. Kim, W.Y., Lee, Y.K. Han, J.: Ccmine: efficient mining of confidence-closed correlated patterns. In: PAKDD, pp. 569–579 (2004)
18. Lee, Y.K., Kim, W.Y., Cao, D., Han, J.: CoMine: efficient mining of correlated patterns. In: ICDM, pp. 581–584 (2003)
19. Liu, B., Hsu, W., Ma, Y.: Mining association rules with multiple minimum supports. In: KDD, pp. 337–341 (1999)
20. Rage, U.K., Kitsuregawa, M.: Efficient discovery of correlated patterns using multiple minimum all-confidence thresholds. J. Intell. Inf. Syst. (2014)
21. Nancy P. Lin., Hao-En Chueh.: Fuzzy correlation rules mining. In: International Conference on Applied Computer Science, Hangzhou, China, 15–17 April (2007)

22. Han, J., Kamber, W.: Data Mining: Concepts and Techniques. Morgan Kaufman, San Francisco (2001)
23. Hong, T.P., Kuo, C.S., Wang, S.L.: A fuzzy aprioritid mining algorithm with reduced computational time. Appl. Soft Comput. **5**, 1–10 (2004)
24. Lin, J.C., Hong, T.P., Lin, T.C.: CMFFP-tree algorithm to mine complete multiple fuzzy frequent itemsets. Appl. Soft Comput. (2015)
25. Frequent Itemset Mining Implementations repository: http://fimi.helsinki.fi

Springer

http://www.springer.com/978-3-319-27211-5

Proceedings of the Fifth International Conference on
Fuzzy and Neuro Computing (FANCCO - 2015)
Ravi, V.; Panigrahi, B.K.; Das, S.; Suganthan, P.N. (Eds.)
2015, XIII, 366 p. 145 illus., 102 illus. in color., Softcover
ISBN: 978-3-319-27211-5