ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

A Framework for Discovering Temporal Coherent Rules

R. Ratchambigah[1], R.Swetha[2], R. Anuradha[3],

[1,2]B.E Student, Department of Computer Science and Engineering, Sri Ramakrishna Engineering College, Coimbatore, Tamilnadu, India

[3]Assistant Professor,Department of Computer Science and Engineering, Sri Ramakrishna Engineering College Coimbatore, Tamilnadu, , India

ABSTRACT: Data mining is the procedure of extracting patterns and knowledge from large dataset. Using different membership functions for different levels, the quantitative data is converted into fuzzy values. In quantitative dataset interesting rules are mined using minimum support and minimum confidence. These threshold values are difficult to set for each dataset. To overcome the above problem, the fuzzy temporal coherent rule mining algorithm has been introduced for a predefined taxonomy. The association in the same level items isfound using coherent logic. Fuzzy value is generated for each level using different low, middle and high membership values (Fig. 1). Positive and negative coherencies are found using coherency formula. Coherency is found between level-wise items which include items at the same level and cross level. In this methodology, both positive and negative coherent rules are mined for same level in a taxonomical dataset. From the resultant data, the maximum number of coherent rules for the items among the levels in different time periods isalso found. The rules mined gives knowledge on association of items without the usage of interesting measure.

KEYWORDS: temporal coherent rules; membership function; quantitative dataset;

I. INTRODUCTION

Data mining is used to discover valuable information which eventually helps in decision making.Theactualdatamining taskisfullyautomaticorsemi-automaticanalysisoflarge volumesofdatato extricatepreviouslyunknowninteresting patternssuchasgroupsofdata records(clusteranalysis),uncommonrecords(anomalydetection)anddependencies(association rule mining). In data mining, association rules are useful for analyzing and predicting customer behavior, credit card fraud detection, decision making and genomic analysis. It also plays an important part in market basket data analysis, product clustering, and catalog design and store layout.

Association rule mining is a process of efficiently mining both positive and negative rules. Most of association rule mining techniques used binary data for extracting information. But the real world scenario includes quantitative data. Fuzzy learning methodologies are used in deriving membership functions and rule generation. Fuzzy association rule mining techniques are discovered to mine significant rules from quantitative data.

In actual scenario, the knowledge that is utilized for helping decision-making is always time fluctuating. For supporting better decision making, it is favorable to have the capacity to really identify the temporal features with the interesting patterns or rules. It focuses on finding all interesting contiguous time intervals during which a specific association holds.

Some association rules might be infrequent at times and therefore do not cross the minimum support threshold required for them to appear in the frequent items sets and from now on in the association rules found by the mining framework. Hence, taking occasionally into consideration is important for deriving more robust association rules. All the more exactly, association rules show temporal natures and along with these esteemed in-complete and incorrect if not associated by

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

appropriate temporal patterns demonstrating the periods in which they are legitimate. In this way, we are not just interested in mining association rules which are actually valid but also those rules that we have to apply for only limited time intervals. Additionally, we are keen on deciding all the frequent item sets in the transaction data alongside the maximal time interims for which they are applicable.The idea used in this paper is to extend the concept of fuzzy coherent mining for taxonomical data associated with temporal data mining.

II. REVIEW OF RELATED WORKS

This section analyses related works in the field of temporal rules, fuzzy association rule mining and coherent rule mining.

The concept of association rule was first introduced and discovery of association rules has been extensively studied [1] [2]. Association rule mining is a well-established method of data mining which identifies significant correlations between the data items in large databases [3]. For finding out the frequency of the desired association rules using frequent pattern growth algorithm has been discussed [4].Transaction data in real-world applications usually consist of fuzzy and quantitative values. The designof sophisticated data-mining algorithms able to deal with various types of data.The concepts and techniques of some fuzzy mining are along with association-rule discovery [5].

The frequent fuzzy pattern tree (fuzzy FP-tree) is found for getting frequent fuzzy itemsets from the transactions with quantitative values. There are few issues in mining association rules and related time intervals, where an association rule holds either in all or portion of the interims. To confine to valid time intervals, some calendar based examples is been utilized with the assistance, the processing turn out to be a great deal than the original since fuzzy intersection in each transaction should be taken care of. The fuzzy FP-tree construction algorithm is hence composed, and the mining process based on the tree is displayed [6].

In some mining techniques, the complexity and efficiency has not been considered, which will lead to a great drawback in algorithms. The effectiveness and efficiency of Fuzzy Association Rules Mining from Ubiquitous Data Streams is been evolved by developing a new technique named Fuzzy Frequent Pattern Ubiquitous Stream (FFP_USTREAM) [7]. It includes the discovery of Temporal Association of finding efficient algorithms for this mining problem by extending the well-known Apriori algorithm with effective pruning techniques [8].

In emerging trend, transactions may include quantitative values and each item may have a life expectancy from a temporal database. The data mining algorithm converts each quantitative value into a fuzzy set utilizing the given membership functions. The algorithm then figures the scalar cardinality of each linguistic term of everything. A mining procedure based on fuzzy numbers and item lifespans is then performed to discover fuzzy temporal association rules [9].

The calendar-based pattern [10] has already been proposed by researchers to restrict the time-based association-ships. It proposes a novel algorithm to discover association rule on time dependent data utilizing efficient T tree and P-tree data structures. The algorithm explains the noteworthy advantage in terms of time and memory while joining time dimension. This approach of filtering in light of time-intervals yields small dataset for a given valid interval along these lines reducing the running time.

There are lot of fuzzy data mining approaches have been proposed for discovering fuzzy association rules with the predefined minimum support from the given quantitative values. However, those approaches are that an appropriate minimum support is hard to set. Sometimes, the interesting patterns are missing or infrequent patterns are also including in transactions. By overcoming those problems, coherent rule mining algorithm is been proposed by generating fuzzy sets [11].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering
(An ISO 3297: 2007 Certified Organization)
Website: www.ijircce.com
Vol. 5, Issue 3, March 2017

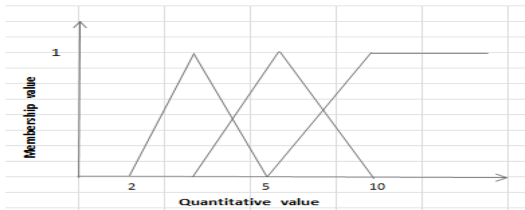

Fig. 1. The membership functions

III. FUZZY TEMPORAL COHERENT RULE MINING ALGORITHM (FTCRM)

The proposed algorithm first mines level-wise fuzzy coherent rules and cross-level fuzzy coherent rules for a predefined taxonomy. Later it retrieves the irregular coherent rules that exist in the database.

INPUT: A body of n temporal quantitative transaction data, a given item y, and a given set of membership functions.

OUTPUT: A set of fuzzy temporal coherent rules (FTCR).

STEP 1: Transform the temporal quantitative value $v^{(i)}_j$ of each transaction datum $D^{(i)}$, i = 1 to n, for each temporal item I_j, j = 1 to m, into fuzzy values represented as $\left(\frac{f\ (i)j1}{Rj1}\right)+\left(\frac{f(i)j2}{Rj2}\right)+\ldots+\left(\frac{f\ (i)jl}{Rjl}\right)$using the given membership functions, where the fuzzy region R_{jk} is defined as item. It means the k^{th} fuzzy region of item I_j. The fuzzy value $f^{(i)}_{jk}$ is $v^{(i)}_j$'s fuzzy membership value in region R_{jk}, k = 1 to l, and l is the number of linguistic terms for I_j. A fuzzy region is defined as item term.

STEP 2: Record the number of temporal transactions in time period p_t, denoted as $|p_t|$, and the starting period of each temporal item, denoted as $H(Ij)$, into a Temporal Information table, denoted as TI table. Each item in a transaction thus has its own life time.

STEP 3:Calculate the scalar cardinality of each fuzzy region (linguistic term) R_{jk} in the temporal transaction data:
$count_{jk} = \sum_{i=1}^{n} f_{jk}^{(i)}$.

STEP 4: For each fuzzy region R_{jk}, calculate its complement value according to the definition of fuzzy complement which is: If $f^{(i)}_{jk}$ is the fuzzy value of kth fuzzy region for item I_j, then its fuzzy complement is calculated as $(1 - f\ (i)jk\ )$. The result is represented as follows$\left(\frac{1-f(i)j1}{Rj1}+\frac{1-f(i)j2}{Rj2}+\frac{1-f(i)jl}{Rjl}\right)$,where i is the transaction id number, the $f^{(i)}_{jk}$ is $v^{(i)}_j$' s membershipvalue in fuzzy region R_{jk}.

STEP 5: Collect all fuzzy regions into the set B. In other words, the $set\ B = \{R11, R12, \ldots, R1l, R21, \ldots, Rjk, \ldots, Rml\}$, where j ranges from 1 to m and k ranges from 1 to l, m is the number of items and l is the number of linguistic terms in the given membership functions.

STEP 6: Collect all fuzzy regions of the given item y into the set P. In other words, the $set\ P = \{Ry1, Ry2, \ldots, Ryk, \ldots, Ryl\}$.

STEP 7: Remove the given item y's fuzzy regions from the set B to from the set K. In other words, the $set\ K = B - P$. The two sets P and K are then used to generate the candidate coherent fuzzy rules in the following steps.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

STEP 8: Set g = 1, where g means the length of antecedent X and form candidate fuzzy coherent rule $X \rightarrow Y$, where X is an element of K, and Y is an element of P.

STEP 9: Execute the following sub-steps to generate fuzzy temporal coherent rules:

STEP 9.1: Calculate the contingency table for antecedent X and consequent Y belonging to a same period. Here, four count values will be found out by calculating it, including Q1:$count_{XY}$, Q2:$count_{X\sim Y}$, Q3:$count_{\sim XY}$ and Q4:$count_{\sim X\sim Y}$. From the above calculations,$count_H = \sum_{i=1}^{n} f_H^{(i)}$,where the fuzzy value of an item set H in each transaction is calculated as $f^{(i)}_H = f^{(i)}_{H1} \wedge f^{(i)}_{H2} \wedge ... \wedge f^{(i)}_{Hm}$, and $f^{(i)}_{Hj}$is the membership value of fuzzy item Hj in ith transaction. If the minimum operator is used for the intersection, then:$f_H^{(i)} = Min_{j=1}^{m} f_{Hj}^{(i)}$.

STEP 9.2: Check the candidate fuzzy coherent rule meets the four conditions, which are $Q1 > Q2, Q1 > Q3, Q4 > Q2$ and $Q4 > Q3$ or not. If yes, let $FTCRall = FTCRs \cup (X, Y)$, go to STEP 9.1 to calculate the contingency table of next candidate fuzzy coherent rule. If all candidate rules are checked, go to STEP 9.3. Otherwise, back to STEP 9.1.

STEP 9.3: Check the $FTCR_g$ is empty or not. If yes, go to STEP13.Otherwise, go to STEP 10.

STEP 10: Collect the fuzzy regions of antecedent and consequence parts of the derived fuzzy temporal coherent rules in $FTCR_g$ to form new set K and P, respectively.

STEP 11: Set $g = g + 1$.Form candidate fuzzy coherent rule $X \rightarrow Y$ according to P and K, where the length of X is g. Note that fuzzy regions with the same item cannot be used to form candidate fuzzy coherent rules.

STEP 12: If there is null candidate fuzzy coherent rule, go to STEP 13. Otherwise, go to STEP 9.1.

STEP 13: Output the Fuzzy Temporal Coherent rules $FTCR_{all}$.

Table 1: The temporal quantitative transactions. (Mentioned in Step1, IV)

Period	TID	Items	Period	TID	Items
P1	T1	(I_1,2),(I_2,13), (I_3, 7)	P2	T4	(I_1,10),(I_2,6), (I_4,16)
	T2	(I_1,8),(I_3,5), (I_4,16)			
	T3	(I_3,6),(I_4,7), (I_5,7)		T5	(I_1,10),(I_2,10), (I_5, 8)

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

Table 2: The fuzzy set transformation. (Mentioned in Step1, IV)

Period	TID	Items
P1	T1	$(0.8/I1.Low + 0.2/I1.Middle), (1/I2.High), (0.75/I3.Middle + 0.25/I3.High);$
	T2	$(0.5/I1.Middle + 0.5/I1.High), (0.2/I3.Low + 0.8/I3.Middle), (1/I4.High);$
	T3	$(1/I3.Low), (0.75/I4.Middle + 0.25/I4.High), (0.75/I5.Middle + 0.25/I5.High)$
P2	T4	$(1/I1.High), (1/I2.Middle), (1/I4.High);$
	T5	$(1/I1.High), (1/I2.High), (0.5/I5.Middle + 0.5/I5.High)$

Table 3: The count values of fuzzy regions.(Mentioned in Step 3, IV)

Items	Count	Items	Count
I_1.Low	0.8	I_2.Low	0
I_1.middle	0.7	I_2.Middle	1
I_1.High	2.5	I_2.High	2

Table 4: The complement fuzzy set transformed from Table 3.(Mentioned in Step 4, IV)

Period	TID	Items
P1	T1	$(0.2/I1.Low + 0.8/I1.Middle + 1/I1.High), (1/I2.Low + 1/I2.Middle), (1/I3.Low + 0.25/I3.Middle + 0.75/I3.High);$
	T2	$(1/I1.Low + 0.5/I1.Middle + 0.5/I1.High), (0.8/I3.Low + 0.2/I3.Middle + 1/I3.High), (1/I4.Low + 1/I4.Middle);$
	T3	$(1/I3.Middle + 1/I3.High), (1/I4.Low + 0.25/I4.Middle + 0.75/I4.High), (1/I5.Low + 0.25/I5.Middle + 0.75/I5.High)$
P2	T4	$(1/I1.Low + 1/I1.Middle), (1/I2.Low + 1/I2.High), (1/I4.Low + 1/I4.Middle);$
	T5	$(1/I1.Low + 1/I1.Middle), (1/I2.Low + 1/I2.Middle), (1/I5.Low + 0.5/I5.Middle + 0.5/I5.High)$

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

Table 5: The membership values for $I1.High \cap I4.High$. (Mentioned in Step 9.1, IV)

TID	I1.High	I4.High	$I1.High \cap I4.High$
T1	0	0	0
T2	0.5	1	0.5
T3	0	0.25	0
T4	1	1	1
T5	1	0	0

IV. AN EXAMPLE

In this section, an example is given to get an idea of how the proposed fuzzy temporal coherent rule mining algorithm works. This is a simple example to demonstrate how the proposed algorithm can be used to generate coherent rules from temporal quantitative transactions. The dataset includes five transactions (Table 1) and it will be used to illustrate the proposed approach. In this example, we assume that each itemhas its own membership functions.

Step 1: The temporal quantitative transactions in two time periods, P1 and P2 are collected (Table 1). Each quantitative transaction in the database is transformed into corresponding fuzzy sets. Temporal item 1 of T1 in time period P1 will be used as an example. The quantitative value "2" is converted into a fuzzy set $(0.8/1.Low + 0.2/1.Middle)$ using the given membership functions. This step is followed for all transactions and time periods. The results are shown in Table 2.

Step 2: The number of temporal transactions in each time period and the starting period of each temporal item are recorded in a Temporal Information (TI) table. Time period P1 will be taken as an example. The number of transactions is five, and the starting period of temporal items I_1, I_2, I_3, I_4 and I_5 is time period P1.

Step 3: The scalar cardinality of each fuzzy region in the temporal transaction data is calculated. Here by taking I_1.Middle, the scalar cardinality can be calculated as $0.2 + 0.5 + 0.0 + 0.0 + 0.0\ is$ 0.7. This step will be same for all fuzzy regions. The results are shown in Table 3.

Step 4: The complement fuzzy set is derived from the transformation of fuzzy values. Take an example of the second item in transaction T3 using the membership functions. The fuzzy region $(0.75/I4.Middle + 0.25/I4.High)$ item I_4 is then converted into the complement fuzzy set $(1/\ I4.Low + 0.25/\ I4.Middle + 0.75/\ I4.High)$. The same steps will be followed for each item. The results are shown in Table 4.

Step 5: All fuzzy regions are then collected into a set B, which is B = {I_1.Low, I_1.Middle, I_1.High,I_2.Low, I_2.Middle, I_2.High,I_3.Low, I_3.Middle, I_3.High,I_4.Low, I_4.Middle, I_4.High,I_5.Low, I_5.Middle, I_5.High}.

Step 6: The fuzzy regions of item I_1 are collected to form the setP = {I_1.Low, I_1.Middle, I_1.High}.

Step 7: The set K is generated from the set B by removing fuzzy regions of item I_1, which is K = $\{I2.Low, I2.Middle, I2.High, I3.Low, I3.Middle, I3.High, I4.Low, I4.Middle, I4.High, I5.Low, I5.Middle, I5.High\}$.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

Step 8: Let g=1, the candidate fuzzy coherent rule set is generated as follows: $(I1.Low \rightarrow I2.Low)$, $(I1.Low \rightarrow I_2.Middle \ldots I_{4.High} \rightarrow I5.High.$

Step 9: The fuzzy coherent rules are derived by the following substeps:

Step 9.1: Take a candidate fuzzy coherent rule "(I1.High → I4.High)" as an example. Since the consequence part is I4.High and the antecedent part is I1.High, the contingency table of (I1.High, I4.High) is then calculated. For instance, the count value of count (I1.High, I4.High) is 1.5 = (0 + 0.5 + 0 + 1 + 0) is shown in Table 5. Similarly all candidate fuzzy set is calculated in the same way.

Step 9.2: From the Table 5, based on the count value of all the candidate fuzzy sets, Check out this four conditions, $Q1 > Q2, Q1 > Q3, Q4 > Q2$ and $Q4 > Q3.$ Thus, (I1.High → I4.High) is put into the $FTCR_1$ and $FTCR_{all.}$

Step 9.3: After Step 9.2, there are two fuzzy coherent rules are generated which are $FTCR_1=\{(I1.Low \rightarrow I3.Middle), (I1.Middle \rightarrow I3.Middle)\}$. Since the $FTCR_1$ is not empty, it then go to STEP 10.

Step 10: From the derived $FTCR_{1,}$ the sets K and P is renewed as $K = \{I3.Middle, I3.High\}$and $P = \{I1.Low, I1.Middle\}$.

Step 11: set g=2.Since there are only two fuzzy regions with the same item in the set K, it cannot be used to form any candidate fuzzy coherent rule in this step.

Step 12: Because there is no candidate rule be formed, it then go to STEP 13.

Step 13: In this example, totally two fuzzy coherent rules are outputted in Table 6.

Table 6: The fuzzy temporal coherent rules.

Period	TID	$FTCR_{all}$	$count_{xy}$	$count_{x\sim y}$	$count_{\sim xy}$	$count_{\sim x\sim y}$
P1	T1	$(I1.Low \rightarrow I3.Middle)$	0.75	0.25	1	3.4
P2	T4	$(I1.High \rightarrow I2.High)$	1	2.5	2	1.5

V. EXPERIMENT EVALUATION

In our experiment, we have generated a synthetic dataset contains five thousands of transactions including 18 items in this proposed methodology. The items are used along with their purchase quantity. The membership function is used for generating fuzzy regions low, middle and high for each item.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

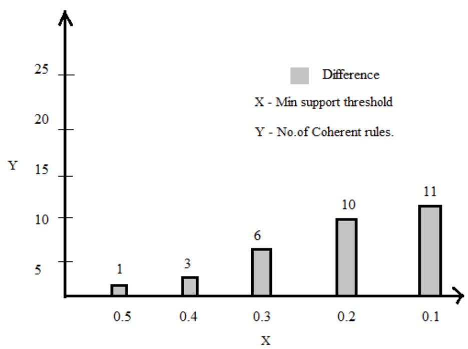

Fig. 2.The difference in the number of coherent rules between the FTCR approach and previous approach.

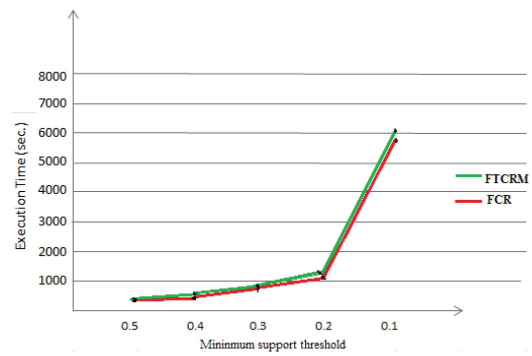

Fig. 3.The execution times of proposed and previous methodology.

The number of coherent rules obtained with the given minimum support threshold by the comparison of our proposed method with Fuzzy Coherent approach is shown in Fig.2. The execution time variations of each transaction with the given minimum support threshold of two approaches is displayed in Fig. 3.

VI. CONCLUSION AND FUTURE WORKS

In this paper, we have proposed a fuzzy temporal coherent rule mining algorithm from quantitative transactions without giving minimum support threshold. In this algorithm, initially converts quantitative transactions into fuzzy sets using the predefined membership functions. After, those fuzzy sets are gathered to produce candidate fuzzy coherent rules. For each candidate fuzzy coherent rule, the contingency table is then calculated and checks whether it satisfies the four conditions or not.

ACKNOWLEDGMENT

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

REFERENCES

1. R. Agrawal, T. Imielinski, A.N. Swami, Mining association rules between sets of items in large databases, Proceedings of the ACM-SIGMOD 1993 International Conference on Management of Data, pp. 207–216, 1993.
2. R. Agrawal, R. Srikant, Fast algorithm for mining association rules, The International Conference on Very Large Data Bases, pp. 487–499, 1994.
3. A. Savasere, E. Omiecinski, S.B. Navathe, An efficient algorithm for mining association rules in large databases, in: Proceedings of the 1995 International Conference on Very Large Data Bases, pp. 432–444, 1995.
4. MeeraNarvekar. ShafaqueFatma Syed, An Optimized Algorithm for Association Rule Mining Using FP Tree, Procedia Computer ScienceVol. 45, pp. 101-110, 2015.
5. C.C. Chan, W.H. Au, Mining fuzzy association rules, in: The Conference on Information and Knowledge Management, Las Vegas, pp. 209–215, 1997.
6. Chun-Wei Lin, Tzung-Pei Hong, Wen-Hsiang Lu, Linguistic data mining with fuzzy FP-trees, in: Expert Systems with Applications, Vol. 37, Issue 6, pp. 4560-4567, 2010.
7. AmalMoustafa, BadrAbuelnasr, Mohamed Said Abougabal, Efficient mining fuzzy association rules from ubiquitous data streams, in: Alexandria Engineering Journal, Vol. 54, pp. 163–174, 2015.
8. Yingjiu Li, PengNing, X. Sean Wang, SushilJajodia, Discovering calendar-based temporal association rules, in: Data & Knowledge Engineering 44, pp. 193–218, 2003.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

9. Chun-Hao Chen, Guo-Cheng Lan, Tzung-Pei Hong , Shih-Bin Lin, Mining fuzzy temporal association rules by item lifespans in : Applied Soft ComputingVol. 41, pp. 265–274, 2016.
10. K. Verma, O. P. Vyas, and R. Vyas, Temporal approach to association rule mining using T-tree and P-tree , Lecture Notes in Computer Science, pp. 651-659, 2005.
11. Chun-Hao Chen, Ai-Fang Li, Yeong-Chyi Lee, A fuzzy coherent rule mining algorithm in: Applied Soft Computing 13, pp. 3422–3428,2013.

BIOGRAPHY

R.Ratchambigah and R.Swetha are currently pursuing Final year Computer Science and Engineering in Sri Ramakrishna Engineering College, Coimbatore.

R. Anuradha is working as an Assistant Professor in Sri Ramakrishna Engineering College and currently pursuing research in Anna University. She has published papers in Journal of Intelligent and fuzzy system, IEEE Conference etc.