Deterministic and Fuzzy Model for Temporal Association Rule Mining

Anjana Pandey
University Institute of Technology, RGPV Bhopal

ABSTRACT
This paper explores the usage of deterministic and soft computing approaches in frequent item set mining in temporal data. In deterministic approach TPASCAL and PPCI algorithms are discussed in this paper. TPASCAL is based on counting inference method and PPCI combines progressive partition approach with counting inference method to discover association rules in temporal database. For effective knowledge discovery both Soft Computing and Data Mining can be merged. Soft Computing techniques such as fuzzy logic, rough sets aims to reveal the tolerance for imprecision and uncertainty for achieving tractability, robustness and low-cost solutions. Temporal fuzzy association rule on quantitative database and RSMAR and RSHAR which are used for mining of multidimensional association rules with rough set technology are discussed. It can be seen the algorithms is effective to settle with some problems. All the models developed here lead to superior performance and efficiency of mining temporal patterns as compared to existing algorithms.

Keywords
Data mining, temporal association rule, fuzzy logic, rough set, counting inference method

1. INTRODUCTION
Many data mining problems involve temporal aspects, with examples ranging from engineering to scientific research, finance and medicine. Temporal data mining [1] is a single step in the process of knowledge discovery in temporal databases that enumerates structures (temporal patterns or models) over the temporal data, and any algorithm that enumerates temporal patterns from, or fits models to, temporal data is a temporal data mining algorithm. Temporal data mining covers a wide spectrum of paradigms for knowledge modeling and discovery [2]. Since temporal data mining is relatively a new field of research, there is no widely accepted taxonomy yet. In view of above it is proposed to study and develop algorithms for mining temporal association rules. The objective of temporal association rules is to discover time-dependent correlations, patterns or rules between events in large volumes of data. The mining of temporal association rules is usually a two phase's process. The first phase is for generating all frequent temporal itemset with their support values .The second phase generates all temporal association rules that satisfy user defined parameter min_conf.

A number of algorithms have been developed for searching temporal association rules. One of the main challenges in mining association rules is developing fast and efficient algorithms that can handle large volumes of data, because most association rule algorithms perform computations over entire databases, which are very large. A Method like

Temporal-Apriori [3] takes more time to generate frequent patterns in Association rule Mining. This problem is mainly related to the number of operations required for counting pattern supports in the database. Huge space is required to perform the mining in Apriori based temporal association rule and it generates a huge number of candidates in case of a dateset, which is large and/or sparse. To overcome the problem of Temporal-Apriori, TPASCAL and PPCI algorithms are proposed in this paper.TPASCAL is based on counting inference method and PPCI combines progressive partition approach with counting inference method to discover association rules in temporal database where the items are allowed to have different exhibition periods. Partition algorithm to further improve efficiency, since it effectively reduces the number of database scans;

Another interesting problem concerning temporal association rule mining is the problem of mining temporal fuzzy association rule on quantitative database. Existing algorithms [4-7] are not well suited for this problem because most of them are based on Apriori algorithm and each of researchers treated all attributes (or all the linguistic terms) as uniform. However, in real-world applications, the users have more interest in the rules that contain fashionable items. As a remedy these problems we proposed a model for discovering interesting temporal patterns from weighted quantitative data. The mined patterns are expressed in fuzzy temporal association rules which satisfy the temporal requirement specified by the user.

Above mention algorithms are single dimensional and these algorithms having their own limitations. To overcome the problem of single dimension association rule, multidimensional algorithms have been proposed in this thesis. These algorithms are RSMAR and RSHAR which are used for mining of association rules with rough set technology. It can be seen the algorithms is effective to settle with some problems.

The remainder of the paper is organized as follows. Section 2 examines the concepts of temporal association rule mining and fuzzy calendar. Section 3 discusses the temporal association rule mining methodology and above mention algorithm. Result and discussion are explained in Section 4. Finally, Section 5 concludes the paper.

2. RELATED WORK
A Temporal Association Rule for d is an expression of the form $X \Rightarrow Y[t_1, t_2]$, where $X \subseteq R, Y \subseteq R \setminus X$, and $[t_1, t_2]$ is a time frame corresponding to the lifespan of $X \cup Y$.A temporal association rule has three factors associated with it: support, temporal support, and confidence[8].

The support of X in D, denoted by $S(X, D)$, is the fraction of the transactions in d that contains $X :| V(X, d) | / | d |$. The frequency of a set X is its support. Given a support threshold $\sigma \in [0,1], X$,is frequent if $s(X,d) \geq \sigma$. In this case, it is said that X has minimum support. Temporal support as the amplitude of the lifespan of an itemset.

Definition 1: The confidence of a rule $X \Rightarrow Y[t_1, t_2]$, denoted by conf $(X \Rightarrow Y,[t_1,t_2],d)$ is the conditional probability that a transaction of d , randomly selected in the time frame[t_1, t_2], that contains X also contains Y .

$$conf(X \Rightarrow Y,[t_1,t_2],d) = s(X \cup Y, l_{X \cup Y}, d) / s(X, l_{X \cup Y}, d)$$

Where $l_{X \cup Y} = \{[t_1,\ t_2]\}$. ……………………….(1)

Definition 2: Let P be a finite set of items, O a finite set of objects (e.g., transaction ids) and R $\subseteq$ O × P a binary relation (where (o, p) ε R may be read as "item p is included in transaction o").The triple D=(O, P, R) is called dataset [9]. Each subset p of P is called a pattern.

Definition 3. A pattern P is a key pattern if P ε min [P]. A candidate key pattern is a pattern where all its proper sub patterns are frequent key patterns.

Theorem 1. Let P be a pattern.

(i) Let pε P .then P ε [P \{p}] if and only if supp(P)= supp(P\{p})…………………. (2)

(ii) P is a key pattern if and only if supp(P) $\neq$ min pε P (supp(P\{p})) [15]………… (3)

Theorem 2 Let P is a non key pattern then supp (P) = min pεP (supp (P\{p})) [9]…..(4)

Calendar Schema: When temporal information is applied in terms of date, month, year & week form the term Calendar schema. It is introduced in temporal data mining. A calendar schema is a relational Schema R =(Gn : Dn,Gn-1 : Dn-1,……, G1 : D1) together with a valid constraint[2]. Each attribute G_i is a granularity name like year, month and week. Each domain D_i is a finite subset of the Positive integers. A calendar schema (year: {2007,2006….}, month: 1,2,3,4…12}, day{1,2,3…..31}) with the constraints is valid if that evaluates(yy,mm,dd) to true only If the combination gives a valid date while <1996,2,31> is not

Definition 4: A *basic* fuzzy calendar (FC), *A*, characterizes a fuzzy proposition about the collection of time intervals in a time granularity *U*, described by a membership function μ_A where

$\mu_A : U \to [0,1]$……………………(5)

for every time interval $T_i \in$ *U*. The function value $\mu_A(T_i)$ indicates the matching degree of T_i to A [8].

Definition 5*:* A fuzzy calendar (FC) is defined recursively as follows [10]:

1) A basic fuzzy calendar is a fuzzy calendar.

2) Let *A* and *B* be two fuzzy calendars. Then *A* **and** *B* ,*A* **or** *B* , **not** *A* , A **xor** *B* , and A **sub** *B* are also fuzzy calendars.

Definition 6: Let FC be the specified fuzzy calendar and w_i be the matching degree, or weight, of the time period T_i corresponding to partition P_i in the database D *and average weight of fuzzy region (Fw).* For a given itemset I $\subseteq I$, a transaction t is said to contain I if and only if I $\subseteq t_c$ .Let | P_i (I) | be the number of transactions containing itemset *I* in partition P_i .The *weighted count* of an itemset *I* in D denoted σ_D(I) ,is defined to be

$$\sigma_D(I) = \sum_{K=1}^{n}(|P_K(I)| \times w_K) \times fw \quad …(6)$$

The rough set method operates on data matrics, so called "Information System.It contains data about the universe U of interest, condition attributes and decision attributes. The goal is to derive rules that give information how the decision attributes depend on the condition attributes. By an information system S, S= {U, *At, V, f*}, where U is a finite set of objects, U= { $x_1, x_2, \ldots x_n$ }, *At* is a finite set of attributes, the attribute in *At* is further classified into two disjoint subsets, condition attributes C and decision attribute D, *At* = C $\cup$ D where

C= $c_1 \wedge c_2 \wedge c_3 \ldots\ldots \wedge c_n$ and

D= $d_1 \wedge d_2 \wedge d_3 \ldots\ldots \wedge d_n$ …………(7)

V= $\bigcup_{p \in A} V_p$,and V_p is a domain of attribute p…………………………..(8)

The function f performs a mapping code of condition attributes such that $c_1, c_2 \ldots c_n$ into one simple attribute C which can be added directly into the information system as one certain attribute, it will only posses one column in the information system, analogous

an item.

$f : U \times At \to V$ is a total function such that $f(x_i, q) \in V_q$ ………………….(9)

Let $B \subseteq At, x_i, x_j \in U$ .Define a binary relation *IND*, called an indiscernibility relation as follow[11]:

$$IND = \{(x_i, x_j) \in U^2 : p \in IND, p(x_i) = p(x_j)\}$$
……………………………..(10)

It can be said that x_i and x_j are indiscernible by a set of attributes B in *S if and only if*

$$p(x_i) = p(x_j) \quad …………………(11)$$

for every $p \in B$. One can check that *IND* is an equivalence relation on U for every $B \subseteq At$ .Equivalence classes of relations are called elementary sets in S . For any element x_i of U , the equivalence class of x_i in relation B is represented as $[x_i]_{IND.}$

3. METHODOLOGY OF RESEARCH WORK

A dataset [2] can be regarded as set of measurements taken from some environment or process. Each database consists of |D| transactions and, on the average, each transaction has |T| items. To simulate the characteristic of the exhibition period in each item, transaction items are uniformly distributed into database D with a random selection. Time dependent databases can be mined for deterministic model and fuzzy model. The methodology for different problems is given in subsections below.

3.1 Counting Inference Approach to Discover Calendar based Temporal Association Rules

To overcome the problem of Temporal-Apriori, we propose an algorithm Temporal-PASCAL (TPASCAL) to discover calendar –based temporal Association Rules. The TPASCAL uses Pattern counting inference that minimize as much as possible the number of pattern support counts performed when extracting frequent patterns. It is based on the notion of key patterns, where a key pattern is minimal pattern of equivalence class gathering all patterns common to the same objects of the database relation. With pattern counting inference, only the supports of the frequent key patterns (and some infrequent ones) are determined from the database, while supports of the frequent non key pattern are derived from those of the frequent key patterns.

Algorithm

The TPASCAL works in three Phases.

First Phase

i [Initialization]supp=1,key =true and P_0={Ǿ}

ii For all basic time interval e_0 do begin P_1 (e_0) $\leftarrow$ { frequent 1 pattern in T[e_0]}.

iii. Find large P_1 (1-itemsets) and divide them into set of key and non-key patterns

iv for all star pattern e that cover e_0 do update P_1 (e) using P(e_0) ;

Second Phase

i. If the number of key patterns is small then database scans only for superset of key patterns and finds all large itemsets by help of non-key patterns.

ii. If number of key pattern is large then algorithm has two options.

In ***option 1*** scan the database for calculating support of candidate 2-itemset and each key pattern store address of records (transactions) that contain particular key pattern.

In ***option 2*** rather than storing address of records, create new database for these records.

After completion of ***option 1*** or ***option 2***, it moves to phase *third*.

Third Phase

1) Finds rest of large itemset P_k (e_0) (for $k \geq 3$) only accessing those records that contain key patterns. for all star pattern e that cover e_0 do update P_k (e) using P_k (e_0)

3.2 PPCI Algorithm for Mining Temporal Association Rules in Large Database

Progressive Partition Counting Inference (PPCI) method combines progressive partition approach with counting inference method to discover association rules in temporal database where the items are allowed to have different exhibition periods and the determination of their supports is made in accordance with their exhibition periods. Suppose $(X{=}{>}Y)^{t,n}$ is temporal association rule where t is the latest -exhibition start time of both itemsets X and Y and n denotes the end time of the transaction in the database.

An association rule X=>Y is termed to be a frequent temporal association rule $(X{=}{>}Y)^{t,n}$ if and only if its probability is larger than minimum support required,i.e.P($X^{t,n}$U $Y^{t,n}$)> min_supp,and the conditional probability P($Y^{t,n}$|$X^{t,n}$) is larger than minimum confidence needed ,i.e., P($Y^{t,n}$|$X^{t,n}$) > min_conf.Instead of using the absolute support threshold S^A = $\lceil |D| * \min\_\sup p \rceil$ as a minimum support threshold for each item.

Algorithm

Input:PPCI(p,q,min_supp,min_leng)

p-Starting time is fixed,

q-ending time is q,

min_supp-user defined threshold

min_leng-filtering threshold for frequent itemsets to satisfy the minimal length

Output:Frequent itemset and temporal association rules

The PPCI works in Three Phases.

First Phase

1. A database D is partitioned into n partitions based on the exhibition periods of items

and set PS = Ǿ where n=q-p+1

2. Generates 1-itemset with progressive screen. Let C1 be the set of progressive

candidate 1-itemsets generated by database D.

$P_1 \leftarrow$ { frequent 1 pattern };

For all $p \in P_1$ do begin

p.pred_supp $\leftarrow$ 1; p.key $\leftarrow$(p.supp $\neq$ 1);

end;

Second phase

//Generate Candidate C_k-Itemset with the counting inference method

1. for(K=2,$P_{k-1} \neq \Phi$,K^{++}) do begin $C_k \leftarrow$TPASCAL-Gen(P_{k-1})
2. If $\exists$ C $\in$ C_k | C.key then for all partition starting from p to q

   call Subprocedure 1 //PS contains supp for those pattern which are key pattern

3. for all item $\in$ PS,if (item $\in$ PS) and item.supp = pred_supp then P_k.item = item;
4. P_k.item.supp = supp;P_k.item.key = F
5. return $\cup_k P_k$

Third phase

// TPASCAL-Gen

1. $P_{k-1,}$ the set of frequent (K-1) patterns p with their support p.supp and the p.key flag.

2. Insert into C_k select p.item$_1$,p.item$_2$,……p.item$_{k-1}$,q.item$_{k-1}$ from P_{k-1} p, P_{k-1} q

Where p.item$_1$ = q.item$_1$, p.item$_{k-2}$ = q.item$_{k-2}$, p.item$_{k-1}$ < q.item$_{k-1;}$

3. for all (k-1) subsets s of c do begin ,if s $\notin P_{k-1}$ then delete c from $C_{k\,;}$

4. else begin c.pred_supp ← min(c.pred_supp,s.supp);

5. if not s.key then c.key ←false ; if not c.key then c.supp ←c.pred_supp;

6. return C_k

(C_k ,the set of candidate k patterns c each with the flag c.key, the value c.pred_supp,and the support c.supp if c is not a key pattern)

3.3 A Model for Mining Temporal Weighted Association Rule

The problem of mining temporal weighted association rules based on fuzzy approach. Temporal requirement specified by human beings tend to be ill-defined or uncertain. To deal with this kind of uncertainty, fuzzy calendar algebra is applied to allow the users to describe desired temporal requirements in fuzzy calendar [10] easily and naturally. The mined patterns are represented in terms of fuzzy temporal association rules and are subject to the time requirements specified by the user. Users can define complicated calendars with multiple time granularities and different preferences. Different time intervals can have different weights corresponding to their matching degrees to the specified fuzzy calendar. This can be of great help for users to discover the knowledge in the time intervals that are of interest to them. A fuzzy data mining algorithm to discover fuzzy temporal weighted association rules for quantitative data is proposed. The existing algorithm is not well suited for mining fuzzy temporal weighted association rules from quantitative data. In proposed algorithm to generate the Candidate itemset we are using scan reduction techniques .This technique help to reduce unnecessary scans over the database. The candidate itemsets can be computed in an efficient way and unnecessary scans over the database can be saved.

Algorithm

1. Get membership functions, minimum support, minimum confidence
2. Assign weight to each fuzzy membership for each attribute (categorical)
3. Calculate support for each fuzzy region
4. If support > minimum, OK
5. If confidence > minimum, OK
6. If both OK, generate rules

3.4 Rough Set Model for Discovering Hybrid Association Rules

The problem of mining hybrid association rules with rough set [11] approach is investigated as the algorithm RSHAR.The RSHAR algorithm consists of two steps mainly. At first, to join the participant tables into a general table to generate the rules which is expressing the relationship between two or more domains that belong to several different tables in a database. Then we apply the mapping code on selected dimension, which can be added directly into the information system as one certain attribute. To find the association rules, frequent itemsets are generated in second step where candidate itemsets are generated through equivalence classes and also transforming the mapping code in to real dimensions. The searching method for candidate item set is similar to Apriori algorithm.

Algorithm

The RSHAR works in two Phases.

First phase

// First we apply the Combine Dims *algorithm* to combine the selected dimensions in order to provide the framework for mining hybrid association rules

1. Select the dimension d_2, ….. d_m From the general tables where ($d_2 = duser_2$) And ($d_m = duser_m$).

2. Then we combine the two dimensions and provide one mapping code

Second phase

// in second phase we use the table in the GenFI algorithm to discover frequent itemsets on hybrid association rules in transaction database

1. Frequent itemsets are generated in second phase where candidate itemsets are generated through equivalence classes.
2. And also transforming the mapping code in to real dimensions.

3.5 Rough Set Model for Discovering Intradimensional Association Rules

The problem of mining of intradimensional association rules with rough set approach is investigated as the algorithm RSMAR.The RSMAR algorithm consists of two steps. At first, we join the participant tables into a general table to generate the rules which is expressing the relationship between two or more domains that belong to several different tables in a database. Then we apply the mapping code on selected dimension, which can be added directly into the information system as one certain attribute. To find the association rules, frequent itemsets are generated in second step where candidate itemsets are generated through equivalence classes and also transforming the mapping code in to real dimensions.

4. RESULTS AND DISCUSSION

We have made an effort in direction of building deterministic and fuzzy models concerning time based databases. The key results are summarized as under:

4.1 To assess the performance of Algorithm TPASCAL, several experiments are performed on a computer with a CPU clock rate of 450 MHz and 512 MB of main memory. The transaction data resides in the NTFS file system and is stored on a 30GB IDE 3.5” drive with a measured sequential throughput of 10MB/second. The simulation program was coded in C++. We generated several different transaction databases from a set of potentially frequent item sets to evaluate the performance of TPASCAL. These transactions mimic the grocery items in a grocery sales data. We show the experimental results from synthetic transaction data so as to obtain results of different workload parameters. Each database consists of |D| transactions and, on the average, each transaction has |T| items. To simulate the characteristic of the exhibition period in each item, transaction items are uniformly distributed into database D with a random selection.

As will be shown by our experimental results, the execution time of TPASCAL is, in orders of magnitude, smaller than that required by Temporal-Apriori. As shown in Fig. 1, the experimental results are consistent for various values of n, |L| and N on data set T10-I4-D100, e.g., T10-I4- D100(N20-L4-n20). For interest of space, we only report the results on |L| = 2, 000 and N = 10, 000 in the following experiments. Fig. 1 shows the relative execution times for both two algorithms as the minimum support threshold is decreased from 1 percent support to 0.1 percent support. When the support threshold is high, there are only a limited number of frequent item sets produced. However, as the support threshold decreases, the performance difference becomes prominent in that TPASCAL significantly outperforms Temporal-Apriori. Explicitly, TPASCAL is in orders of magnitude faster than Temporal-Apriori, and the margin grows as the minimum support threshold decreases. In fact, TPASCAL outperforms Temporal_Apriori in both CPU and I/O costs.

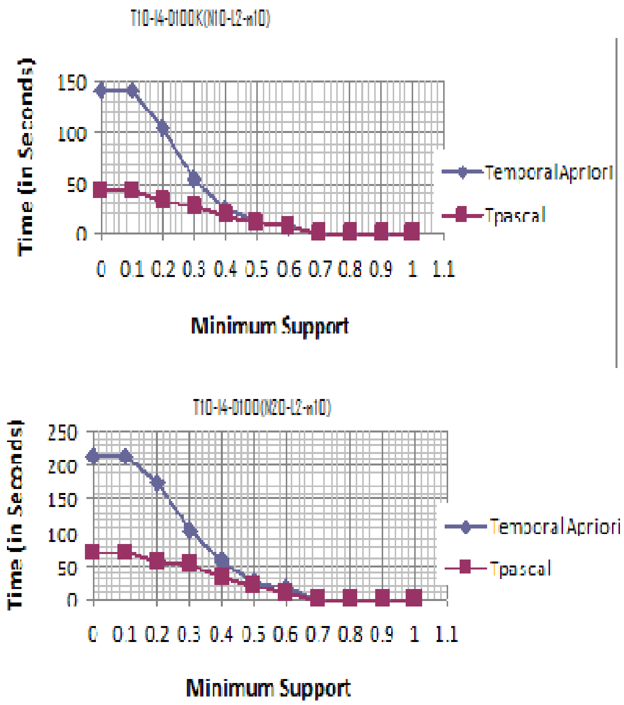

Fig 1 Relative Performance Studies of TPASCAL and TemporalApriori

4.2 PPCI algorithm, efficiently extracts frequent pattern in large database with partitioned database and time constraints such as calendar pattern .We are comparing the performance of our method with that of the "Temporal Apriori" algorithm. Under PPCI the cumulative information of mining previous partitions is selectively carried over toward the generation of candidate item set for subsequent partitions. Algorithm PPCI not only significantly reduces input/output cost and CPU cost by the concept of progressive counting and counting inference technique but also effectively control memory utilization by proper partitioning. It was noted that by improvement achieved as the size of the databse increases. We will evaluate and compare the performance of this combination of algorithm on synthetic datasets, T20I6D100K, T25I10D10K and T25I20D100K that mimic market basket data. The simulation program was coded in C++.

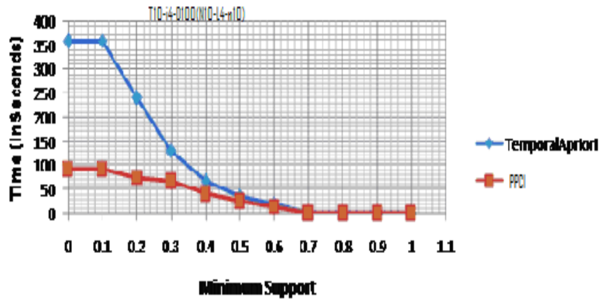

Fig 2 Relative Performance Studies of PPCI and TemporalApriori

4.3 A fuzzy temporal weighted association rule mining method is proposed for quantitive or categorical data, which is more flexible, natural and understandable. This is an extension of the fuzzy quantitive association mining problem. In this generalization, the fuzzy calendar algebra is applied which helps users to describe desired temporal requirements in fuzzy calendars easily and naturally. Fuzzy calendars capture the way in which human beings reason with time intervals in daily life. For the proposed method, it is easy to add User's guide to data mining procedure, so it is quicker to find association rules interested by users, and reduce the work burden of choosing interesting rules. The proposed method satisfies partitions method and scan reduction technique which decreases computation time, at the same time, uninteresting rules can be pruned early because of assigning weights to items, which also results in the reduction of execution time. It is assumed that transactions in the databases are collected and ordered by time. Therefore, the oldest data always expire first and the newest data are inserted into the last partition of the database. In the proposed algorithm, we assign fuzzy regions to each categorical attribute, which can describe categorical attribute in linguistic language more flexible according to situations especially when the number of categorical regions is large. The proposed fuzzy data mining method can also resolve conventional binary value problem by using a degraded membership functions and let all weights be 1.

The effectiveness of our approach is assessed by experimenting with a real life datasets. The data had 5 quantitative attributes: Age, monthly-income, credit, Risk and Result. There are 5000 records in the data .The above five quantitative attributes are used where three fuzzy sets are defined for Age, Income Risk and two fuzzy sets are defined for credit and result. Figure 3 shows the increase of the number of rules as a function of average weight, for five different support thresholds. The minimum confidence was set to 0.25.We used five intervals from which random weights generated. The increase of the number of rules is close to linear with respect to the average weight.

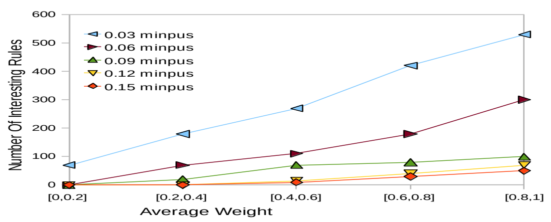

Figure 3: Random weight Intervals

Figure 4, shows how the performance varies with the number of records from 1000 to 5000 for five different support thresholds. The graphs show that the method scales quite linearly for this dataset

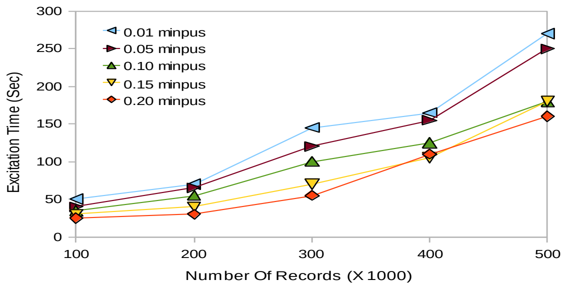

Figure 4: Scale-up: Number of Records

4.4 The RSHAR is proposed for mining of hybrid association rules. Mining rules with the RSHAR algorithm is two step processes: First the CombineDims *algorithmis employed* to combine the selected dimensions in order to provide the framework for mining hybrid association rules. Then, the GenFI algorithm is employed to discover frequent itemsets in the transaction database. For the new information system, the searching of frequent itemsets is easy based on the concept of equivalence class. The algorithm provides better performance improvements. The gap between the RSHAR and Apriori algorithms becomes evident with the number and size of patterns identified and the searching time reduced.

To evaluate the efficiency of the proposed method, the RSHAR, along with the Apriori algorithm, is implemented at the same condition. A sample sales database is used which contains three dimensions (i.e. customer dimension, product, dimension, Promotions dimension) and one sales fact table.We perform our experiments using a Pentium IV 1,8 Gigahertz CPU with 512MB.The minimum support of Apriori algorithm is 0.45%, and the computation times and the numbers of frequent itemsets found by the two algorithms are shown in Figure 5.

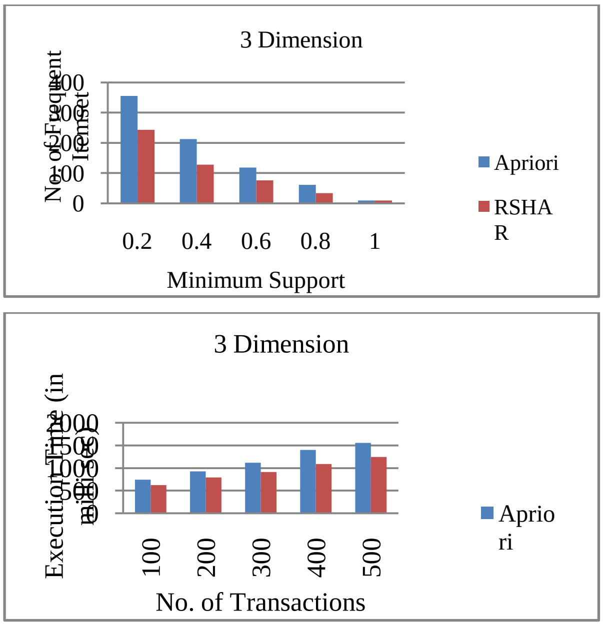

Figure 5. (a) No of frequent itemset. (b) Computation Time

The experimental results in Figure 5 show that the RSHAR performs batter and more rapid than the Apriori algorithm. The RSHAR is not only eliminating considerable amounts of data, but also decreasing the numbers of database scanning, thus reducing the computation quantities to perform data contrasts and also memory requirements.

4.5 Multidimensional rules with no repeated predicates are called *interadimension association* rules .On the other hand, multidimensional association rules with repeated predicates, which contain multiple occurrences of some predicates, are *called hybrid-dimension association rules* .The rules may be also considered as combination (hybridization) between intradimension association rules and interdimension association rules.The problem of mining of intradimensional association rules with rough set approach is investigated as the algorithm RSMAR.RSMAR and RSHAR both perform better than Apriori algorithm. The RSMAR and RSHAR have different models which are used in algorithm.

5. CONCLUSION

This paper deals with the problem of mining temporal association rules in large databases for the deterministic situations as well as the problems involving uncertainty. In this paper new algorithms for deterministic situations TPASCAL and PPCI were proposed and extended to variable number of snapshots. Both algorithms are based on the Counting Inference principle. The second one uses Partitioning method to reduce the database scan. Both algorithms improve the performance in terms of speed, accuracy and resource utilization as compared to existing algorithms.

Also, a new problem definition for uncertainty was introduced for temporal association rules mining, which uses the concept of Fuzzy calendar. This algorithm was known as fuzzy temporal weighted association rule for quantitive database. The proposed method satisfies partition method and scan reduction technique which decreases computation time, at the same time; uninteresting rules can be pruned early because of assigning weights to items, which also result in the reduction of execution time. This algorithm is tested on real-life datasets with variable values of input parameters.

All above mentioned problems are single dimension, multidimensional problems like hybrid association rule mining and Intra association rule mining are also discussed in this thesis. These algorithms are RSHAR and RSMAR respectively. Both algorithms are based on rough set method and perform better than Apriori algorithm.

These models can handle more realistic problems of deterministic situations as well as situations involving uncertainty of relationship among items. Further the new approaches and algorithms can lead to increase in speed of searching and matching, reduction in number of transactions (a kind of instance selection), reduction in subsets per transaction, and reduction in number of candidates by pruning .All the models developed here lead to superior performance and efficiency of mining temporal patterns as compared to existing algorithms. It is needless to say that, it will lead to contribution of knowledge in the area of data mining especially temporal association rule mining.

6. REFERENCES

[1] David L.Olsen ,Dursan Delen 2003"Advance Data Mining Techniques" ISBN: 978-3-540-76916-3 Springer

[2] Hipp. J., U. Güntzer, Nakhaeizadeh G. 2000. "Algorithms for association rule mining – a general survey and comparison", *SIGKDD Explorations* 2:1, 58–64.

[3] Lee Chang-Hung, Ming-Syan Chen, July/August 2003 "Progressive Partition Miner: An Efficient Algorithm for Mining General Temporal Association Rules" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 15, NO. 4,

[4] Roddick, J. F. & Spiliopoulou, M. 2002, 'A survey of temporal knowledge discovery paradigms and methods', IEEE Transactions on Knowledge and Data Engineering 14(4), 750–767.

[5] Han, J. & Kamber, M. 2001, Data Mining : Concepts and Techniques, Academic Press, San Diego.

[6] Wang X. S., Jajodia S., Y. Li, and Ning P. June 2001. "Discovering calendar-based temporal association rules". In Eigth International Symposium on Temporal Representation and Reasoning,TIME-01, pages 111–118, Civdale del Friuli, Italy,

[7] Zhang W., 2007 " Mining fuzzy quantitative association rules", *Proceedings of IEEE International Conference on Tools with Artificial Intelligence 1999 Piscataway, NJ*, IEEE Press, 1999, pp. 99-102. Proceedings of the 40th Hawaii International Conference on System Sciences

[8] Kuok C. , Fu A. and H. Wong, 2007 " Mining fuzzy association rules in databases", *ACM SIGMOD Record*, 27, 1998, pp. 41- 46. Proceedings of the 40th Hawaii International Conference on System Sciences - [9] P. Bosc and O. Pivert, On some fuzzy extensions of association rules, *Proceedings of IFSA-NAFIPS 2001,Piscataway, NJ*, IEEE Press, 2001, pp. 1104-9

[9] Au W. H. and Chan K. C. C., Aug. 1999 "Farm: A data mining system for discovering fuzzy association rules," in *Proc. 8th IEEE Int. Conf. Fuzzy Systems*,Seoul, Korea, , pp. 1217–1222

[10] Lee Wan-Jui and Lee Shie-Jue" December 2004 "Discovery of Fuzzy Temporal Association Rules" IEEE Transactions on Systems, Man, And Cybernetics-B: Cybernetics,Vol.34, No.6,

[11] Lee Wan-Jui and Shie-Jue Lee DECEMBER 2004 "Discovery of Fuzzy Temporal Association Rules" IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 34, NO. 6.