IADIS International Journal on Computer Science and Information Systems
Vol. 12, No. 1, pp. 126-142
ISSN: 1646-3692

A DISSIMILARITY MEASURE FOR MINING SIMILAR TEMPORAL ASSOCIATION PATTERNS

Vangipuram Radhakrishna[1], P. V. Kumar[2], V. Janaki[3] and Aravind Cheruvu[1]
[1]*Department of Information Technology, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India -500096*
[2]*Professor(Retd), Computer Science and Engg Department, University College of Engineering, Osmania University, Hyderabad, India*
[3]*Professor, Computer Science and Engg Department, Vaagdevi College of Engineering, Warangal, India*

ABSTRACT

This research address the design of a new dissimilarity measure and applying it to find all valid similarity profiled patterns in a temporal database defined over finite number of time slots. The proposed dissimilarity measure is a function of the reference sequence, threshold and standard deviation. Given, a reference time sequence and allowable dissimilarity limit, unearthing all eccentric (similar) temporal association patterns requires a similarity or correlation measure that can estimate similar association patterns accurately, efficiently, and is computationally optimal. This research also proposes a method to estimate temporal pattern support bounds. The experiment result shows the advantage of our proposed measure and bound estimation approach and also proves that our method is computationally efficient when compared to naïve, sequential and Spamine approaches.

KEYWORDS

Temporal, Dissimilarity, Association Pattern, Outliers, Time Stamp

1. INTRODUCTION

Temporal data mining is defined as a "single step knowledge discovery process that enumerates temporal models or patterns over temporal data in temporal databases. An algorithm which discovers temporal patterns from temporal data or fits models to temporal data is called as temporal data mining algorithm. Some of the important temporal data mining techniques are temporal classification, temporal clustering and induction. Rule induction and decision trees are two approaches in data mining that fall under induction (Weiqiang Lin, Mehmet A. Orgun, & Graham J. Williams, 2015). Other common techniques in temporal data

mining are handling time series temporal pattern discovery, sequential, sequence and event data, temporal data search, temporal data retrieval and segmentation of time series and temporal data (Erich Fuchs, Thiemo Gruber, Helmuth Pree, & Bernhard Sick, 2010). Temporal data mining has practical applications in predicting next stage medical drugs, weather forecasting, finance analysis, and biomedical temporal analysis. Applications involving time series data (Weiqiang Lin et al., 2015) implicitly require similarity metrics to predict and analyze temporal trends. "Sub space learning" is one of the important learning mechanisms which are essentially related and concerned to dimensionality reduction aiming at upholding the statistical information. In Erich Fuchs et al. 2010, a subspace representation termed as "polynomial shape space representation" is proposed to handle time series (temporal) data by considering univariate time series. Polynomial shape space representation allows the use of similarity measures to time series which is not possible otherwise. Machine learning technique based algorithms fail in many situations because of the inability to generalize correctly or due to over fitting. The "Polynomial shape space representation ", suggested helps to overcome this dis-advantage when handling time series data. It also facilitates to find similarity between two time series data adopting a distance metric (either existing or any proposed distance measure). A distance metric named as "shape space distance measure" is proposed in their work to find similarity among two time series data.

Motif detection (Erich Fuchs, Thiemo Gruber, Helmuth Pree, Bernhard Sick, 2009) is a temporal data mining task which requires time series similarity measure. An approach named as "Swift Motif" is addressed in Erich Fuchs et al. (2009) for the online discovery of motif patterns in time series data. The segmentation of time series data is achieved using data stream segmentation technique. This approach is extended, to detect outlier and frequent temporal patterns.

Temporal data mining applications in biomedical and bioinformatics fields also received wide importance. This is because the medical data consists of readings that involve clinical readings, biological values each recorded at various time points that are temporal in nature. Analyzing such temporal data requires applying temporal data mining techniques.

Functional genomics (Mario Stefanelli, 2016) is also one of the several areas that are receiving practical significance. The consequent and antecedent of traditional ARs (association rules) have no temporal relation among them. In contrast to these traditional association rules, temporal association rules contain temporal ordering and such an ordering is represented using "cause-effect" relationship. This property of temporal rules gains importance from genomics perspective. Maria Carla Calzarossa & Daniele Tessera (2015), applies temporal data mining to web page contents and represents all information related to webpage content changes as a "periodic time series". Contents of webpage undergo updating regularly such as additions of new content or deletion of the existing content. This changing web content, web access, webpage interaction by users, affects the temporal change patterns.

Temporal Association Rules (TARs) are generated for the web usage data (Stephen G. Matthews, Mario A. Gongora, Adrian A. Hopgood, & Samad Ahmadi, 2013) using genetic algorithms. Conventional fuzzy TARs do not consider the membership degree of association rules. In several instances, association rules are generated at a relatively less membership degree area. Such rules usually loose when they are not handled correctly. One approach to address this issue is addressed by generating FTARs using genetic algorithm approach (Stephen et al., 2013). Each transaction carried is generally recorded in a database. Itemsets of transactions may be qualitative or quantitative. Items which are quantitative in nature are associated with a life span w.r.t a given temporal base. Itemset life spans (Chun-Hao Chen,

Guo-Cheng Lan, Tzung-Pei Hong, Shih-Bin Lin, 2016) are found and fuzzy TARs are generated. In their work, the itemset lifespan are considered and a temporal information table (TIT) is generated through transformation process. Each quantitative value of temporal itemset is transformed to an equivalent fuzzy value using fuzzy membership function. A new approach for discovering evolutionary behavior of objects is discussed in Manish Gupta, Jing Gao, Yizhou Sun, & Jaiwen Han (2012). These objects are termed as "community trend outliers". Temporal data may be distributed data which includes temporal distributed data, spatial sensor data, time series data, stream data, network data, and spatio-temporal data.

A recent survey carried out by M. Gupta, J. Gao, C. C. Aggarwal, & J. Han (2014) discuss extensively on "mining outliers in temporal data". The authors address specific challenges for detecting outliers from temporal data, temporal outlier analysis classification and prediction models to outline some of the main contributions. They also discuss a framework for "community outlier detection" in M. Gupta et al. (2014). When performing analysis, infrequent itemsets may also provide key insights into the datasets, contrary to that discovered by deriving frequent itemsets. This fact is the basis for the approach carried out in W. Ouyang, S. Luo, & Q. Huang (2007). In this, authors propose an approach for discovering both direct association patterns and indirect association patterns. The limitation is that the generated association rules are actually based on using only single support values. The work in W. Ouyang et al. (2007) is extended by considering the use of multiple support values instead of single support values. In Weimin Ouyang & Qinhua Huang (2010), problem of rare item is discussed and association rules are generated by considering multiple minimum support values for discovering both direct and indirect association rules. In J. S. Yoo & S. Shekhar (2009); Yoo JS (2012), consider the temporal data (time stamped), to retrieve all TAPs (temporal association patterns) using Euclidean distance.

Some closest related research motivated from J. S. Yoo et al. (2009) includes Vangipuram Radhakrishna, P. V. Kumar, & V. Janaki (2015) and subsequently Vangipuram Radhakrishna et al.(2015); S. Aljawarneh, V. Radhakrishna, P. V. Kumar and V. Janaki (2016); S.Aljawarneh (2017). An approach for discovering and retrieval of routine tasks (Oliver Brdiczka, Norman Makoto Su, & James Begole, 2010) also termed as the "temporal task foot printing" is addressed. For this temporal irregularities are considered. Vangipuram Radhakrishna et al. (2016) proposed a dissimilarity measure for predicting temporally similar patterns extending (J. S. Yoo et al., 2009). Deep learning and data mining techniques (Saaed Mehrabi et al., 2015) are applied to generate temporal association rules (TARs).

Vangipuram Radhakrishna, Shadi A. Aljawarneh, Puligadda Veereswara Kumar, & Kim-Kwang Raymond Choo (2016) come up with a novel dissimilarity measure holding monotonicity. An approach for estimating supports is proposed in Aravind Cheruvu & V. Radhakrishna (2016). Abdelsalam M. Maatuk, M. Akhtar Ali, & Shadi Aljawarneh (2015) proposes a solution to generate XML Schema from relation database. Negin Keivani, Abdelsalam M Maatuk, Shadi Aljawarneh, & Muhammad Akhtar Ali (2015) reviews the promises of object relational databases and unification of OR databases with relational technology. Muneer Bani Yassein, Shadi Aljawarneh, Esra'a Masa'deh (2017) propose an elastic trickle algorithm for IoT. Shadi A. Aljawarneh, Mohammed R. Elkobaisi, & Abdelsalam M. Maatuk (2016) and Mohammed R Elkobaisi et al. (2015) address recognizing research trends in wearable systems. A solution for translating relational database schema to object based schemas in Abdelsalam M Maatuk, Muhammad A Ali, & Shadi Aljawarneh (2015). Shadi A. Aljawarneh, Raja A. Moftah, & Abdelsalam M. Maatuk (2016) address an approach for worm signature detection.

The paper is outlined as follows. The proposed measure is discussed in section-2 and its formal proof in section-3. Section-4 explains bound computations and algorithm is outlined in section-5. Results are recorded in section-6 and section-7 concludes this paper.

2. TEMPORAL DISSIMILARITY MEASURE

Previous works (J. S. Yoo, 2009) addressing similarity profiled mining used Euclidean measure. In our earlier research, we addressed the problem of "mining similarity profiled temporal patterns" by proposing a novel distance measure (Vangipuram et al., 2016) applying principles of soft computing. This work extends our previous research by introducing a novel dissimilarity measure for time stamped temporal databases. The basic terms and terminology (V. Radhakrishna, 2015 and Vangipuram, 2016) is followed in this paper. The motivation is from J. S. Yoo (2009) that addressed mining similarity patterns in temporal databases (time stamped). The present temporal dissimilarity measure is inspired from Jung-Yi Jiang, Ren-Jia Liou, & Shie-Jue Lee (2011); Y. S. Lin, J. Yi. Jiang, & S. J. Lee (2014). The membership function is a function of support value of temporal and reference patterns at a given time slot and deviation in fuzzy space. The dissimilarity measure is defined as a function of average membership value.

2.1 Proposed Measure

Let, T_s and R_s be the temporal and reference pattern. T_{s_m} and R_{s_m} denote support of patterns at m^{th} timeslot. $\overrightarrow{T_s} = \left(T_{s_1}\,;\, T_{s_2}\,;\, T_{s_3}\,\ldots\ldots\ldots;\, T_{s_m}\right)$ is the support time sequence of T_s and $\overrightarrow{R_s} = \left(R_{s_1}\,;\, R_{s_2}\,;\, R_{s_3}\,\ldots\ldots\ldots;\, R_{s_m}\right)$ is the reference time sequence. The membership degree, $M_k^{T,R}$ of temporal pattern, T_s w. r. t R_s for k^{th} timeslot is given by equation (1)

$$M_k^{T_s,\ R_s} = 0.5 * \left[1 + e^{-\left[\frac{T_{s_k} - R_{s_k}}{\sigma_g}\right]^2}\right] \tag{1}$$

The above equation gives membership for one time slot. Extending this to patterns defined over 'm' time slots, the membership degree is given by equation (2)

$$M_{avg} = \frac{\sum_{k=1}^{k=m} M_k^{T_s,\ R_s}}{|k|} \tag{2}$$

Equation (3) gives the formal expression for true dissimilarity, D_{T_s,R_s}^{true} between T_s and R_s as a function of average dissimilarity, M_{avg}

$$D_{T_s,R_s}^{true} = \frac{(1 - M_{avg})}{0.5} \tag{3}$$

In equations (2) and (3), M_{avg} is the average similarity for 'm' time slots.

2.2 Threshold (δ^g) and Deviation (σ_g)

The expressions for computing threshold and deviation are given by equations (4) and (5). In both these equations the notation,'δ' is the user threshold specified in Euclidean space.

$$\sigma_g = \frac{\delta}{\sqrt{log_e\left(\frac{1}{1-\delta}\right)}} \tag{4}$$

The expression for δ^g (transformed threshold) is a function of two variables ('δ' and σ_g) whereas σ_g is a function of only 'δ'.

$$\delta^g = 1 - e^{-\left[\frac{\delta}{\sigma_g}\right]^2} \tag{5}$$

3. FORMAL PROOF

3.1 Dissimilarity Function

Consider the expression for membership function denoted by equation (6)

$$M_k^{T_s,\ R_s} = 0.5 * \left[1 + e^{-\left[\frac{T_{s_k} - R_{s_k}}{\sigma_g}\right]^2}\right] \tag{6}$$

We define the true dissimilarity using equation (7).

$$D_{T_s,R_s}^{true} = \frac{1-M_{avg}}{1+Q} = \mathrm{D}(T_s\ ,\ R_s) \tag{7}$$

For best case, $D_{T_s,R_s}^{true} = 0$. Equation (7), reduces to

$$\frac{1-M_{avg}}{1+Q} = 0 \tag{8}$$

Solving equation (8), we get M_{avg} as unity. This is represented using Equation (9).

$$M_{avg} = 1 \tag{9}$$

Similarly, $D_{T_s,R_s}^{true} = 1$ in the worst case. This gives

$$\frac{1-M_{avg}}{1+Q} = 1 \tag{10}$$

From (10), we have

$$Q = -M_{avg} = -0.5 * \left[1 + e^{-\left[\frac{T_{s_k} - R_{s_k}}{\sigma_g}\right]^2}\right] \quad (11)$$

Since, $T_{s_k} - R_{s_k} = 1$ for the worst case, equation (11) reduces to equation (12)

$$Q = -M_{avg} = -0.5 * \left[1 + e^{-\left[\frac{1}{\sigma_g}\right]^2}\right] \quad (12)$$

Since the minimum and maximum deviation value is 0 and 1, we have two cases following from these constraints.

Case-1:$\sigma_g = 0$

Equation (12) reduces to Q = -0.5 ($e^{-\infty} = 0$) as shown using equation (13)

$$Q = -M_{avg} = -0.5 * \left[1 + exp\left[-1 * \left(\frac{1}{0}\right)^2\right]\right] = -0.5 \quad (13)$$

Case-2:$\sigma_g = 1$

The maximum value of deviation results in Q = -0.68395 as shown below.

$$Q = -M_{avg} = -0.5 * \left[1 + exp\left[-1 * \left(\frac{1}{1}\right)^2\right]\right] = -0.68395 \quad (14)$$

From these two cases, by choosing the negative maximum i.e $Q = -0.5$. The true dissimilarity, $D^{true}_{T_s, R_s}$ is given by

$$D^{true}_{T_s, R_s} = \frac{1 - M_{avg}}{0.5} \quad (15)$$

3.2 Standard Deviation

Using equation (15) and substituting expression for M_{avg} from equation (1), considering kth time slot, we have

$$D^{true}_{T_{s_k}, R_{s_k}} = \frac{0.5 * \left[1 - exp\left[-1 * \left(\frac{T_{s_k} - R_{s_k}}{\sigma_g}\right)^2\right]\right]}{0.5} \quad (16)$$

We know, $\delta = T_{s_k} - R_{s_k}$ for k[th] time slot. Equation (16) hence reduces to

$$D^{true}_{T_{s_k}, R_{s_k}} = 1 - e^{-\left[\frac{\delta}{\sigma_g}\right]^2} \quad (17)$$

Equating distances in both spaces,

$$1 - e^{-\left[\frac{\delta}{\sigma_g}\right]^2} = \delta \tag{18}$$

From (18), we obtain deviation in fuzzy space given by equation (19).

$$\sigma_g = \frac{\delta}{\sqrt{\ln_e(\frac{1}{abs(1-\delta)})}} \tag{19}$$

3.3 Threshold

Using equations (13) to (17), we can derive the equivalent threshold,δ^g

$$\delta^g = 1 - e^{-\left[\frac{\delta}{\sigma_g}\right]^2} \tag{20}$$

Equation (20), gives expression for threshold (δ^g). It is verified that thresholds in both space are equal.

4. PATTERN SUPPORT AND DISTANCE BOUNDS

Section 4.1 gives the expressions for estimation of maximum and minimum possible support values at kth time slot for association pattern of size,|S|=2 and Section 4.2 gives expressions for estimation of maximum and minimum possible support values of association pattern at kth time slot for size, |S|>2.

4.1 Support at kth Time Slot for Pattern Size, |S| =2

Assume P_{sk} and Q_{sk} are support values at kth time slot. The values for $P_{sk}{}'$ and $Q_{sk}{}'$ are given by $P_{sk}{}' = 1 - P_{sk}$ and by $Q_{sk}{}' = 1 - Q_{sk}$ respectively. For, kth time slot, the bound computation for temporal association pattern PQ is given by equation (21)

$$[PQ_{sk}]_{max} = P_{sk} - maximum\{(1 - P'_{sk} - Q_{sk}), 0\} \tag{21}$$

$$[PQ_{sk}]_{min} = maximum\{(1 - P'_{sk} - Q'_{sk}), 0\} \tag{22}$$

4.2 Support at kth Time Slot for Pattern Size, |S| >2

In this case, P_{sk} denotes support valueat kth time slot for patterns with size equal to |S|-1 and Q_{sk}denotes support value of singleton pattern at kth time slot. The values for $P_{sk}{}'$ and $Q_{sk}{}'$ are given by $P_{sk}{}' = 1 - P_{sk}$ and by $Q_{sk}{}' = 1 - Q_{sk}$ respectively. For, kth time slot, the bound computation for temporal association pattern PQ is given by equation (23). Here, $P_{sk}$and Q_{sk}

denotes all possible pattern combinations of size, |S|-1 and 1 respectively for temporal association pattern, PQ.

$$[PQ_{sk}]_{max} = P_{sk} - maximum\{(1 - P'_{sk} - Q_{sk}), 0\} \quad (23)$$

$$[PQ_{sk}]_{min} = maximum\{(1 - P'_{sk} - Q'_{sk}), 0\} \quad (24)$$

The resultant maximum and minimum values for association pattern, $[PQ]$ are obtained using equations (25) and (26)

$$[PQ_{sk}]_{max} = min([(P^{c1}Q)_{sk}]_{max}, [(P^{c2}Q)_{sk}]_{max}, \ldots\ldots\ldots\ldots, [(P^{cl}Q)_{sk}]_{max}) \quad (25)$$

$$[PQ_{sk}]_{min} = max([(P^{c1}Q)_{sk}]_{min}, [(P^{c2}Q)_{sk}]_{min}, \ldots\ldots\ldots\ldots, [(P^{cl}Q)_{sk}]_{min}) \quad (26)$$

P^{c1}, P^{c2}, ……., P^{cl} in equations (25) and (26) denote subset temporal patterns of size, |S|-1 for superset temporal pattern of size, |S|. The variable, k denotes kth time slot. $(P^{cl}Q)_{sk}$, denote the support at kth time slot for association pattern, PQ formed from sub patterns, P^{cl} and Q. $[PQ_{sk}]_{min}$ and $[PQ_{sk}]_{max}$ denotes the minimum and maximum support values at the kth time slot for association pattern, PQ. The support time sequences of resultant temporal patterns are sequences of support values at each time slot.

4.3 Dissimilarity Bound Computations

4.3.1 Upper-Lower Dissimilarity Bound,($D^{ULB}(U_s, R_s)$)

Let the maximum support sequence of a temporal association pattern, is denoted as $U_s = (U_{s_1}; U_{s_2}; U_{s_3} \ldots\ldots\ldots U_{s_m})$ and $R_s = (R_{s_1}; R_{s_2}; R_{s_3} \ldots\ldots\ldots R_{s_m})$ is the reference sequence. The distance computation at k[th] timeslot is given by equation (27)

$$D^{ULB}(U_s, R_s)^k = \begin{cases} 0 & ; \ R_{s_k} \leq U_{s_k} \\ 1 - e^{-\left[\frac{T_{s_k} - R_{s_k}}{\sigma g}\right]^2} & ; \ R_{s_k} > U_{s_k} \end{cases} \quad (27)$$

The upper-lower distance bound, D^{ULB} is hence given by equation (28).

$$D^{ULB}(U_s, R_s) = \frac{\sum_{k=1}^{k=m} D^{ULB}{}_{(U_s, R_s)}{}^k}{|k|} \quad (28)$$

4.3.2 Lower-Lower Dissimilarity Bound, ($D^{LLB}(L_s, R_s)$)

Let the minimum support sequence of a temporal association pattern is denoted by $L_s = (L_{s_1}; L_{s_2}; L_{s_3} \ldots\ldots\ldots, L_{s_m})$. The distance bound, $D^{LLB}(L_s, R_s)$ at k[th] timeslot is given by equation (29)

$$D^{LLB}(L_s, R_s)^k = \begin{cases} 0 & ; \ R_{s_k} > L_{s_k} \\ 1 - e^{-\left[\frac{T_{s_k} - R_{s_k}}{\sigma_g}\right]^2} & ; \ R_{s_k} \leq L_{s_k} \end{cases} \quad (29)$$

The distance bound, $D^{LLB}(L_s, R_s)$is given by equation (30)

$$D^{LLB}(L_s, R_s) = \frac{\sum_{k=1}^{k=m} D^{LLB}{}_{(L_s, R_s)}{}^k}{|k|} \quad (30)$$

4.3.3 Lower Bound, ($D^{LB}(T_s, R_s)$)

The lower bound distance between T_s, R_s denoted by $D^{LB}(T_s, R_s)$is defined as

$$D^{LB}(T_s, R_s) = D^{LLB}(L_s, R_s) + D^{ULB}(U_s, R_s) \quad (31)$$

4.4 Pruning

This section outlines the procedure to prune all infeasible temporal patterns without computing their true supports.

4.4.1 Pruning using $D^{ULB}(U_s, R_s)$

The distance bound, $D^{ULB}(U_s, R_s)$between true support sequence of temporal pattern and reference is compared to δ^g. If this distance, $D^{ULB}(U_s, R_s) \leq \delta^g$, temporal pattern is retained. In case, $D^{ULB}(U_s, R_s) > \delta^g$, temporal pattern is pruned.

4.4.2 Pruning using $D^{LB}(T_s, R_s)$

Compute the lower bound distance, $D^{LB}(T_s, R_s)$ using equation (31). The pattern is similar if $D^{LB}(T_s, R_s) \leq \delta^g$. If the lower bound distance, $D^{LB}(T_s, R_s) > \delta^g$ then, temporal pattern is not similar. To retain the temporal pattern, $D^{ULB}(T_s, R_s)$ is computed whenever $D^{LB}(T_s, R_s) > \delta^g$. If $D^{ULB}(T_s, R_s) \leq \delta^g$, temporal pattern must be retained. The pattern is pruned when $D^{ULB}(T_s, R_s) > \delta^g$.

4.4.3 Pruning Based on Retaining

A superset pattern is considered dissimilar if there exists at least one proper subset pattern that is both dissimilar and not retained. According to monotonicity of , $D^{ULB}(U_s, R_s)$, all such superset pattern temporal patterns are pruned.

5. ALGORITHM

Step-1
Find true support sequence of every singleton temporal pattern, (T_s).

Step-2

Obtain σ_g and δ^g for the transformed Gaussian space using equations (4) and (5).

Step-3

Using proposed measure, compute true distance for each level-1 (singleton) temporal pattern w.r.t reference pattern. If temporal pattern is not similar, then compare its$\boldsymbol{D^{ULB}}$. Retain or prune it based on $\boldsymbol{D^{ULB}}$. Go to step-7.

Step-4

For each next level, compute the maximum and minimum support sequence for temporal pattern of next level using true supports of temporal patterns at the previous level.

Step-5

Compute $\boldsymbol{D^{LB}(T_s, R_s)}$. If $\boldsymbol{D^{LB}(T_s, R_s)}$ exceeds δ^g , then $\boldsymbol{T_s}$ is not similar. If the distance, $\boldsymbol{D^{LB}(T_s, R_s)} \leq \delta^g$, the pattern may be similar, so compute its true support sequence. For this true support sequence of temporal pattern, obtain $\boldsymbol{D^{LB}(T_s, R_s)}$. If the distance, $\boldsymbol{D^{LB}(T_s, R_s)} \leq \delta^g$ pattern is similar. Otherwise it is not similar. When pattern is not similar, compare the distance, $\boldsymbol{D^{ULB}}$ to decide whether the pattern must be retained or not.

Step-6

Repeat step-5 for each temporal pattern. A pattern is similar when all its subset patterns at previous level are similar. A superset pattern is not similar if there is at least one subset pattern that is both not similar and is also not retained. A superset pattern has chances for being similar, even if its subset patterns are not similar but are retained. Go to step-7.

Step-7

Generate the next level temporal association pattern combinations and repeat the process from step-3 through step-6. Stop when pattern size, |S| = number of items in finite itemset.

Step-8

Output all valid similar temporal patterns

6. EXPERIMENT RESULTS AND DISCUSSIONS

The computation results of naïve, sequential, SPAMINE (2012) and our approach are recorded and discussed in this section. We compare experiment results using proposed measure to approaches in (J. S. Yoo, 2009 and Yoo J. S, 2012). Figure 1 depicts true support computations required for a randomly generated temporal database denoted as TTS1000-TS100-I20. TTS indicates number of transactions per time slot, TS is number of time slots, I is the total number of items in finite itemset. The generated temporal database hence comprises of one lakh transactions. The total number of possible temporal association patterns possible is 2^{20} which are 1 billion temporal patterns. For example, a database generated over 10 items has 1024 different possible pattern combinations. Figure 2 shows total

number of true support computations required using proposed support bound estimation procedure for different thresholds 0.15, 0.25 and 0.35.

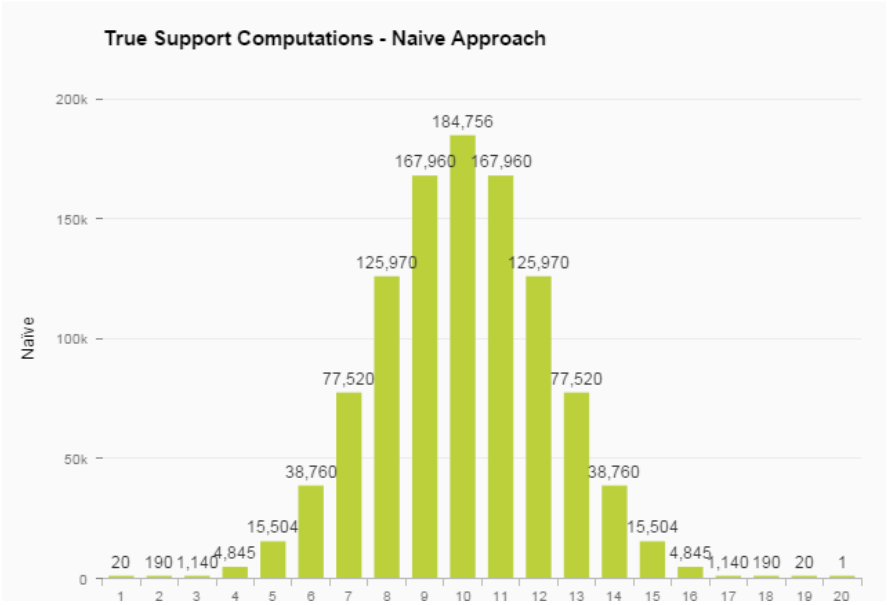

Figure 1. True Support Computations – Naïve Approach

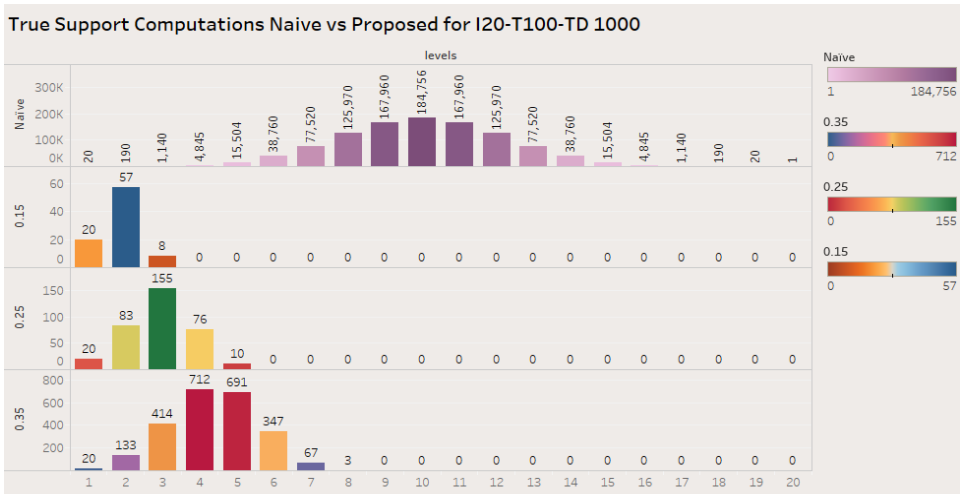

Figure 2. True Support Computations – Proposed Method

Figure 3 depicts the graph comparing the true support computations carried using naïve and proposed approaches for a threshold, $\delta = 0.35$.

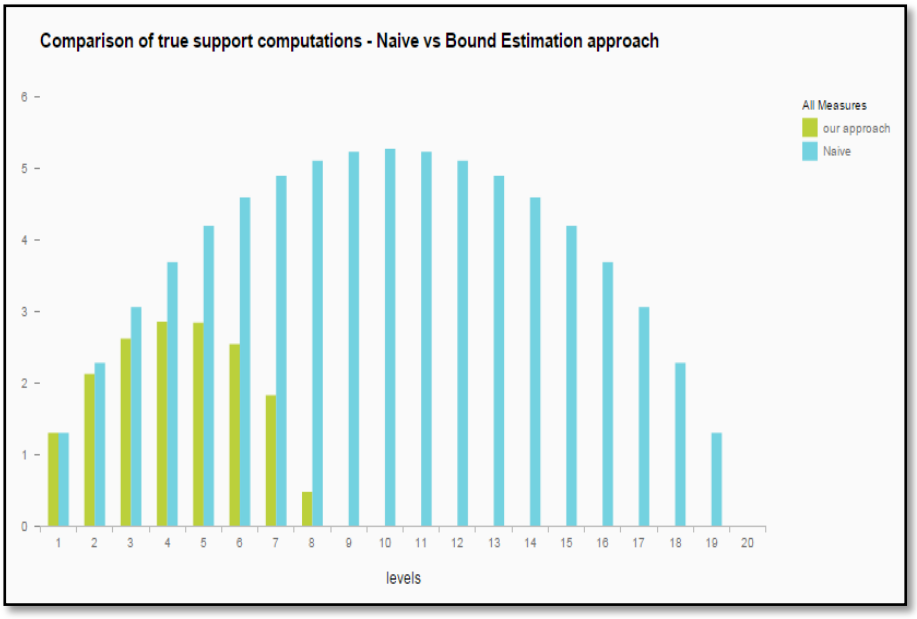

Figure 3. Naïve vs Proposed for Threshold for I20-T100-TS1000 and δ = 0.35

Figure 4 depicts the graph comparing retained association patterns to consider for similarity when adopting the naïve and proposed approaches for thresholds, δequal to 0.15, 0.25 and 0.35.

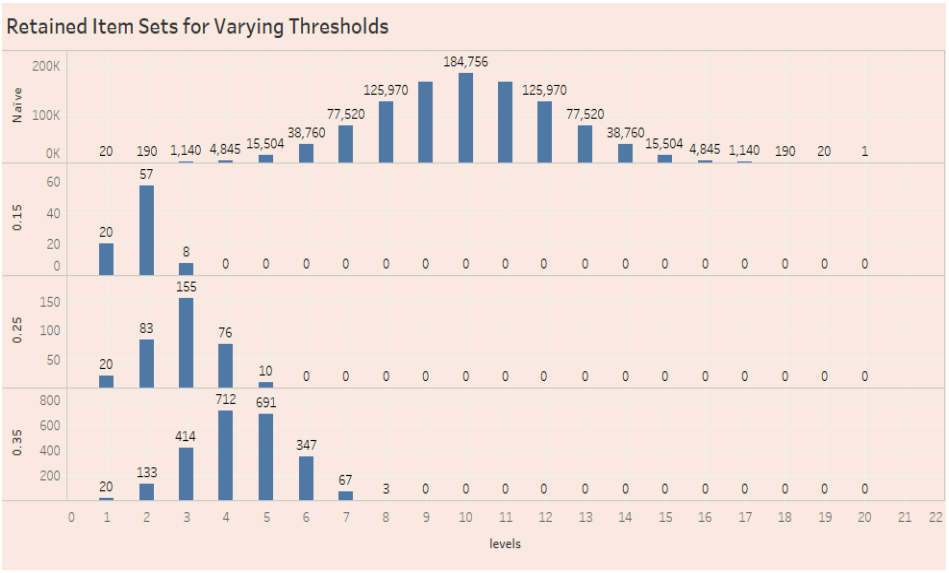

Figure 4. Retained Patterns - Naïve vs. Proposed by Level for threshold, δ = 0.25

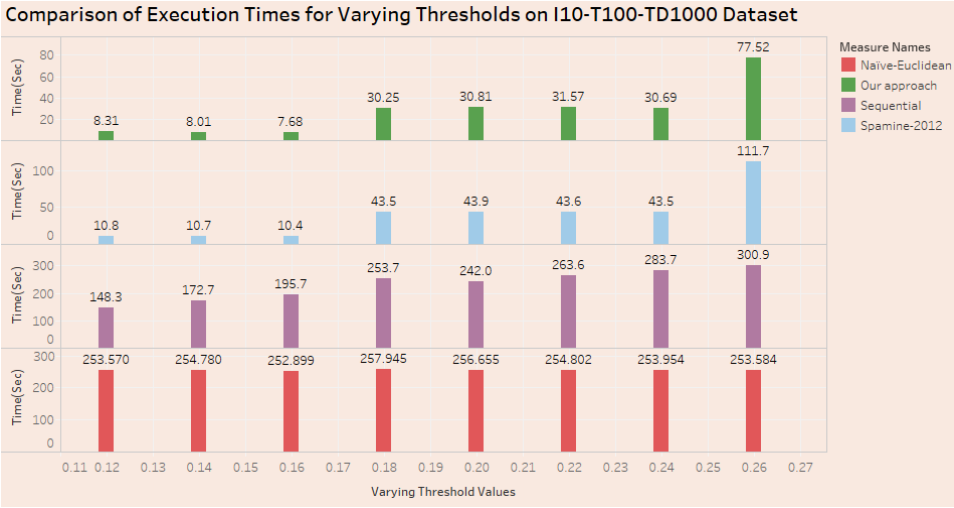

Figure 5. Execution Times over Temporal Database I20-T100-TD1000 for Varying Thresholds

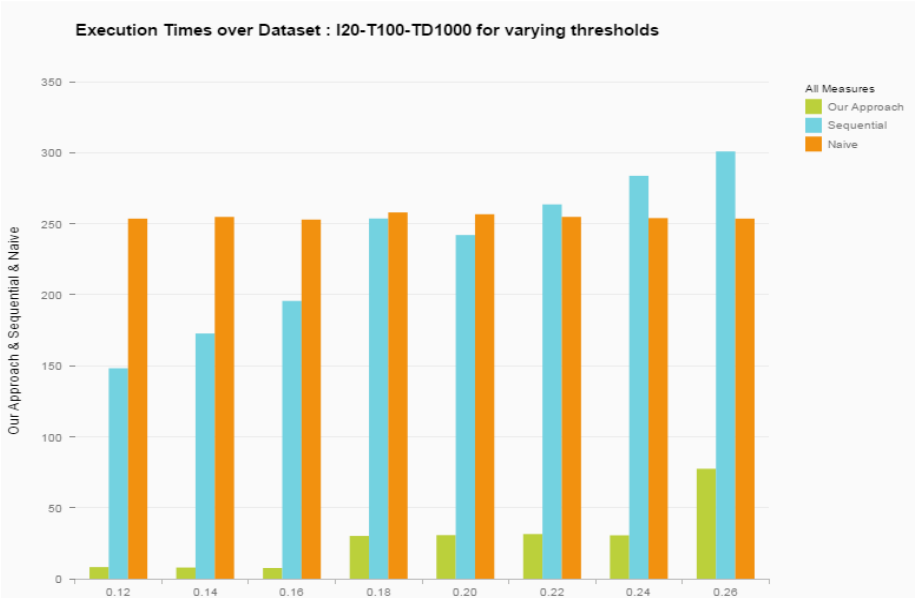

Figure 6. Execution Times over Temporal Database I20-T100-TD1000 for Varying Thresholds

The execution times of our approach is compared to naïve, sequential and Spamine approaches for a random temporal database generated over 100 time slots, 1000 transactions per time slots with 20 items for varying thresholds. These are represented in Figure 5. It is seen that the proposed approach is comparatively better to naïve, sequential and Spamine approaches. Figure 6 shows the comparison of execution times for our approach w.r.t naïve and sequential approaches for different thresholds considering one lakh transactions defined over 100 time slots.

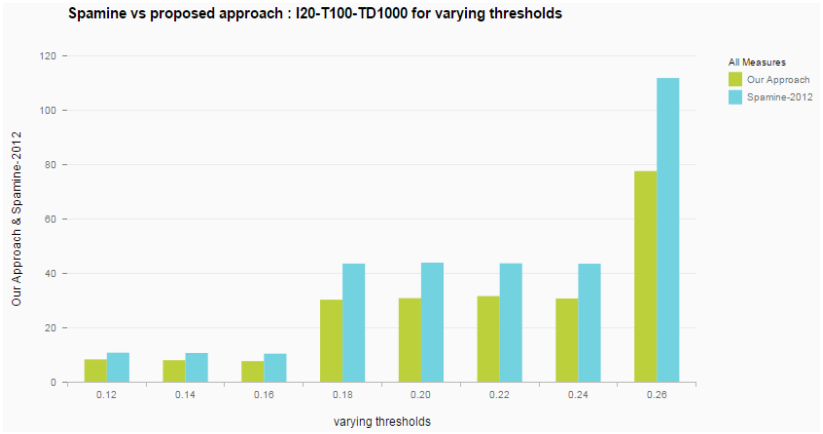

Figure 7. SPAMINE vs. Proposed Approach for Varying Thresholds

Figure 7 shows comparison of execution times of Spamine and our approach. Figure 8 compares execution times of naïve and sequential approaches to our approach for varying number of transaction items for a temporal database consisting of 100 time slots and 1000 transactions per time slot for a constant threshold equal to 0.2.

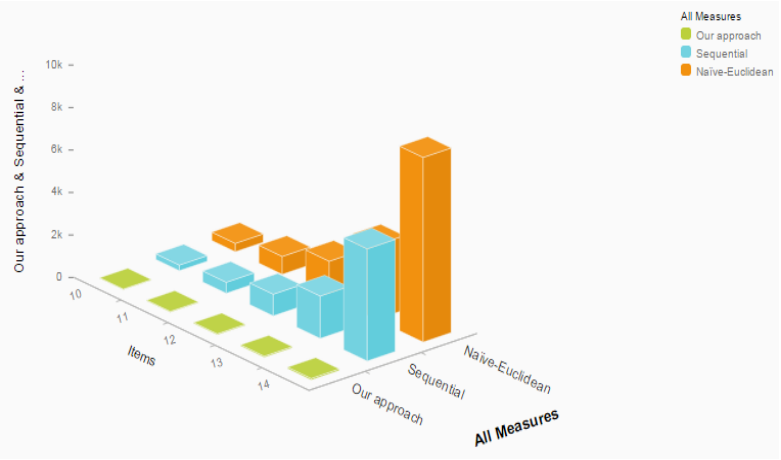

Figure 8. Execution Times of Various Approaches for Varying Items

The execution time of our approach is compared to Spamine in figure 9. In the experimental results recorded and analyzed for various datasets generated it is seen that our approach is comparatively better to other approaches.

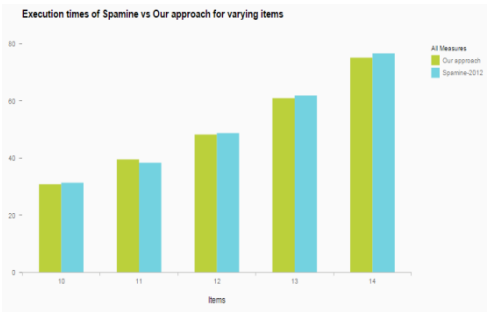

Figure 9. Execution Times of SPAMINE and our approach

7. CONCLUSIONS

The contribution in this paper includes designing a novel dissimilarity measure to discover pattern trends that varies similar to another temporal pattern usually termed as the reference. Expressions are defined for standard deviation, the threshold in transformed space and an approach for estimating support limits of temporal patterns is proposed. The monotonicity property of dissimilarity measures helps to prune infeasible patterns at early stages eliminating unnecessary true support computations. Experiment results prove that using the proposed approach true support scans and execution time are reduced significantly when compared to naïve, sequential and Spamine approaches.

REFERENCES

Abdelsalam M Maatuk, Muhammad A Ali, & Shadi Aljawarneh. (2015). Translating relational database schemas into object-based schemas: university case study. *Recent Patents on Computer Science, 8(*2), 122-131.

Abdelsalam M. Maatuk, M. Akhtar Ali, & Shadi Aljawarneh. (2015). An algorithm for constructing xml schema documents from relational databases. *Proceedings of the International Conference on Engineering & MIS 2015 (ICEMIS '15)*, 1-6. doi : 10.1145/2832987.2833007

Aravind Cheruvu & V. Radhakrishna. (2016). Estimating temporal pattern bounds using negative support computations. *Proceedings of the International Conference on Engineering and MIS*, 1-4. doi: 10.1109/ICEMIS.2016.7745352

Chun-Hao Chen, Guo-Cheng Lan, Tzung-Pei Hong, & Shih-Bin Lin. (2016). Mining fuzzy temporal association rules by item life spans. *Applied Soft Computing, 41*, 265-274.

Erich Fuchs, Thiemo Gruber, Helmuth Pree, & Bernhard Sick. (2010). Temporal data mining using shape space representations of time series. *Neuro Computing, 74*(1-3), 379-393.

Erich Fuchs, Thiemo Gruber, Jiri Nitschke, & Bernhard Sick. (2009). On-line motif detection in time series with Swift Motif. *Pattern Recognition, 42*(11), 3015-3031.

J. S. Yoo & S. Shekhar. (2009). Similarity-profiled temporal association mining. *IEEE Transactions on Knowledge and Data Engineering, 21*(8), 147-1161. doi: 10.1109/TKDE.2008.185

Jung-Yi Jiang, Ren-Jia Liou, & Shie-Jue Lee. (2011). A fuzzy self-constructing feature clustering algorithm for text classification. *IEEE Transactions on Knowledge and Data Engineering,23*(3), 335-349.

M. Gupta, J. Gao, C. C. Aggarwal, & J. Han. (2014). Outlier detection for temporal data: a survey. *IEEE Transactions on Knowledge and Data Engineering, 26*(9), 2250-2267.

Manish Gupta, Jing Gao, Yizhou Sun, Jaiwen Han. (2012). Community trend outlier detection using soft temporal pattern mining. *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases, 1*(2), 692-708. doi: 10.1007/978-3-642-33486-3_44

Maria Carla Calzarossa, Daniele Tessera. (2015). Modeling and predicting temporal patterns of web content changes. International Journal of Network and Computer Applications, 56, 115-123. http://dx.doi.org/10.1016/j.jnca.2015.06.008

Mario Stefanelli. (2016). Temporal Data Mining. http: //www.labmedinfo.org. Retrieved on October15, 2016.

Mohammed R Elkobaisi, Abdelsalam M Maatuk, & Shadi Aljawarneh. (2015). A proposed method to recognize the research trends using web-based search engines. *Proceedings of the International Conference on Engineering and MIS,*1-4. doi : 10.1145/2832987.2833012

Muneer Bani Yassein, Shadi Aljawarneh, Esra'a Masa'deh. (2017). A new elastic trickle timer algorithm for Internet of Things. Journal of Network and Computer Applications. Available online 27 January 2017. Paper retrieved from http://dx.doi.org/10.1016/j.jnca.2017.01.024

Negin Keivani, Abdelsalam M Maatuk, Shadi Aljawarneh, & Muhammad Akhtar Ali. (2015). Towards the maturity of object-relational database technology: promises and reality. *International Journal of Technology Diffusion (IJTD),6*(4),1-19. doi: 10.4018/IJTD.2015100101

Oliver Brdiczka, Norman Makoto Su, & James Begole. (2010). Temporal task foot printing: identifying routine tasks by their temporal patterns. *Proceedings of the 15th International Conference on Intelligent User Interfaces (IUI '10), 281-284.* doi:10.1145/1719970.1720011

S. Aljawarneh, V. Radhakrishna, P. V. Kumar and V. Janaki. (2016). A similarity measure for temporal pattern discovery in time series data generated by IoT. *Proceedings of the International Conference on Engineering and MIS,* 1-4.doi: 10.1109/ICEMIS.2016.7745355

Saaed Mehrabi et.al. (2015). Temporal Pattern and Association Discovery of Diagnosis Codes Using Deep Learning. International Conference on Healthcare Informatics. *Proceedings of the International Conference on Engineering and MIS,* 408-416. doi: 10.1109/ICHI.2015.58

Shadi A. Aljawarneh, Mohammed R. Elkobaisi, & Abdelsalam M. Maatuk. (2016). A new agent approach for recognizing research trends in wearable systems. *Computers & Electrical Engineering.* Paper retrieved from http://dx.doi.org/10.1016/j.compeleceng.2016.12.003

Shadi A. Aljawarneh, Raja A. Moftah, & Abdelsalam M. Maatuk. (2016). Investigations of automatic methods for detecting the polymorphic worms signatures. *Future Generation Computer Systems, 60,* 67-77.

Shadi A. Aljawarneh, Vangipuram Radhakrishna, P.V.Kumar, V. Janaki. (2017). G-SPAMINE: An approach to discover temporal association patterns and trends in internet of things. *Future Generation Computer Systems.* http://dx.doi.org/10.1016/j.future.2017.01.013

Stephen G. Matthews, Mario A. Gongora, Adrian A. Hopgood, & Samad Ahmadi. (2013). Web usage mining with evolutionary extraction of temporal fuzzy association rules. *Knowledge-Based Systems, 54,* 66-72. http://dx.doi.org/10.1016/j.knosys.2013.09.003

V. Radhakrishna, P. V. Kumar & V. Janaki. (2016). Mining of outlier temporal patterns. *Proceedings of the International Conference on Engineering and MIS,* 1-6. doi: 10.1109/ICEMIS.2016.7745343

V. Radhakrishna, P. V. Kumar, V. Janaki & S. Aljawarneh. (2016). A computationally efficient approach for temporal pattern mining in IoT. *Proceedings of the 2016 International Conference on Engineering & MIS (ICEMIS),* 1-4. doi:10.1109/ICEMIS.2016.7745354

V. Radhakrishna, P. V. Kumar, V. Janaki & S. Aljawarneh. (2016). A similarity measure for outlier detection in time stamped temporal databases. *Proceedings of the 2016 International Conference on Engineering & MIS (ICEMIS),* 1-5.doi: 10.1109/ICEMIS.2016.7745347

Vangipuram Radhakrishna, P.V.Kumar, & V.Janaki. (2015). A novel approach for mining similarity profiled temporal association patterns using Venn diagrams. *Proceedings of the ACM International Conference on Engineering & MIS (ICEMIS).* doi: 10.1145/2832987.2833071

Vangipuram Radhakrishna, P.V.Kumar, & V.Janaki. (2016). A novel similar temporal system call pattern mining for efficient intrusion detection. *Journal of Universal Computer Science, 22*(4), 475-493.

Vangipuram Radhakrishna, P.V.Kumar, & V.Janaki. (2016). An efficient approach to find similar temporal association patterns performing only single database scan. *Revista Tecnica De La Facultad de Ingeneria Universidad Del Zulia, 39*(1), 241-255. doi:10.21311/001.39.1.25

Vangipuram Radhakrishna, P.V.Kumar, &V. Janaki. (2015). A survey on temporal databases and data mining. *Paper presented at the International Conference on Engineering & MIS (ICEMIS),* Istanbul, Turkey. Abstract retrieved from http://dx.doi.org/10.1145/2832987.2833064

Vangipuram Radhakrishna, P.V.Kumar, &V.Janaki. (2015). A novel approach for mining similarity profiled temporal association patterns. *Revista Tecnica De La Facultad de Ingeneria Universidad Del Zulia, 38*(3), 80-93.

Vangipuram Radhakrishna, P.V.Kumar, &V.Janaki. (2015). An approach for mining similarity profiled temporal association patterns using Gaussian based dissimilarity measure *Paper presented at the International Conference on Engineering & MIS (ICEMIS),* Istanbul, Turkey. Abstract retrieved from http://dx.doi.org/10.1145/2832987.2833069

Vangipuram Radhakrishna, P.V.Kumar, &V.Janaki. (2016). Looking into the possibility of novel dissimilarity measure to discover similarity profiled temporal association patterns in IoT. *Proceedings of the International Conference on Engineering & MIS,* 1-5.

Vangipuram Radhakrishna, P.V.Kumar, &V.Janaki. (2016). Mining Outlier Temporal Association Patterns. *Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies (ICTCS '16).* doi: 10.1145/2905055.2905320

Vangipuram Radhakrishna, P.V.Kumar, &V.Janaki. (2016). Normal Distribution Based Similarity Profiled Temporal Association Pattern Mining (N-SPAMINE). *Database Systems Journal, 7*(3), 22-33. http://www.dbjournal.ro/archive/25/25_3.pdf

Vangipuram Radhakrishna, Shadi A. Aljawarneh, Puligadda Veereswara Kumar, Kim-Kwang Raymond Choo. (2016). A novel fuzzy Gaussian-based dissimilarity measure for discovering similarity temporal association patterns. *Soft Computing*. First Online: 18 November 2016. doi:10.1007/s00500-016-2445-y

W. Ouyang, S. Luo, & Q. Huang. (2007). Discovery of direct and indirect association patterns in large transaction databases. *Proceedings of the International Conference on Computational Intelligence and Security,* (pp 167-170). Harbin. doi: 10.1109/CIS.2007.112

Weimin Ouyang & Qinhua Huang. (2010). Mining Direct and Indirect Association Patterns with Multiple Minimum Supports. *Proceedings of the International Conference on Computational Intelligence and Software Engineering,* Wuhan. doi: 10.1109/CISE.2010.5677032

Weiqiang Lin, Mehmet A. Orgun, Graham J. Williams. (2015). An overview of temporal data mining. *Paper presented at the meeting of the Australasian Data Mining Workshop*. Retrieved from http://togaware.com/papers/adm02.pdf

Y. S. Lin, J.Yi.Jiang, & S.J.Lee. (2014). A similarity measure for text classification and clustering. *IEEE Transactions on Knowledge and Data Engineering, 26*(7), 1575-1590. doi: 10.1109/TKDE.2013.19

Yoo JS. (2012). Temporal data mining: similarity-profiled association pattern. *Data mining: Foundations and Intelligent Paradigms, 23*, 29–47. doi: 10.1007/978-3-642-23166-7_3