

doi:10.21311/001.39.1.25

An Efficient Approach to find Similar Temporal Association Patterns Performing Only Single Database Scan

Vangipuram Radhakrishna¹, Dr.P.V.Kumar², Dr.V.Janaki³

¹ Faculty of Information Technology, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India

²Former Professor, Department of Computer Science and Engineering, Osmania University, Hyderabad, India

³Professor & Head, Department of Computer Science and Engineering, Vaagdevi Engineering College, Warangal, India.

¹Email :vrkrishna2014@gmail.com

Abstract. Traditional frequent pattern mining algorithms use concept of support which is evaluated to a single numeric value. This concept of traditional support value does not hold good to discover similar temporal association patterns for a known reference sequence and threshold value. In temporal sense, support value is sequence of support values and not a single support value. This requires use of a suitable distance function to find the degree of dissimilarity. In this proposed method, we mainly aim at four research objectives. The first objective is to obtain temporal patterns in a single database scan. The second objective is to introduce the concept of negative support sequence to find temporal association patterns. The third objective is to design novel expressions to find the lower bound support sequence and upper bound support sequence by defining two functions called sum and XOR over temporal patterns. The fourth objective is to validate designed expressions. The proposed approach of finding similar temporal patterns is space efficient as it performs only a single database scan and time efficient as it finds temporally similar patterns using only a single database scan.

Keywords: Temporal, Association Pattern, Transaction, Upper bound, Lower bound

1. INTRODUCTION

For the past four decades, significant research contribution from database community has been towards studying temporal databases, temporal data mining techniques and various aspects of temporal information systems. In the year 1986, a summary of temporal database research discussed at various symposiums, workshops and research carried out at various research labs was first published as the ACM SIGMOD Record. The importance of temporal databases coined and came in to existence with IEEE Data Engineering devoting a complete issue for temporal databases in 1988. Consequently in year 1990 and 1992, two research papers that contributed to survey on temporal databases were published. In (Gultekin Ozsoyoilu et.al, 1995), authors perform survey on temporal and real time databases.

A detailed literature survey on temporal databases, temporal database methodologies and techniques, applications is discussed in (Srivatsan Laxman et.al, 2006). The theory of temporal databases, design and implementation is discussed in very interesting manner by (Tansel et.al, 1993) which is edited volume of papers presented at symposiums and workshops as various chapters. In (Matteo Golfarelli et.al, 2009), authors perform detailed literature survey on temporal data warehouses. (Alexander et.al, 2003) discussed database support for database applications. (Tansel et.al, 2007) work involves finding temporal association rules from temporal databases. An approach to discover temporal association rules from publication databases is addressed in research contributed by authors (Chang-Hung Lee et.al, 2001). In (Toon Calders et.al, 2001), authors define the approach for finding frequent items using the approximation method by computing upper bound and lower bound of items. An approach for finding frequent items in spatio temporal databases is discussed in (Cheqing Jin et.al, 2004). A method for finding temporal frequent patterns through using TFP-tree is discussed in (Long Jin et.al, 2006).

The authors (Jin Soung Yoo et.al 2009), use the Euclidean distance measure as dissimilarity measure to find temporally similar association patterns of interest to the user. This work is further extended in (Jin Soung Yoo et.al, 2012). The authors approach includes finding upper-lower bound, lower-lower bound and lower bound distances to discover similar temporal patterns. They define two approaches for discovering temporally similar patterns. The drawback of this approach is that the algorithm designed requires knowing true support values of temporal pattern of size, k , for deciding if a pattern of size, $k+1$ is temporally similar or temporally dissimilar. This initiates the necessity for scanning database more than once to obtain true support sequences of temporal pattern at level- k .

In present approach for discovering similar temporal association patterns, we overcome this disadvantage and also eliminate the necessity to compute true support values of its sub-patterns and the need to consider support sequence, support values of all possible subsets of size- k temporal patterns to find lower and upper bound support sequences of temporal patterns of size, $k+1$. A recent survey on temporal databases and data mining techniques is discussed in (Vangipuram et.al, 2015). A method for discovering temporally similar patterns is discussed in

(Vangipuram Radhakrishna et al., 2015) which uses representation of Venn diagram.

2. MOTIVATION

This research is mainly motivated from research contribution of authors (Jin Soung Yoo et al. 2012). The present approach which we propose is the novel approach which can be used to efficiently find similar temporal association patterns of interest w.r.t a given reference support time sequence, R and a user specified threshold constraint to overcome following disadvantages of (Jin Soung Yoo et al. 2012).

1. We eliminate, multiple scan of temporal database required to find support sequences of temporal patterns generated at different levels.
2. We overcome, the disadvantage of approach followed in (Jin Soung Yoo et al. 2012), which requires repeated scanning of the temporal database to know true support values of temporal association patterns of level-($k-1$) when finding temporal association patterns at level- k .
3. We overcome the requirement to know the true support sequence values of all subsets of an itemset as followed in (Jin Soung Yoo et al. 2009).

The problem of mining frequent patterns in databases is addressed extensively in the literature. Conventional algorithms designed to obtain frequent patterns are not applicable to find frequent patterns from temporal database of time stamped transaction sets. This is because, all these algorithms which are used to find frequent patterns are essentially based on support values which are actually independent of time stamp of transactions.

In our case i.e. for temporal database of time stamped transactions, these transactions have time stamps and hence the support of the itemsets is in the form of a vector representing support values computed for each time slot. This makes the conventional approach unsuitable to obtain frequent temporal patterns. Also, the popular Euclidean distance measure which is used to find distance between any two vectors does not satisfy the monotonicity property (Jin Soung Yoo et al. 2012). In the present work, we consider the problem of mining similarity profiled temporal patterns from set of time stamped transaction sets of a given temporal database. We demonstrate using a case study how the proposed approach may be used to find similar temporal frequent patterns.

3. MOTIVATION

3.1. Problem Definition

The objective is to use and apply concept of Venn diagrams to find temporally similar patterns by introducing two types of support sequences namely positive and negative support sequences so as to retrieve and discover similar temporal association patterns which are of user interest by performing only single database scan. This is achieved through designing and validating generalized formal expressions for computing support sequences and validate suitability of proposed approach to find temporal association patterns. The proposed approach of finding temporal patterns is based on algorithm design strategy called dynamic programming technique as it makes use of support time sequence of temporal patterns obtained in previous stage aiming at both time and space optimizations.

3.2. Problem Statement

Given a

- i) Finite set of singleton items (elements) denoted by I ,
- ii) D_{temporal} , formally represented as $D_{\text{temporal}} = D_1 \cup D_2 \dots \dots \cup D_n$ is a temporal database of time stamped Transaction sets such that for any value of i and j , whenever $i \neq j$, we have both D_i and D_j to be disjoint.
- iii) $T = t_1 \cup t_2 \dots \dots \cup t_n$, is the time period such that for any two values i, j whenever $i \neq j$, we have both T_i and T_j to be disjoint sets.
- iv) Each transaction record, d in D_{temporal} , is formally represented as a 2-tuple $\langle \text{timestamp}, \text{itemset} \rangle$, where timestamp and itemset, I are subsets of T and I respectively.
- v) $R = \langle r_1, r_2, r_3, \dots, r_n \rangle$ is a reference support sequence over time slots t_1, t_2, \dots, t_n and a user specified dissimilarity value denoted by θ_{th} . We denote reference as **Ref** in this paper.
- vi) A distance measure, denoted as $f_{\text{similarity}}(P, Q): \rightarrow R^n$ is a function of two support sequences P and Q

The objective is to “Discover set of all temporal patterns, I or I_{temporal} , which are subsets of I such that each of these itemsets represented by I , satisfy the condition $f_{\text{similarity}}(S_I, \text{Ref}) \leq \theta_{\text{th}}$ where S_I is the sequence of support values of I at time slots t_1, t_2, \dots, t_n

3.3. Objective

The main objective of the present research is to essentially retrieve or discover similar temporal association patterns or frequent patterns with respect to the specified reference sequence which satisfy the condition $f_{\text{similarity}}(S_i, \text{Reference}) \leq \theta_{\text{th}}$ by scanning the temporal database only once. The basic idea is to make use of Venn diagram concept and effectively use the support sequences of itemsets found initially after the first scan of the database. In order, to find the upper-lower bound and lower-lower bound, lower bound distances we adopt the method outlined in (Jin Soung Yoo et.al 2009). For dissimilarity computation, we use the traditional distance measure (Euclidean) to find the distance between itemset support sequence and the reference sequence, however normalized Euclidean distance must be considered, if distance exceed upper bound limit (usually > 1). Since we use the distance measure (also called as similarity measure) to find the temporal association patterns, we call these temporal patterns as similarity profiled temporal association patterns.

3.4. Basic Terms and Terminology

This section covers the definitions and introduction to basic terms and terminology used for discovering temporally similar patterns.

Positive itemset: Any finite itemset, $I' \subseteq I$, is called a positive itemset, if we compute the probability of its existence in the temporal database. For example, the itemsets represented as A, AB are called positive itemsets. We use terms itemset and pattern interchangeably in this paper.

Negative Itemset: Any finite itemset denoted by, $I' \subseteq I$, said to be a negative itemset, if we compute the probability of its absence in the temporal database. For example, the item sets represented as $\bar{A}, \overline{AB}, \bar{B}$ are called negative itemsets. We may also denote, $\bar{A}, \overline{AB}, \bar{B}$ as $A', (AB)'$ and B' respectively.

Positive Support: The support value computed for positive itemset is called as positive support.

Negative Support: The support value computed for negative itemset is called as negative support.

Support Sequence: The support sequence of a temporal pattern is an n-tuple denoted by $S_{\theta}(I') = \langle S_{t_1}, S_{t_2}, S_{t_3}, \dots, S_{t_n} \rangle$, where each S_{t_i} is the support value of itemset, I' at time slot, t_i . Formally, $S_{\theta}(I') = \bigcup_n \{ S_{t_i} \mid \text{time slot, } t_i, \text{ varies from } t_1 \text{ to } t_n \}$ where the symbol, \bigcup_n , denotes union of all support values computed for each time slot, t_i .

Positive Support Sequence, $S_{\theta_p}(I')$: The support time sequence, S_{θ} , of a temporal pattern $I' \subseteq I$ denoted by, $S_{\theta}(I')$, is defined as positive support sequence, denoted by $S_{\theta_p}(I')$, if and only if, the support of each element, S_{t_i} , in the support sequence is obtained for the existence of itemset I' in temporal database.

Negative Support Sequence: The support time sequence, S_{θ} , of a temporal pattern $I' \subseteq I$ denoted by, $S_{\theta}(I')$, is called negative support sequence, denoted by $S_{\theta_N}(I')$, if and only if, the support of each element, S_{t_i} , in the support sequence is computed for the absence of itemset I' in temporal database. The negative support sequences denote the probability of the temporal pattern not existing in the time slots, t_1, t_2, \dots, t_n .

True Distance: It is formally defined as the actual distance between support sequence vector of a temporal pattern and the reference support sequence vector, obtained using Euclidean distance (normalized).

Upper Lower Bound (ULB): The maximum possible distance between upper bound support sequence of a temporal pattern and the reference sequence vector is formally defined as upper-lower bound distance (Jin Soung Yoo et.al 2012).

Lower Lower Bound (LLB): The minimum possible distance between lower bound support sequence of a temporal pattern and the reference sequence vector is formally defined as lower-lower bound distance (Jin Soung Yoo et.al 2012). This value indicates that the support value at time slot t_i , cannot be less than this value.

Pruning: The process of discarding or eliminating all those temporal patterns which do not satisfy user defined constraints is called pruning. A temporal pattern, denoted by I' , is considered temporally similar pattern if and only if, every subset I' of I is also temporally similar.

3.5. Proposed Algorithm to Find Temporal Patterns

The algorithm to find temporal association a pattern using proposed approach is discusses below

Input:

- L or I Finite set consisting of single items
- $D_{temporal}$ Temporal database of transactions
- R Reference Sequence
- $f_{similarity}$ Euclidean Distance measure
- θ Threshold
- k Size of Itemset
- I Finite set of itemsets
- I' Itemset, subset of I
- N Number of items in Set, L

Output:

Set of all temporal patterns, I' , which are subsets of I, such that each of these temporal itemsets represented by I' , satisfy the condition $f_{similarity}(S_I, R) \leq \theta$ with S_I being support sequence of temporal pattern, I' at time slots t_1, t_2, \dots, t_n

Step1: Find Negative and Positive support values

Consider each item from finite itemset and obtain positive and negative support values for every time slot. We call these singleton pattern as positive and negative singleton temporal pattern of size one, $|S|=1$

Step2: Find Negative and Positive support sequences

Obtain negative and positive temporal support sequences from support values of negative and positive singleton temporal patterns. These negative and positive support sequences are mathematically denoted using notations S_{θ_N} and S_{θ_p} respectively. We classify temporal patterns obtained from temporal databases in to three classes. Temporal patterns of class-1 include all singleton items. Temporal patterns of class-2 include, all temporal patterns of size, $|S|=2$. Finally, temporal patterns of class-3 include all patterns whose size, is more than two, denoted as $|S|>2$.

Step-3: Find Similar Singleton Temporal Patterns (P, Q, R)

This includes determining if a singleton temporal pattern is similar or dissimilar w.r.t a given reference support sequence. To achieve this, we must consider each singleton temporal pattern and obtain Euclidean distance of this pattern to specified reference support sequence. Compare obtained distances with user defined threshold. All singleton temporal patterns, whose distance is less than user defined threshold value, are marked similar temporal patterns and remaining set of temporal patterns are marked as dissimilar temporal patterns. All these temporal patterns are called class-1 temporal patterns.

However, to generate class-2 ($|S|=2$) and class-3 temporal patterns ($|S|>2$), we choose to retain all singleton temporal patterns whose approximate upper bound distance value (ULB_{approx}) w.r.t reference sequence is less than or equal to user defined threshold value.

Step-4: Find Similar Singleton Temporal Patterns (P, Q, R) // Class-2 or Level-2 temporal association patterns (PQ, PR, QR....)

Set $|S|=|S|+1$ i.e Size, $|S|=2$. This step includes finding temporal patterns of size, equal to two. We call all such patterns as class-2 or level-2 patterns. To find similar temporal association patterns of at level-2, generate all possible combinations of class-2 temporal patterns from class-1 temporal patterns marked in step-3. We denote a class-2 temporal association patterns using notation $T_m T_n$. Each temporal pattern generated shall be of the form $T_m T_n$ with T_m and T_n denoting singleton patterns. For temporal patterns of size, $|S|=2$, denoted by $T_m T_n$, we compute both maximum and minimum bound on support sequences. To compute support sequences of temporal itemset of size, $|S|=2$, we use the equation defined by equation (1),

$$T_m T_n = \frac{1}{2} [T_m + T_n - T_m \overline{T_n} - T_n \overline{T_m}] \tag{1}$$

For each temporal pattern denoted by $T_i T_j$, we must find upper bound and lower support time sequence of temporal patterns $T_m \overline{T_n}$ and $T_n \overline{T_m}$ using equation 2 to equation 5. The idea is to store lower and upper bounds of support sequences of such temporal patterns to eliminate multiple scanning of temporal database.

$$UBSTS(T_m \overline{T_n}) = \langle \min(S_{m1}, \overline{S_{n1}}), \min(S_{m2}, \overline{S_{n2}}), \min(S_{m3}, \overline{S_{n3}}) \dots \dots \dots \min(S_{mk}, \overline{S_{nk}}) \rangle \quad (2)$$

$$LBSTS(T_m \overline{T_n}) = \langle \max(S_{m1} + \overline{S_{n1}} - 1, 0), \max(S_{m2} + \overline{S_{n2}} - 1, 0), \dots \dots \dots \min(S_{mk} + \overline{S_{nk}} - 1, 0) \rangle \quad (3)$$

$$UBSTS(T_n \overline{T_m}) = \langle \min(S_{n1}, \overline{S_{m1}}), \min(S_{n2}, \overline{S_{m2}}), \min(S_{n3}, \overline{S_{m3}}) \dots \dots \dots \min(S_{nk}, \overline{S_{mk}}) \rangle \quad (4)$$

$$LBSTS(T_n \overline{T_m}) = \langle \max(S_{n1} + \overline{S_{m1}} - 1, 0), \max(S_{n2} + \overline{S_{m2}} - 1, 0), \dots \dots \dots \min(S_{nk} + \overline{S_{mk}} - 1, 0) \rangle \quad (5)$$

Here, for each temporal pattern denoted by $[T_m T_n]$ of size, $|S|=2$, T_m and T_n are singleton patterns, whose support sequence is obtained in Step-3. To compute, support time sequence for temporal pattern, $T_m T_n$, we must compute maximum and minimum possible support sequences of temporal patterns denoted by $T_m \overline{T_n}$ and $T_n \overline{T_m}$. This is followed by computation minimum and maximum possible support sequences for temporal pattern, $T_m T_n$ (size, $|S|=2$).

Now, find the upper-lower bound, lower-lower bound, and lower bound values for temporal pattern, $T_m T_n$. If the lower bound distance value of temporal pattern, lower bound (LB) $\leq \theta$, then consider it as similar temporal association pattern, otherwise treat all such patterns as not similar temporally. However, if, approximate upper lower bound distance value of $T_m T_n \leq \theta$, then, retain all such itemsets of the form $T_m T_n$, to generate itemset support sequences of size, $|S|>2$.

Step-5: $|S|>2$, Level- K // generalized for all temporal patterns with $|S| > 2$ (i.e. $|S|=3, 4, 5, \dots, n$)

Set $|S| = |K|+1$. Generate all temporal patterns of size, $|S|>2$, from temporal association patterns of previous stage. All such temporal patterns generated shall be of the form $T_m T_n$ where T_m must be mapped to first $|S|-1$ sequence of items. Similarly, temporal pattern denoted by T_n denotes singleton temporal pattern not present in T_m . For all temporal patterns of size, $|S|>2$, we have a different and peculiar situation.

This is mainly because of two reasons.

1. For $|S|=1$, i.e for all singleton temporal patterns, we have their true support values computed through scanning the database directly. This finishes first temporal database scan.
2. For $|S|=2$, i.e for all temporal patterns of length two, we obtain maximum and minimum possible support sequences of temporal patterns. We do not have true support values.
3. So, for temporal patterns of size, $|S|>2$, such as $|S|=3, 4, 5, \dots, n$, we have four cases to be considered as shown in Eq.7, for computing support sequences of temporal patterns. Equation.6, gives expression for finding support sequence of temporal patterns of size, $|S|>2$.

In Equation.6 below, $T_m T_n$ is temporal pattern of size, $|S|>2$

$$T_m T_n = \begin{cases} T_{m_{min}} - T_{m_{min}} \otimes \overline{T_n} \\ T_{m_{max}} - T_{m_{max}} \otimes \overline{T_n} \end{cases} \quad (6)$$

From Eq.6, we have four sub cases generated as shown below using Eq.7

$$T_m T_n = \begin{cases} T_{m_{min}} - UBSTS(T_{m_{min}}, \overline{T_n}) \\ T_{m_{min}} - LBSTS(T_{m_{min}}, \overline{T_n}) \\ T_{m_{max}} - UBSTS(T_{m_{max}}, \overline{T_n}) \\ T_{m_{max}} - LBSTS(T_{m_{max}}, \overline{T_n}) \end{cases} \quad (7)$$

From these support sequences, obtain $[T_m T_n]_{\max}$ and $[T_m T_n]_{\min}$, the maximum and minimum support sequence of temporal patterns of size, $|S| > 2$ respectively. These are called maximum support time sequence and minimum support sequence of itemset, $[T_m T_n]$ of size, $|S| > 2$. Now, find the upper lower bound and lower lower bound, lower bound distance values for the pattern of the form $T_m T_n$.

If the value of lower bound distance $<$ user threshold, then consider the corresponding temporal pattern as a similar temporal association pattern, otherwise treat all such patterns as temporally dissimilar. However, if, the approximate upper lower bound distance of temporal pattern, $T_m T_n$ is less than θ , user threshold, then retain all such patterns of form $T_m T_n$, to find support time sequences of patterns of next size, $|S| = |S| + 1$. Repeat step-5 till size of temporal itemset is equal to number of items in set, denoted by I or until no further temporal patterns are generated.

Step-6: Display all temporal patterns satisfying subset constraints

Output all similar temporal association patterns w.r.t reference support sequence satisfying user specified constraints.

3.6. Algorithm Explanation

Step1: Let $T = \{t_1, t_2, t_3, \dots, t_n\}$ denote set of all time slots. For every time slot, t_j where j takes any value from 1 to n , find support values of positive and negative singleton temporal patterns. i.e of size, $|S|=1$

This step mainly includes following steps.

- i. Find probability of existence of the positive singleton pattern for every time slot. The probability is also called the support value.
- ii. Obtain probability of support values for negative singleton pattern from support values of positive items. We use the probability of pattern and support interchangeably in this paper.

Step2: Temporal Patterns with Size, $|S|=1$. Obtain support sequences, $S_{\theta_p}, S_{\theta_n}$ for temporal patterns of size, $|S|=1$, from support values of temporal patterns obtained in step-1.

Let the set I , denoting itemset consists of singleton items P, Q and R . From support values of positive items (P, Q and R) and negative singleton items (\bar{P}, \bar{Q} , and \bar{R}) computed for every time slot, obtain support sequences for items $P, Q, R, \bar{P}, \bar{Q}, \bar{R}$. Positive support time sequences obtained from items P, Q, R are denoted by $S_{\theta_p}(P), S_{\theta_p}(Q), S_{\theta_p}(R)$. Similarly, negative support sequences obtained from items \bar{P}, \bar{Q} , and \bar{R} are denoted respectively by $S_{\theta_n}(\bar{P}), S_{\theta_n}(\bar{Q}), S_{\theta_n}(\bar{R})$.

Step-3: Find level-1 similarity profiled temporal patterns.

Compute the true distance, upper-lower bound, lower-lower bound distances between the single item support sequences obtained in step-2 and reference sequence using the method adopted in (Jin Soung Yoo et.al 2012). If the true distance \leq threshold value, then this item is considered as similar temporal pattern. However if the true distance computed exceeds the threshold, and corresponding upper lower bound (ULB) distance is less than the user specified threshold, then we retain this itemset without killing, so that it may be used to compute temporal association patterns of the next stage. All such item sets which are considered for retaining in the previous stage are not treated similar temporal association patterns but are only retained for purpose of finding temporal patterns of higher levels.

Step-4: Set $k = k+1$. Find the level- k similarity profiled temporal association patterns. Here ($k=2, 3, 4, \dots, n$)

Generate all possible non-empty itemset combinations of size, $k = k+1$, using itemsets retained at level- k . For these temporal patterns of size equal to $k+1$, obtain the maximum possible support sequence (upper bound support sequence) and minimum possible support sequence (lower bound support sequence). This is followed by computation of the lower bound distance of the support sequence of itemset w.r.t the reference sequence. If the lower bound distance, satisfies the threshold constraint then the pattern is temporally similar w.r.t reference or similar temporal pattern. Otherwise, it is not treated as the similar temporal association pattern.

The algorithm or procedure to compute the support sequence is given below.

Algorithm to generate upper bound support time sequence (UBSTS), lower bound support time sequence (LBSTS) , compute lower bound (LB) distance of temporal pattern of the form $T_m T_n$ of size, $k \geq 2$
 {

1. Consider the Equation (8) to find support time sequence of temporal pattern $T_m T_n$ as given below

$$T_m T_n = \frac{1}{2} [T_m + T_n - T_m \overline{T_n} - T_n \overline{T_m}] \quad (8)$$

For each generated itemset (also called as association pattern) combination of the form $T_m T_n$ which is of size equal to k , T_m must be mapped to the level-($k-1$) itemset of length equal to ($k-1$) and T_n indicates the singleton item of length equal to 1 which is not present in T_m using Equation (8).

2. Determine the upper bound and lower bound support sequence for association pattern of the form $T_m \overline{T_n}$ using the procedure outlined in section 3.7. From these maximum and minimum possible support sequences computed for pattern, $T_m \overline{T_n}$. Obtain corresponding minimum $(T_m T_n)_{MIN}$ and maximum $(T_m T_n)_{MAX}$ support sequences for the association pattern denoted by $T_m T_n$ using Equations (9).

$$T_m T_n = \begin{cases} \frac{1}{2} * \{ (T_m)_{UBSTS} + T_n - [(T_m)_{UBSTS} * \overline{T_n}]_{UBSTS} - [(T_n)_{UBSTS} * \overline{(T_m)_{UBSTS}}]_{UBSTS} \} \\ \frac{1}{2} * \{ (T_m)_{UBSTS} + T_n - [(T_m)_{UBSTS} * \overline{T_n}]_{LBSTS} - [(T_n)_{UBSTS} * \overline{(T_m)_{UBSTS}}]_{LBSTS} \} \\ \frac{1}{2} * \{ (T_m)_{LBSTS} + T_n - [(T_m)_{LBSTS} * \overline{T_n}]_{UBSTS} - [(T_n)_{LBSTS} * \overline{(T_m)_{LBSTS}}]_{UBSTS} \} \\ \frac{1}{2} * \{ (T_m)_{LBSTS} + T_n - [(T_m)_{LBSTS} * \overline{T_n}]_{LBSTS} - [(T_n)_{LBSTS} * \overline{(T_m)_{LBSTS}}]_{LBSTS} \} \end{cases} \quad (9)$$

Here $[T_m T_n]_{MAX}$ and $[T_m T_n]_{MIN}$ denote maximum and minimum support time sequences of temporal pattern, $T_m T_n$.

3. Compute the approximate upper-lower bound and approximate lower-lower bound distance values of temporal pattern, denoted by $T_m T_n$ using maximum and minimum support sequence vectors generated in step-3 and using procedure outlined in sub sections 3.7.1, 3.7.2, 3.7.3.

4. Finally, obtain the lower bound distance by summing both upper-lower bound and lower -lower bound distances
 i.e. LB (Lower bound) = ULB (Upper lower bound) + LLB (Lower lower bound)
 }

Step-5: Repeat step-4 till itemset size is equal to number of items in set I or till no further itemsets can be generated.

For generating candidate items for level $k+1$ we use Equation 1 and Equation 2. At every stage of generating the itemset combinations as in step-4 and step-5, prune all the itemset combinations even if any subset of these item sets is not considered frequent in previous stages.

For example if XY is not frequent, then even if PR and QR are frequent, we can say PQR is not frequent. However on the fly from stage- i to stage - ($i + 1$), all those itemsets which satisfy upper-lower bound distance may be retained.

3.7. Generating UBSTS, LBSTS support time sequences and ULB, LLB and LB distances

Generating itemset support time sequences is crucial to find the similar temporal association patterns. To generate upper bound and lower bound support sequences we follow the procedure outlined in (Jin Soung Yoo et.al 2012). However the computation of support time sequence for a given itemset or pattern is done using different method using the Equations 1, 6 and 7. The earlier approach for finding support time sequence of an itemset requires

knowing the support values of all its subsets and also requires scanning the database for actual support values at previous stage to find support sequence of temporal association pattern of next stage.

In our present approach, computation of support sequence for a given temporal pattern denoted by $T_m T_n$ requires computing only the support time sequences for itemsets denoted by $T_m \overline{T_n}$ and $\overline{T_m} T_n$. This eliminates the need to maintain support of all subsets of itemset of size k. We generate the upper bound and lower bound support sequences for the itemset combination $T_m \overline{T_n}$ and $\overline{T_m} T_n$ and use these support sequences, to compute support time sequences of $T_m T_n$.

3.7.1 Computation of Upper Bound and Lower Bound Support Time Sequences

Let

$$S(I_i) = \langle S_{i1}, S_{i2}, S_{i3}, \dots, S_{im} \rangle$$

$$S(I_j) = \langle S_{j1}, S_{j2}, S_{j3}, \dots, S_{jm} \rangle \tag{10}$$

be support time sequences of items I_i and I_j . We use I_i and I_j and T_m and T_n interchangeably.

The upper bound and lower support time sequences of itemset $I_i I_j$ are computed using equations (11) and (12)

$$UBSTS(I_i I_j) = \langle \min(S_{i1}, S_{j1}), \min(S_{i2}, S_{j2}), \min(S_{i3}, S_{j3}), \dots, \min(S_{im}, S_{jm}) \rangle \tag{11}$$

$$LBSTS(I_i I_j) = \langle \max(S_{i1} + S_{j1} - 1, 0), \max(S_{i2} + S_{j2} - 1, 0), \dots, \min(S_{im} + S_{jm} - 1, 0) \rangle \tag{12}$$

3.7.2 Computation of Upper-Lower Bound distance

Let $R = \langle r_1, r_2, r_3, \dots, r_m \rangle = \langle r_i \mid i \leftarrow 1 \text{ to } m \rangle$ be a reference sequence and $U = \langle U_1, U_2, U_3, \dots, U_m \rangle$ be the upper bound support time sequence. Now, if we assume R^{Upper} and U^{Lower} to be the subsequence of reference and upper bound support time sequences of length, k, such that for all i varying from 1 to k, the condition $R_i > U_i$ holds good, then the upper-lower bound distance value is computed as $ULB\text{-distance}(R, U) = \text{distance between } R^{Upper} \text{ and } U^{Lower}$ of length k using Euclidean distance of k-dimensions.

3.7.3 Computation of Lower-Lower bound distance

Let $R = \langle r_1, r_2, r_3, \dots, r_m \rangle = \langle r_i \mid i \leftarrow 1 \text{ to } m \rangle$ be a reference support sequence and $L = \langle L_1, L_2, L_3, \dots, L_m \rangle$ be the lower bound support time sequence. Now, if we assume R^{lower} and L^{upper} to be the subsequence of reference and lower bound support time sequences of length k, such that for all i varying from 1 to k, the condition $R_i < L_i$ holds good, then the lower-lower bound distance value is computed as $LLB\text{-distance}(R, L) = \text{distance between } R^{lower} \text{ and } L^{upper}$ of length k using Euclidean distance of k-dimensions.

3.7.4 Computation of Lower Bound distance

The lower bound distance is the sum of ULB (upper lower bound) and LLB (lower lower bound) distances. Mathematically, we may define the lower bound value as given by Equation.13

$$LB \text{ (Lower-bound distance)} = ULB \text{ (Upper Lower Bound)} + LLB \text{ (Lower Lower Bound)} \tag{13}$$

4. CASE STUDY

The Table.1 denotes sample temporal database defined over time stamped transactions with finite set, I consisting items {P, Q, R}. The temporal database is partitioned into two groups of transaction sets performed at time slots, t_1 and t_2 respectively. Here we assume user specified threshold value as 0.2 and reference support sequence as $\langle 0.4, 0.6 \rangle$ as shown in Table.3. Table.2 contains values of true support time sequences for all possible itemset combinations. Since there are 3 items, there are 7 possible combinations as shown in Table.2.

Step-1:

Initially, we start by scanning temporal database to find the positive support value of singleton items P, Q, R and also corresponding negative support of items \overline{P} , \overline{Q} , \overline{R} for each time slot. The support values at time slots t_1 , and t_2 , is shown in the table 4 for each positive and negative singleton temporal pattern.

Table 1. Temporal Database

Table 2. True support sequences

Database, D1		Database, D2	
Time Slot, T1	Items	Time Slot, T2	Items
[T1]	P	[T11]	Q,R
[T2]	P,Q,R	[T12]	Q
[T3]	P,R	[T13]	P,Q,R
[T4]	P	[T14]	P,Q,R
[T5]	P,Q,R	[T15]	R
[T6]	R	[T16]	P,QR
[T7]	R	[T17]	P,R
[T8]	P,QR	[T18]	R
[T9]	R	[T19]	Q
[T10]	R	[T20]	Q,R

Table 4. Support values of singleton items at t_1, t_2

	Singleton Itemset	Support at t_1	Support at t_2
Positive Singleton item	P	0.6	0.4
	Q	0.3	0.7
	R	0.8	0.8
Negative singleton item	\bar{P}	0.4	0.6
	\bar{Q}	0.7	0.3
	\bar{R}	0.2	0.2

Table 6. ULB and actual distance of itemset w.r.t Ref

Pattern	PSTS, S_{θ_p}	ULB-approx	actual distance
P	<0.6,0.4>	0.2✓	0.28 *
Q	<0.3,0.7>	0.1✓	0.14 ✓
R	<0.8,0.8>	0.0✓	0.45 *

Table 8. Detailed LB distance of pattern [PQ]

Patterns	STS	Upper-lower & Lower-lower bounds w.r.t reference		
		ULB	LLB	LB
P	<0.6,0.4>			
Q	<0.7,0.3>			
$P\bar{Q}_{UBSTS}$	<0.6,0.3>			
$P\bar{Q}_{LBSTS}$	<0.3,0.0>			
PQ_{MIN}	<0.0,0.1>			
PQ_{MAX}	<0.3,0.4>			
Ref	<0.4,0.6>			
		0.2236	0.0	0.2236
Similarity of temporal pattern [PQ] w.r.t Ref				✗
Move to Level-3				✗

Table 9. Lower Bound distance computation of PR

Pattern	STS	ULB	LLB	LB
PR_{max}	<0.6,0.4>			
PR_{min}	<0.4,0.2>			

S.No	Pattern	Pattern Support Sequence
1	P	<0.6,0.4>
2	Q	<0.3,0.7>
3	R	<0.8,0.8>
4	PQ	<0.3,0.3>
5	PR	<0.4,0.4>
6	QR	<0.3,0.5>
7	PQR	<0.3,0.3>

Table3. Reference Sequence

Reference	Support Time Sequence
Ref	<0.4,0.6>

Table 5. Positive & Negative Support Time Sequences

Positive Itemset	PSTS, S_{θ_p}	Negative Itemset	NSTS, S_{θ_N}
P	<0.6,0.4>	\bar{P}	<0.4,0.6>
Q	<0.3,0.7>	\bar{Q}	<0.7,0.3>
R	<0.8,0.8>	\bar{R}	<0.2,0.2>

Table 7. Lower Bound distance computation of XY

Pattern	STS	ULB	LLB	LB
PQ_{max}	<0.3,0.4>			
PQ_{min}	<0.0,0.1>			
Ref	<0.4,0.6>			
		0.2236*	0	0.2236*
Actual distance w.r.t reference STS = 0.32				

Table 10. Detailed LB distance of itemset [PR]

Patterns	STS	Upper-lower & Lower-lower bounds w.r.t reference		
		ULB	LLB	LB
P	<0.6,0.4>			
\bar{R}	<0.2,0.2>			
$P\bar{R}_{UBSTS}$	<0.2,0.2>			
$P\bar{R}_{LBSTS}$	<0.0,0.0>			
PR_{MIN}	<0.4,0.2>			
PR_{MAX}	<0.6,0.4>			
Ref	<0.4,0.6>			
		0.2	0.0	0.2
Similarity of temporal pattern [PR] w.r.t Ref				✓
Move to Level-3				✓

Table 12. Detailed LB distance of itemset [QR]

Ref	<0.4,0.6>			
		0.2✓	0	0.2✓
Actual distance w.r.t reference STS = 0. 2				

Table 11. Lower Bound distance computation of QR

Pattern	STS	ULB	LLB	LB
QR _{max}	<0.3,0.7>			
QR _{min}	<0.1,0.5>			
Ref	<0.4,0.6>			
		0.1✓	0	0.1✓
Actual distance w.r.t reference STS = 0. 1414				

		Upper-lower & Lower-lower bounds w.r.t reference		
Pattern	STS	ULB	LLB	LB
Q	<0.3,0.7>			
\bar{R}	<0.2,0.2>			
QR _{UBSTS}	<0.2,0.2>			
QR _{LBSTS}	<0.0,0.0>			
QR _{MIN}	<0.1,0.5>			
QR _{MAX}	<0.3,0.7>			
Ref	<0.4,0.6>			
		0.1	0.0	0.1
Similarity of temporal pattern [QR] w.r.t Ref				✓
Move to Level-3				✓

Step2:

Compute Positive and Negative support time sequences ($S_{\theta_p}, S_{\theta_n}$) of singleton temporal items from support values of corresponding positive and negative items obtained in step-1. We can obtain the positive and negative support sequences of singleton items as shown in Table.5 for time slots t_1 and t_2 from Table.4.

Step-3:

Find Level -1 similarity profiled temporal items. In this step, we compute both the approximate upper lower bound (ULB) and actual Euclidean distance between reference vector and singleton items. If the actual distance is less than the user threshold, this means that the singleton item is temporally similar. Otherwise it is temporally dissimilar pattern. We also find the approximate upper lower bound distance value which is the deciding factor to consider or discard the corresponding singleton item. In the present example, even though true distances of P and R do not satisfy the threshold constraint, we consider retaining these singleton items, as these may be helpful to generate level-2 itemsets. We retain all such items whose upper-lower bound distance satisfies the threshold constraint as shown in Table.6.

Step-4: Find Temporal Patterns of Size, |S|=2

This step involves generating support sequence for item sets of length, |S|=2, as denoted by [PQ], [PR] and [QR]. From these support sequences generated, we must find whether itemsets satisfy threshold constraint. This must be done without scanning the database. We now show computations of these itemsets.

A. Finding Support Time Sequence for temporal pattern [PQ]

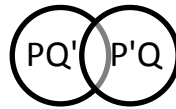


Figure.1 Computation of Support sequence for PQ using Venn diagram

The figure.1 shows the Venn diagram for pattern [PQ]. Here P' and Q' indicates complementary pattern. We have expression to find support sequence as $T_m T_n = \frac{1}{2} [T_m + T_n - T_m \bar{T}_n - T_n \bar{T}_m]$ where $T_m = P$ and $T_n = Q$. Now, using procedure to find upper and lower bound support sequences as discussed in section 3.7, we have

$$\bar{P}\bar{Q}_{UBSTS} = < \min(0.6, 0.7), \min(0.4, 0.3) > = < 0.6, 0.3 >$$

$$\bar{P}\bar{Q}_{LBSTS} = < \max(0.6 + 0.7 - 1, 0), \max(0.4 + 0.3 - 1, 0) > = < 0.3, 0.0 >$$

$$\bar{P}Q_{UBSTS} = < \min(0.4, 0.3), \min(0.6, 0.7) > = < 0.3, 0.6 >$$

$$\bar{P}Q_{LBSTS} = < \max(0.6 + 0.7 - 1, 0), \max(0.4 + 0.3 - 1, 0) > = < 0.0, 0.3 >$$

Further using the equation (1) and (6), we have

$$PQ = \begin{cases} = \frac{1}{2} [< 0.6, 0.4 > + < 0.3, 0.7 > - < 0.6, 0.3 > - < 0.3, 0.6 >] = < 0.0, 0.1 > = PQ_{\text{MIN}} \\ = \frac{1}{2} [< 0.6, 0.4 > + < 0.3, 0.7 > - < 0.3, 0.0 > - < 0.0, 0.3 >] = < 0.3, 0.4 > = PQ_{\text{MAX}} \end{cases}$$

From these minimum and maximum support sequences obtained for the temporal pattern $[P\bar{Q}]$ and $[\bar{P}Q]$, we can deduce maximum and minimum possible support sequences for temporal pattern $[PQ]$ defined as, $PQ_{\text{min}} = < 0.0, 0.1 >$ and $PQ_{\text{max}} = < 0.3, 0.4 >$. Table.7 gives computation of lower bound distance of temporal pattern $[PQ]$. Since, upper lower bound and lower bound distances of support sequence of temporal itemset $[PQ]$ w.r.t reference support sequence, R do not satisfy user defined threshold value constraint; we consider the temporal itemset $[PQ]$ as temporally dissimilar pattern w.r.t *Ref*. The Table.8 shows detailed computation process of support sequence, support values for temporal pattern $[PQ]$.

B. Finding Support Time Sequence for temporal pattern $[PR]$

The Figure .2 shows the Venn diagram for pattern $[PR]$. Here P' and R' indicates complementary pattern.

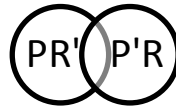


Figure.2 Computation of Support sequence for PR using Venn diagram

Using the expression defined in Eq.(1) to compute temporal support sequence to estimate temporal supports of patterns defined as $T_m T_n = \frac{1}{2} [T_m + T_n - T_m \bar{T}_n - T_n \bar{T}_m]$, where $T_m = P$ and $T_n = R$. Now, using procedure to find upper and lower bound support sequences as discussed in section 3.7, we have

$$\bar{P}\bar{R}_{\text{UBSTS}} = < \min(0.6, 0.2), \min(0.4, 0.2) > = < 0.2, 0.2 >$$

$$\bar{P}\bar{R}_{\text{LBSTS}} = < \max(0.6 + 0.2 - 1, 0), \max(0.6 + 0.2 - 1, 0) > = < 0.0, 0.0 >$$

$$\bar{P}R_{\text{UBSTS}} = < \min(0.4, 0.8), \min(0.6, 0.8) > = < 0.4, 0.6 >$$

$$\bar{P}R_{\text{LBSTS}} = < \max(0.4 + 0.8 - 1, 0), \max(0.6 + 0.8 - 1, 0) > = < 0.2, 0.4 >$$

Further using the equation (1) and (6), we have

$$PR = \begin{cases} = \frac{1}{2} [< 0.6, 0.4 > + < 0.8, 0.8 > - < 0.2, 0.2 > - < 0.4, 0.6 >] = < 0.4, 0.2 > = PQ_{\text{MIN}} \\ = \frac{1}{2} [< 0.6, 0.4 > + < 0.8, 0.8 > - < 0.0, 0.0 > - < 0.2, 0.4 >] = < 0.6, 0.4 > = PQ_{\text{MAX}} \end{cases}$$

From these lower bound and upper bound support sequences obtained for the temporal pattern $[P\bar{R}]$, we can deduce maximum and minimum possible support sequences for temporal pattern $[PR]$ defined as, $PR_{\text{min}} = < 0.4, 0.2 >$ and $PR_{\text{max}} = < 0.6, 0.4 >$. Table.9 gives computation of lower bound distance of temporal pattern $[PR]$. Since, the upper lower bound and lower bound distances of support sequence of temporal itemset $[PR]$ w.r.t reference support sequence, R satisfies user defined threshold value constraint; we consider the temporal itemset $[PR]$ as temporally similar pattern w.r.t *Ref*. The table 10 shows the detailed computation process of support sequence, support values for temporal pattern $[PR]$.

C. Finding Support Time Sequence for temporal pattern $[QR]$

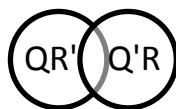


Figure.3 Computation of Support sequence for QR using Venn diagram

The figure.3 shows the Venn diagram for pattern [QR]. Here \bar{Q} and \bar{R} indicates complementary pattern. We have expression to find support sequence as $T_m T_n = \frac{1}{2}[T_m + T_n - T_m \bar{T}_n - T_n \bar{T}_m]$ where $T_m = Q$ and $T_n = R$. Now, using procedure to find upper and lower bound support sequences as discussed in section 3.7, we have

$$Q\bar{R}_{UBSTS} = \langle \min(0.3, 0.2), \min(0.7, 0.2) \rangle = \langle 0.2, 0.2 \rangle$$

$$Q\bar{R}_{LBSTS} = \langle \max(0.3 + 0.2 - 1, 0), \max(0.7 + 0.2 - 1, 0) \rangle = \langle 0.0, 0.0 \rangle$$

$$\bar{Q}R_{UBSTS} = \langle \min(0.7, 0.8), \min(0.3, 0.8) \rangle = \langle 0.7, 0.3 \rangle$$

$$\bar{Q}R_{LBSTS} = \langle \max(0.7 + 0.8 - 1, 0), \max(0.8 + 0.3 - 1, 0) \rangle = \langle 0.5, 0.1 \rangle$$

Further using the equation (1) and (6), we have

$$QR = \begin{cases} = \frac{1}{2}[\langle 0.3, 0.7 \rangle + \langle 0.8, 0.8 \rangle - \langle 0.2, 0.2 \rangle - \langle 0.7, 0.3 \rangle] = \langle 0.1, 0.5 \rangle = QR_{MIN} \\ = \frac{1}{2}[\langle 0.3, 0.7 \rangle + \langle 0.8, 0.8 \rangle - \langle 0.0, 0.0 \rangle - \langle 0.5, 0.1 \rangle] = \langle 0.3, 0.7 \rangle = QR_{MAX} \end{cases}$$

From these lower bound and upper bound support sequences obtained for the temporal pattern [QR], we can deduce maximum and minimum possible support sequences for temporal pattern [QR] defined as, $QR_{min} = \langle 0.1, 0.5 \rangle$ and $QR_{max} = \langle 0.3, 0.7 \rangle$. Table.11 gives computation of lower bound distance of temporal pattern [QR]. Since, the upper lower bound and lower bound distances of support sequence of temporal itemset [QR] w.r.t reference support sequence, R satisfies user defined threshold value constraint; we consider the temporal itemset [QR] as temporally similar pattern w.r.t *Ref*. Table.12 shows detailed process of computation of support sequence, support values for the temporal association pattern [QR].

Step-5: Generate Temporal Pattern of Size, |S|=3, PQR

Consider upper lower bound distance and lower bound distance of support sequences of association patterns [PQ], [PR] and [QR] w.r.t reference sequence, *Ref*. Here \checkmark and \times indicates satisfying and not satisfying the threshold constraint respectively. Since, only [PR] and [QR] satisfy threshold constraint, and [PQ] is not similar w.r.t the reference sequence, we consider association pattern [PQR] to be temporally dissimilar. This is because subsets of the pattern [PQR] which are [PQ], [PR], and [QR] must also be temporally similar which is not true w.r.t pattern [PQ]. Thus, final set of temporal association patterns which are similar to reference are [Q], [PR] and [QR]. Alternately, we may compute lower bound distance of [PQR] by using the Equation.1 and verify, if the association pattern [PQR] is temporally similar or not. This process is explained below.

Here $I_i = PQ$ and $I_j = R$. We may obtain support sequence for temporal pattern [PQR] as explained below

We give equation.1 again as shown below,

$$I_i I_j = \begin{cases} I_{i_{min}} - I_{i_{min}} \otimes \bar{I}_j \\ I_{i_{max}} - I_{i_{max}} \otimes \bar{I}_j \end{cases}$$

Here $I_i = PQ$ and $I_j = R$. We can find support sequence for $I_i I_j$ using Equation.1 as shown below

$$PQR = \begin{cases} PQ_{min} - PQ_{min} \otimes \bar{I}_j \\ PQ_{max} - PQ_{max} \otimes \bar{I}_j \end{cases}$$

We map $I_{i_{max}}$, $I_{i_{min}}$ and \bar{I}_j to PQ_{MAX} , PQ_{MIN} and \bar{R} respectively as shown below

$$\begin{aligned} I_{i_{max}} &= PQ_{MAX} = \langle 0.3, 0.4 \rangle \\ I_{i_{min}} &= PQ_{MIN} = \langle 0.0, 0.1 \rangle \\ \bar{I}_j &= \bar{R} = \langle 0.2, 0.2 \rangle \end{aligned}$$

So, the support sequence for [PQR] may be obtained using computations as shown below

$$\begin{aligned} [PQR] &= \\ &= \begin{cases} \langle 0.0, 0.1 \rangle - \langle 0.0, 0.1 \rangle \otimes \langle 0.2, 0.2 \rangle \\ \langle 0.3, 0.4 \rangle - \langle 0.3, 0.4 \rangle \otimes \langle 0.2, 0.2 \rangle \end{cases} \\ &= \begin{cases} \langle 0.0, 0.1 \rangle - \langle 0.0, 0.1 \rangle \\ \langle 0.0, 0.1 \rangle - \langle 0.0, 0.0 \rangle \\ \langle 0.3, 0.4 \rangle - \langle 0.2, 0.2 \rangle \\ \langle 0.3, 0.4 \rangle - \langle 0.0, 0.0 \rangle \end{cases} \\ &= \begin{cases} \langle 0.0, 0.0 \rangle \\ \langle 0.0, 0.1 \rangle \\ \langle 0.1, 0.2 \rangle \\ \langle 0.3, 0.4 \rangle \end{cases} \end{aligned}$$

From these support sequences obtained, we can obtain maximum and minimum possible support sequences for PQR as

$$\begin{aligned} [PQR]_{max} &= \langle 0.3, 0.4 \rangle \text{ and} \\ [PQR]_{min} &= \langle 0.0, 0.0 \rangle \end{aligned}$$

Table 13. Lower Bound distance computation of PQR

Itemset	STS	ULB	LLB	LB
PQR_{max}	$\langle 0.3, 0.4 \rangle$			
PQR_{min}	$\langle 0.0, 0.0 \rangle$			
Ref	$\langle 0.4, 0.6 \rangle$			
		0.2236*	0	0.2236*
Actual distance w.r.t reference STS = 0. 3162 *				

From Table.13, we can see that the lower bound distance is 0.2236 and exceeds threshold value. Hence, the pattern [PQR] is not temporally similar w.r.t reference, **Ref**.

5. CONCLUSIONS

In this research, given a reference support time sequence defined for a finite number of time slots, we aim to discover all such temporal association patterns from temporal database which are similar to the reference vector. We introduce the concept of negative itemset support values and negative itemset support sequences. In essence, we maintain two types of supports called positive and negative supports. In addition, we also compute two types of support sequences namely positive and negative support sequences. The idea is to use the concept of Venn diagrams, to find similar temporal association patterns satisfying given threshold value, with primary objective of eliminating multiple temporal database scans. This is achieved by formalizing the expressions required to estimate support sequences and through the reuse of support sequences of temporal sub patterns computed in previous stages. The proposed approach thus reduces the space and time complexities. In future, there is a scope for extending present approach by designing a novel similarity measures to find similar temporal patterns and we are in this process.

ACKNOWLEDGEMENTS

The authors are thankful to all anonymous reviewers for their valuable suggestions and remarks. We are also thankful for the editor, Miguel Sanchez for his cooperation and allowing us to extend the present work. This research would not have been possible without the inspiration from professors, Dr.P.V.Kumar and Dr.V.Janaki.

REFERENCES

- Alexander Hinneburg, Dirk Habich, and Wolfgang Lehner, "COMBI-operator - database support for data mining applications", Proceedings of the 29th international conference on Very large data bases – Vol. 29, 2003, pp.429-439.
- Artur Bykowski, Jouni K.Seppanen, Jaakko Hollomen, "Model Independent bounding of the supports of Boolean formulae in binary data", Database Support for Data Mining Applications Lecture Notes in Computer Science Vol. 2682, 2004, pp. 234-249.
- Chang-Hung Lee, Cheng-Ru Lin, Ming-Syan Chen, "On mining general temporal association rules in a publication databases", Proceedings of IEEE ICDM 2001, pp.337-344.
- Cheqing Jin et.al, "Mining Frequent Items in Spatiotemporal Databases", Advances in Web-Age Information Management Lecture Notes in Computer Science Vol.31, Issue 29, 2004, pp. 549-558.
- Gultekin Ozsoyolu and Richard T. Snodgrass, "Temporal and Real-Time Databases: A Survey". IEEE Transactions on Knowledge and Data Engineering, Vol. 7, Issue 4, 1995, pp. 513-532.
- Jin Soung Yoo, Shashi Sekhar, "Mining Temporal association patterns under a similarity constraint", Scientific and Statistical Database Management Lecture Notes in Computer Science, Vol. 5069, 2008, pp. 401-417.
- Jin Soung Yoo, "Temporal Data Mining: Similarity profiled association pattern", Data Mining Found & Intel paradigms, Vol. 23, 2012, pp. 29-47.
- Jin Soung Yoo, Shashi Sekhar, "Similarity-Profiled Temporal Association Mining" , IEEE Transactions on Knowledge and Data Engineering, Vol. 21, No.8, 2009, pp.1147-1161.
- Long Jin, Yongmi Lee, Sungbo Seo, Keun Ho Ryu, "Discovery of Temporal Frequent Patterns Using TFP Tree", Advances in Web-Age Information Management Lecture Notes in Computer Science, Vol. 40, issue 16, 2006, pp 349-361
- Matteo Golfarelli, Stefano Rizzi, "A Survey on Temporal datawarehousing", International Journal of DataWarehousing and Mining, Vol. 5, Issue 1, 2009, pp.1-7.
- Panagiotis Papapetrou, George Kollios, Stan Sclaroff, Dimitrios Gunopulos, "Mining frequent arrangements of temporal intervals. Knowledge and Information Systems", Vol. 21, Issue 2, 2009, pp 133-171.
- Srivatsan Laxman, P S Sastry, "A survey of temporal data mining", Sadhana Publications, Vol. 31, Part 2, 2006, pp. 173-198
- Tansel Clifford, Gadia JaJodia, Segeu Snodgaass, "Temporal Databases. Theory, Design and Implementation", Benjamin Cummings Publishing, 1993
- Tansel A.U, Imberman S.P, "Discovery of Association Rules in Temporal Databases," in Information Technology Proceedings of ITNG '07, Vol.1, 2007, pp.371-376
- Toon Calders, Jan Paredaens, "Axiomatization of frequent items", Theoretical Computer Science. Vol. 290, Issue 1, 2003, pp. 669-693
- Toon Calders, "Deducing Bounds on support of Itemsets", Database Support for Data Mining applications, Vol.2682, 2004, pp.214-233.
- Vangipuram Radhakrishna, P. V. Kumar, and V. Janaki. 2015. A Survey on Temporal Databases and Data mining. In Proceedings of the The International Conference on Engineering & MIS 2015(ICEMIS '15). ACM, New York, NY, USA, Article 52, 6 pages. DOI=http://dx.doi.org/10.1145/2832987.2833064
- Vangipuram Radhakrishna, P. V. Kumar, and V. Janaki, "A Novel Approach for Mining Similarity Profiled Temporal Association Patterns Using Venn Diagrams", In Proceedings of the The International Conference on Engineering & MIS 2015 (ICEMIS '15). ACM, New York, NY, USA, Article 58, 9 pages. DOI=http://dx.doi.org/10.1145/2832987.2833071
- Vangipuram Radhakrishna, P. V. Kumar, and V. Janaki. 2015. An Approach for Mining Similarity Profiled Temporal Association Patterns Using Gaussian Based Dissimilarity Measure. In Proceedings of the The International Conference on Engineering & MIS 2015 (ICEMIS '15). ACM, New York, NY, USA, Article 57 , 6 pages. DOI=http://dx.doi.org/10.1145/2832987.2833069
- Vangipuram Radhakrishna, P.V.Kumar, V.Janaki, " A Novel Approach for Mining Similarity Profiled Temporal Association Patterns" Rev. Téc. Ing. Univ. Zulia. Vol. 38, Nº 3, 80 - 93, 2015.



V. Radhakrishna is presently associated with Department of Information Technology, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, INDIA. He is a Professional Member of IEEE (MemberID-91086459), IEEE Technical Council on Data Engineering, IEEE Computer Society Technical Committee on Software Engineering and Association of Computing Machinery (ACM, Member.ID-6967456). His research interests include Temporal Databases, Temporal Data Mining, and Algorithm Design. He is certified SQL associate from Cambridge intercontinental university. He received several best paper awards at International Conferences within and abroad and has been an active researcher under the footsteps of professor P.V.Kumar who has been major inspiration for all the present and future achievements. His Scopus Author ID is 56118344300 and Thomson Reuters Researcher ID is I-5990-2014.



Dr. P.V. Kumar, Retired Professor of CSE in Osmania University Hyderabad, obtained M.Tech CSE from Osmania University and awarded Ph.D in Computer Science and Engineering from Osmania University. He has over 30 years of Teaching and R&D experience. A number of research scholars are working under his esteemed guidance towards their Ph.D. He has to his credit around 56 papers in various fields of Engineering, in National and Peer Reviewed International Journals indexed by Scopus, Science Citation Indexed and ISI Thomson Reuters, Web of Science. He presented papers at several National and International conferences, He worked as Chairman BOS in Osmania University College of Engineering and conducted various staff development programs and workshops. He is Life Member of MISTE, Life Member of CSI. His interested area is temporal databases, Bio Informatics, Data mining and Artificial Intelligence. Several Research scholars from various universities are pursuing research in his guidance and also awarded successfully.



Dr. V. Janaki received Ph.D degree from J.N.T. University Hyderabad, India in 2009 and M.Tech degree from R.E.C Warangal, Andhra Pradesh, India in 1988. She is currently working as Head and Professor of CSE, Vaagdevi Engineering College, Warangal, India. She has been awarded Ph.D for her research work done on Hill Cipher. Her main research interest includes Network security, Mobile Adhoc Networks and Artificial Intelligence. She has been involved in the organization as a chief member for various conferences and workshops. She published more than 50 research papers in National and International journals and conferences. She is presently supervising nearly 10 scholars for their research.