Normal Distribution Based Similarity Profiled Temporal Association Pattern Mining (N-SPAMINE)

Vangipuram RADHAKRISHNA[1], P.V.KUMAR[2], V.JANAKI[3]
[1]VNR VignanaJyothi Institute of Engineering and Technology, INDIA
[2]University College of Engineering, Osmania University, INDIA
[3]Vaagdevi Engineering College, INDIA
radhakrishna_v@vnrvjiet.in, pvkumar58@gmail.com , janakicse@yahoo.com

Temporal patterns in time stamped temporal databases are sequences of support values and hence, they are represented as vectors. This makes it challenging to obtain similar association patterns in context of time stamped temporal databases whose support trends change similar to a reference support sequence trend. The main idea of this work is to study the possibility of applying normal distribution concept to find similarly varying temporal patterns. This paper introduces a new approach, called N-SPAMINE for mining similarity profiled temporal association patterns by applying normal distribution which is first of its kind of approach for finding similar association patterns and uses the novel dissimilarity measure for obtaining dissimilarity between chosen temporal pattern and the reference. The results show that the proposed approach is correct and complete.
Keywords: *Temporal, Normal distribution, Z-score, Support Sequence, Outlier*

1 Introduction

A wide number of applications exist in our everyday life which continuously originate temporal data. A temporal dataset may be viewed as a collection of temporal data objects. A database system can be modelled to consist of data objects showing temporal behaviour for a wide range of applications. Such a database system may later be used as a temporal dataset. Consider a non-temporal database, in which the entity “employee” has attributes name, pan number, father name, voterID, adhar Number. These attribute values for a sales person do not change and remains same with respect to the changing time. This may not be true for all attributes of an “employee”.
For example, consider the attributes such as address, mobile number, designation and salary. The attribute value of these attributes does not remain same and may vary over a period of time. Such attributes are called temporal attributes which show temporal behaviour. Some other examples include medical datasets, credit datasets, and co-authorship in DBLP [1]. The details of authorship, co-authorship details for an author keep varying and updating over a period of time and shows temporal behaviour. Some of the examples for temporal database include relationships with time-stamped tuple, databases with versioning. Common examples for temporal data include time series data such as EEG readings, stock market data, event sequences such as weblog data, medical data, sensor generated data and temporal databases. Temporal data mining techniques include temporal prediction, temporal clustering, temporal classification, temporal pattern discovery, search and retrieval. Of all these techniques, pattern discovery has been concentrated more by researchers from view point of temporal and spatio-temporal datasets.
Recently soft temporal pattern mining is studied which involves applying machine learning techniques to understand evolutionary behaviour of temporal data. Currently, researchers are addressing methods for discovering outliers from temporal data in soft temporal perspective which has wide scope for research in coming years. The works in this direction includes initiation from research these contributions [5, 15].

1.1 Motivation for current work

Finding temporal association patterns whose support sequence trends vary similar to a given reference trend is an important research problem in temporal databases and temporal, spatial data mining. Many real time and modern applications require finding those patterns which temporally vary in the same scale to a chosen reference trend. Discovering such temporal trends or patterns is very challenging and is not addressed in the data mining literature which is significant. The pioneer of this work is from the authors [16]. They apply the Euclidean distance measure and use the support sequence values of temporal patterns to find the similarity between chosen pattern and reference. This work of the authors motivated us and we started to work in this direction as there is further research scope in this direction. The approach for estimating temporal pattern support limits are addressed in few of our several works [17, 20, 27-31]. Approaches in [27-29, 31] adopted Euclidean distance but did not address new similarity measures for temporal databases. Also, these works did not consider the feasibility of normal distribution into consideration. In this work, we look for the possibility and feasibility of applying the normal distribution to efficiently obtain all valid temporal pattern trends that are essentially similar to the reference support trend.

Section-2 reviews literature. Section-3 discusses proposed method and measure. Section-4 demonstrates the example problem using case study, Section-5 concludes this paper.

2. Literature Review

Objects with similar temporal behaviour grouped under a common category through unsupervised approach may be called a community. In [1], given a snapshot of temporal dataset, the authors discuss the approach of discovering community outliers. Discovering such outliers is called "Evolutionary Community Outlier Detection ".

Many times it may be required to study the collocation behaviours of data objects whose information is stored as spatio-temporal data. The problem of finding the similar data objects which move together such that intra-distances between data objects do not exceed certain threshold is common in surveillance applications. Typical data mining queries includes, finding the suspiciously moving people based on the pattern movement of data objects, vehicle convoys and flocks of moving data objects. In [2], the problem of flock discovery is studied and a framework for discovering patterns in stream data is proposed.

In [3], Fabrizio Angiulli studied the problem of outlier prediction and detection. The approach proposed is an outlier detection approach which is distance based and uses unsupervised learning approach. The algorithm uses the concept called as "solving set", which is used to decide whether a new incoming stream data object is an outlier or normal.

Finding abnormal patterns is considered more important than finding normal patterns many times in the applications related to security. This is mainly because abnormal patterns are anomalies (outliers) and must be identified without missing them. In temporal data mining, detection of outliers becomes more challenging and interesting. Most of existing works treat the outliers as the binary decision problem where we either decide the data object is outlier or not an outlier at all.

In [4], Markus address, the problem by considering and viewing in different perspective such as assigning "degree of outlierness ", for each data object within a community or group. They call it as "local outlier factor", LOF. The value of LOF decides the degree that the considered data object is an outlier w.r.t other neighbouring objects. Several day to day applications generate a lot of temporal data which consists of data generated from bank transactions, traffic data, generated data from news feeds, sensor data, and medical data. To understand such evolving data, we must study the temporal behaviour of these evolving data objects.

In [5], address a new approach for discovering evolutionary behaviour of objects. Such objects are called "community trend outliers". Temporal data can be distributed data which includes temporal distributed data and spatial sensor data, time series data, stream data, network data, spatio-temporal data. A recent survey on "mining outliers in temporal data", is carried out in [6]. In their work, the authors discuss specific challenges for detecting outliers from temporal data, classification of temporal outlier analysis, various prediction models to specify few of the contributions. A framework for "community outlier detection", is also discussed and in this contribution [6].

Discovery of temporal association rules (TAR) from temporal datasets and then using these generated association rules for the figuring out rare items or outliers or unexpected trends is discussed in [7]. The outliers generally do not violate the semantics but are not frequent as compared to normal. Here, generated TARs are multidimensional, time series and quantitative in nature. The pruning of association rules is carried out by using unexpectedness measure. The outliers are detected from the stock market data considered for mining by finding stock splits. Stock splits indicate sudden change in the stock trend which constitutes outlier behaviour.

A three step approach for detecting spatio-temporal outliers from large databases is proposed in [8] which involve performing unsupervised learning and finding nearest temporal and spatial neighbours. The algorithm considers temporal, spatial and non-spatial data object values to detect outliers.

The concept of minimum spanning tree (MST) is TARs [9] from meteorological data which is one of the forms of spatio-temporal data. The importance and advantage of this approach is that we can predict the damages ahead which may be caused by climate conditions and atmosphere for people or things. The objective of using MST is to achieve optimization.

Outlier detection in most of the applications generally considers single time instance. On contrary, the outlier discovery in vehicular traffic data involves handling multiple time instances, i.e dynamic time. The work in [10] discusses considers the traffic data and propose the approach for temporal outlier discovery considering traffic data. Community outlier discovery has been one of the recent research interests concentrated.

In [11], the authors consider the temporal dataset and discover community outliers using the concept of transition matrix and community belongingness matrix.

The application of association pattern mining (APM) is discussed in context of ontology in [12] which considers semantic property not considered when using conventional text mining techniques.

Contrary to frequent items, there exist situations where the infrequent items can provide important insights into the datasets. This point forms the basis for the work in [13]. Here, the authors propose an approach for discovering both indirect and direct association patterns. Association rules are generated considering only single support values. In [14], association rules are generated by considering multiple minimum support values for discovering both direct and indirect association rules. The problem of rare item is also discussed as part of this work.

In [15], temporal and deep learning techniques are applied to mine the TARs (temporal association rules). In [16], authors consider the temporal data (time stamped temporal data), to retrieve all TAPs (temporal association patterns) that are similar to the reference and satisfy given value of threshold. A dissimilarity measure for predicting similar patterns is discussed in

[17] considering temporal dataset for temporal data. In [18], an approach for identifying routine tasks based on their temporal regularities is proposed termed as the "temporal task foot printing".

Given an "interval based event data" [19], the problem of identifying patterns from such event data is complex and challenging because of the complex relationships that exist in event based data. In this, authors propose a new approach and representation to handle such complex relationships. The approach is called the 'incision strategy", which uses the representation termed as the "coincidence representation ". An algorithm called "CTMiner" is proposed to discover frequent pattern from interval based event data which is scalable and efficient. Intrusion detection using temporal pattern is discussed in [20].

An Itemset is said to be frequent if the support of an Itemset is equal to or exceeds a specified threshold. Sometimes, infrequent Itemset may be frequent only for certain period of time. A hierarchical granular framework is considered and the TARs are generated organizing time into granules [21]. In [22], authors describe temporal sub events using five temporal relationships and extend FP-Tree algorithm for temporal domain, called TFP-growth tree. The objective is to generate association rules from video databases which are event based by using cloud infrastructure. Each item in a temporal transaction database has its life time and transactions often contain quantitative values. In [23], the quantitative values are transformed into equivalent fuzzy values by using membership functions. A fuzzy approach is proposed to generate FTARs (fuzzy temporal association rules).

Most of the works in the literature address to find semantic relationship between entities without considering temporal nature. Implicit and Explicit temporal relationships between entities is studied and extended for web search engine in [24]. Some more recent related works includes [25-31].

3. Proposed Method

Our approach for retrieving similar association patterns from time stamped temporal database uses the concept of z-score in normal distribution and examines the possibility for applying z-score computation to mine all valid temporal association patterns applying novel temporal dissimilarity measure. Section-3 of this paper discusses the Z-score computation process for temporal support sequences. The resultant z-score computed is then transformed by computing probability value for these obtained z-scores from standard normalization tables. Section-3.2, discusses temporal dissimilarity measure which uses the z-score probabilities obtained in Section-3.1. A working example is outlined in Section-4 which explains the procedure in detail with the suitable example. The paper is concluded in Section-5.

3.1 Normalizing Support Sequences

This section explains the procedure to compute the z-score for support sequences obtained for all temporal association pattern combinations which are possible from the input temporal database. Given a temporal support sequence expressed as a vector of support values, the idea is to transform these support values into corresponding z-score values. The computation of z-score requires specifying the value for standard deviation. This value is obtained by using the novel expression designed for standard deviation as part of defining the proposed similarity measure.

3.1.1 Computing Z-score and Normalized Probability for Temporal Support Sequences

Let, TP = [TP_1, TP_2, TP_3 ……., TP_N] denote support values of temporal pattern 'TP' for 'N' time slots and the reference temporal pattern is denoted as V= [V_1, V_2, V_3 ……., V_N]. Here, the values V_1, V_2, V_3……., V_N are the support values for N time slots.

The z-score for TP w.r.t V for each time slot is computed using equation 1

$$Z_i = \frac{TP_i - V_i}{\sigma_g} \quad (1)$$

In equation (1), the variable 'i' represents i^{th}, time slot.

The temporal pattern, TP is now expressed as a sequence of z-score values represented by equation (2)

$$\mathbf{Z^{TP}} = [Z_1, Z_2, Z_3, \ldots\ldots\ldots Z_N] \quad (2)$$

These Z-score sequences obtained are then transformed into probability sequence by computing probabilities of z-score values at each time slot denoted by equation (2).

From equation (2), we get the probability vectors denoted as

$$\mathbf{P(Z^{TP})} = [\mathrm{P(Z_1), P(Z_2,), \ldots. P(Z_N)}\ ] \quad (3)$$

Since, we have used the reference temporal sequence for obtaining the z-score, the resultant reference temporal pattern is now treated as a sequence of zero values (z-score values) expressed as equation (4)

$$\mathbf{P(V^{TP})} = [0, 0, 0 \ldots, 0] \quad (4)$$

The distance between probability vectors, TP and V is a sequence vector and is given by Equation (5) that is obtained by considering values of the probabilities at each time slot.

$$\mathbf{D\ (P(Z^{TP}), P(V^{TP}))} = (P\ (Z_1)\text{–}0,\ P\ (Z_2)\text{–}0, \ldots\ldots\ldots\ P\ (Z_N)\text{–}0) \quad (5)$$

Equation (5) hence reduces to Equation (6) expressed as

$$D\ (P^{Z(TP)},\ P^{V(TP)}) = P^{Z(TP)} \quad (6)$$

Equation (6), expresses fact that the distance vector obtained is same as the probability vector of temporal pattern, $\mathbf{P^{Z(TP)}}$.

3.2 Proposed Measure

The design of proposed measure for finding similar temporal association patterns is discussed in this section. The measure uses support sequence values of temporal and reference pattern at each time slot and obtains the dissimilarity value.

We regard our contributions as follows:

1. Coming out with a novel approach i.e Z-score approach for transforming temporal support values to standard scores.
2. Design of novel dissimilarity measure
3. Defining expression for standard deviation
4. Defining threshold expression

Given a temporal pattern, TP = [TP_1, TP_2, TP_3 ……., TP_N] and the reference temporal pattern, V= [V_1, V_2, V_3 ……., V_N]. The dissimilarity between these two temporal patterns is computed using (7)

$$D_V^{TP} = \frac{(1-\mu_V^{TP})}{2} \quad (7)$$

The value of μ_V^{TP} in the equation (7) is defined in the equation (8),

$$\mu_V^{TP} = \frac{\sum_{k=1}^{k=N} e^{-\left(\frac{\mathrm{P(Z_K)}}{\sigma_g}\right)^2}}{|k|} \quad (8)$$

Equation (7) is a function of normalized probability value. The probability value for every time slot is obtained by considering the z-score value at each time slot of temporal pattern and normalized distribution look up table.

The advantage of the temporal measure using the z-score probability values is that the resultant similarities are always evaluated to tight upper and lower bounds.

An upper bound value is unity and lower bound equals zero. This is not true for Euclidean measure. Also the Euclidean measure fails when we compute the distance in traditional way when using normalized probability values. This is overcome using the proposed measure.

3.2.1 Expressions for Threshold and Standard Deviation

Suppose, threshold in Euclidean space is denoted using the symbol 'δ' and standard deviation by 'σ_g'.
Equation (9) represents expression to compute the deviation which is actually a function of threshold

$$\sigma_g = \frac{\delta}{\sqrt{\ln_e\left(\frac{1}{1-\delta}\right)}} \quad (9)$$

The expression for deviation is the novel expression designed for the purpose of estimating the z-score value for a temporal pattern at each time slot. It is this deviation value which helps us in finding the z-score values.
The conventional expression that exists in the literature does not suit the current purpose. This brings the requirement for design of new standard expression. Also, the standard deviation value used remains same for each time slot, irrespective of data distribution and time slots which is another added advantage.
The threshold in normalized space is given by equation (10)

$$\delta^g = \frac{1 - exp\left[-1 * \left(\frac{\delta}{\sigma_g}\right)^2\right]}{2} \quad (10)$$

The threshold value computed using equation (10) is used as dissimilarity limit allowed for considering whether a temporal pattern is similar or not.

4. Case Study

We explain the working of proposed method to find all valid similar temporal patterns. As discussed in section-3, our approach to find similar patterns is based on Z-score and normalized probability computations for support sequence vectors of temporal patterns. The first step towards finding all valid temporal association patterns hence requires transforming temporal pattern support sequences into normalized probabilities. These values obtained are later used in the similarity measure to find if the corresponding temporal pattern in similar or an outlier.
Consider the transaction database in Fig.1. It contains 20 transactions split across two timeslots. Each timeslot has 10 transactions. Transactions TR1 to TR10 correspond to first timeslot and TR11 to TR20 correspond to second timeslot. There are at most 3 items in a transaction. i.e database is defined over A, B and C only. Such a database in Fig.1 is called time stamped temporal database.

Transactions	Items	Transactions	Items
TR1	A B C	TR11	B C
TR2	A B C	TR12	A B C
TR3	A	TR13	A B C
TR4	B C	TR14	B
TR5	C	TR15	A B C
TR6	A B C	TR16	B C
TR7	C	TR17	A B C
TR8	A C	TR18	B
TR9	B C	TR19	A
TR10	A B C	TR20	B C

Fig. 1. Transaction Database

The possible combinations of association patterns or item sets are hence seven. They are A, B, C, AB, AC, BC and ABC. The support value for these patterns is not a single value. It is a vector sequence defined over the timeslots. In this example, the number of timeslots is two. i.e the pattern is expressed as a vector of support values at two timeslots. Fig.2 shows true support values for all pattern combinations.

Pattern	T1	T2
[A]	0.6	0.5
[B]	0.6	0.9
[C]	0.9	0.7
[AB]	0.4	0.4
[AC]	0.5	0.4
[BC]	0.6	0.7
[ABC]	0.4	0.4

Fig. 2. True Supports

The traditional approach for finding, if these patterns are similar or dissimilar to a reference pattern is to apply Euclidean distance. This is because temporal patterns in our case are vectors and not single values and hence the Euclidean distance measure can be used to find distance between such vectors. However, the Euclidean distance is not directly applicable when considered normal distribution. Also, one more drawback of this measure is it does not contain strict higher bound and can even be not finite value. Fig.3 shows the distance to reference pattern for a given temporal pattern, obtained through applying the Euclidean distance metric.

Pattern	Euclidean	Pattern	Euclidean
[A]	0.1581	[AC]	0.1581
[B]	0.2550	[BC]	0.1581
[C]	0.3606	[ABC]	**0.1414**
[AB]	**0.1414**		

Fig. 3. Pattern Dissimilarity w.r.t Reference

The distance values in Fig.3 are normalized distance values and lie between 0 and 1. These distance values are obtained by dividing the obtained Euclidean distance by $\sqrt{N}$, where N is the number of time slots. This is done to make the euclidean distance have a strict upper bound limit=1. However, even such a normalized distance value computed is also not suitable and fails when we adopt the normal distribution based temporal support sequences. Our approach overcomes the above explained disadvantage and finds all valid patterns whose trend varies similar to the reference.

The computation of z-score values for pattern supports requires the value of standard deviation. Suppose, threshold specified is 0.15. This means that the dissimilarity value may not exceed 15% or it must be at least 85% similar. The deviation value for computing z-score is 0.3721. This value is obtained using the equation (9). Similarly, the threshold for normalized space is also to be computed. This value can be obtained using equation (10) after obtaining the standard deviation. For the example considered, normalized threshold is 0.08. We now show computation of z-scores for support sequences of temporal patterns in section 4.1.

4.1 Z-Score and Normalized Probability Computations

The Z-score value is obtained for support sequence of every temporal pattern. Z-score values are obtained using equation (1). The computation of Z-score and normalized probability from Z-score for all temporal patterns are shown below.

Z-Score for Temporal Pattern: [A]

$Z_A = \left[\frac{0.6-0.4}{0.3121}, \frac{0.5-0.6}{0.3121}\right]$ = [0.54, -0.27]

$P(Z_A)$ = [0.2054, 0.1064]

Z-Score for Temporal Pattern: [B]

$Z_B = \left[\frac{0.6-0.4}{0.3121}, \frac{0.9-0.6}{0.3121}\right]$ = [0.54, 0.81]

$P(Z_B)$ = [0.2054, 0.2910]

Z-Score for Temporal Pattern: [C]

$Z_C = \left[\frac{0.9-0.4}{0.3121}, \frac{0.7-0.6}{0.3121}\right]$ = [1.34, 0.27]

$P(Z_C)$ = [0.4099, 0.1064]

Z-Score for Temporal Pattern: [AB]

$Z_{AB} = \left[\frac{0.4-0.4}{0.3121}, \frac{0.4-0.6}{0.3121}\right]$ = [0, -0.54]

$P(Z_{AB})$ = [0, 0.2054]

Z-Score for Temporal Pattern: [AC]

$Z_{AC} = \left[\frac{0.5-0.4}{0.3121}, \frac{0.4-0.6}{0.3121}\right]$ = [0.27, --0.54]

$P(Z_{AC})$ = [0.1064, 0.2054]

Z-Score for Temporal Pattern: [BC]

$$Z_{BC} = \left[\frac{0.6-0.4}{0.3121}, \frac{0.7-0.6}{0.3121}\right] = [0.54, 0.27]$$

$$P(Z_{AC}) = [0.2054, 0.1064]$$

Z-Score for Temporal Pattern: [ABC]

$$Z_{ABC} = \left[\frac{0.4-0.4}{0.3121}, \frac{0.4-0.6}{0.3121}\right] = [0, -0.54]$$

$$P(Z_{ABC}) = [0, 0.2054]$$

These values obtained are then used to derive similar temporal associations not exceeding the upper dissimilarity limit.

4.2 Dissimilarity Computation

The dissimilarity between temporal pattern and reference can be computed using the equation (11). This equation is obtained by substituting (8) in (7).

$$D_V^{TP} = \frac{\left(1 - \frac{\sum_{k=1}^{k=N} e^{-\left(\frac{P(Z_K)}{\sigma g}\right)^2}}{|k|}\right)}{2} \quad (11)$$

The variable $P(Z_K)$ in the equation (11) represents the normalized probability obtained from corresponding Z-score. The dissimilarity values computed applying the proposed measure are expressed in Table.1

Table.1 Dissimilarity Computations

Pattern	Dissimilarity	Similar
[A]	0.0853	×
[B]	0.1801	×
[C]	0.1953	×
[AB]	0.0657	✓
[AC]	0.0853	×
[BC]	0.0853	×
[ABC]	0.019	✓

Comparison of Table.1 and Fig.3 shows that the proposed approach yields the same set of similar association patterns as that computed using Euclidean measure and also overcomes the dis-advantage that the Euclidean distance do not hold good for z-score approach. The only similar patterns for this example are patterns [AB] and [ABC]. All other remaining temporal patterns are outliers and are dissimilar.

4.2 Completeness and Correctness

Our approach is complete and correct as it does not lose any valid patterns that are similar. This is seen from the values computed in Fig.3 and Table.1 using Euclidean and proposed measure, applying normal distribution concept to support sequences of temporal patterns. Also, the application of normal distribution makes the statistical validation. The similarity measure has tight higher limit (=1) and lower limit (=0). This makes the proposed measure feasible for comparing the similarity degree to the threshold specified.

4.3 Monotonicity

Our measure also satisfies monotonicity w.r.t $D^{UL\ distance}$. This property makes the measure feasible for pruning the invalid temporal patterns much before than the actual point of computation in the process of discovering valid similar patterns. The discussion on monotonicity is restricted in this paper and we concentrated more on the discussion of the current approach in detail. In the future, this research may be extended to address the monotonicity property of the current measure and using this property to reduce the computation cost for improving computational efficiency. In this paper, our main objective has been towards introducing the novel approach for retrieving similar temporal patterns. We call our method as N-Spamine.

5. Conclusions

Research towards mining association patterns from time stamped temporal databases is much understudied in the literature and has a huge scope for researchers working in temporal databases, data mining, spatial databases. Also, the generated data in an IoT environment is implicitly temporal which the future is. The research in this paper suggests two new research directions for discovering valid similarity profiled temporal association patterns which include, looking into the possibility of applying normal distribution concept to mine all valid and similar temporal association patterns and the design of a new similarity measure to time stamped temporal databases. The results show that the proposed method retrieves all valid similar temporal patterns.

6 Acknowledgments

I thank P.V.Kumar, retired professor of Computer Science and Engineering, Osmania University and V.Janaki without whose continuous support this research work would have not been possible. The authors also thank the management VNR VJIET, for providing research facilities in carrying out the present research work. I am also thankful to my beloved and late father Prof. Dr.Narasimhacharyulu who suggested me to think of working towards present problem solution addressed in this research.

References

[1] Manish Gupta, Jing Gao, Yizhou Sun, and Jiawei Han, " Integrating community matching and outlier detection for mining evolutionary community outliers", KDD'12, In the *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining 2012,* pp. 859-867.

[2] Marcos R. Vieira, Petko Bakalov, and Vassilis J. Tsotras, " On-line discovery of flock patterns in spatio-temporal data", In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (GIS '09). ACM, New York, NY, USA, pp. 286-295. DOI=http://dx.doi.org/10.1145/1653771.1653812.

[3] Fabrizio Angiulli, Stefano Basta, and Clara Pizzuti, "Detection and prediction of distance-based outliers", In *Proceedings of the 2005 ACM symposium on Applied computing* (SAC '05), pp.537-542.

[4] Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander," LOF: identifying density-based local outliers", In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data* , 2000, pp.93-104.

[5] Manish Gupta, Jing Gao, Yizhou Sun, and Jiawei Han, " Community trend outlier detection using soft temporal pattern mining", In *Proceedings of the 2012 European conference on Machine Learning and Knowledge Discovery in Databases - Volume Part II* (ECML PKDD'12), 2012, pp. 692-708.

[6] M. Gupta, J. Gao, C. C. Aggarwal and J. Han, "Outlier Detection for Temporal Data: A Survey," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 9, pp. 2250-2267, Sept. 2014.
doi: 10.1109/TKDE.2013.184.

[7] A. M. Rajeswari, G. V. Aishwarya, V. A. Nachammai and C. Deisy, "Temporal outlier detection on quantitative data using unexpectedness measure," *2012 12th International Conference on Intelligent Systems Design and Applications (ISDA)*, Kochi, 2012, pp. 420-424.

[8] D. Birant and A. Kut, "Spatio-temporal outlier detection in large databases," *28th International Conference on Information Technology Interfaces, 2006.*, Cavtat/Dubrovnik, 2006, pp. 179-184. doi: 10.1109/ITI.2006.1708474

[9] E. Cipolla and F. Vella, "Identification of Spatio-Temporal Outliers through Minimum Spanning Tree," Tenth International Conference On *Signal-Image Technology and Internet-Based Systems,2014,*pp.248-255.

[10]X. Li, Z. Li, J. Han and J. G. Lee, "Temporal Outlier Detection in Vehicle Traffic Data," proceedings of the *IEEE 25th International Conference on Data*

Engineering, Shanghai, 2009, pp. 1319-1322.

[11] Y. Hu, J. Xie, C. Wang and Z. Zhou, "An Algorithm for Mining Top K Influential Community Based Evolutionary Outliers in Temporal Dataset," *2013 IEEE 25th International Conference on Tools with Artificial Intelligence*, Herndon, VA, 2013, pp. 524-531.

[12]Chao He and Yu-feng Zhang, "Research on semantic association pattern mining model based on ontology," *3rd International Conference on Advanced Computer Theory and Engineering(ICACTE)*, Chengdu, 2010, pp.V1-497-V1-501.

[13]W. Ouyang, S. Luo and Q. Huang, "Discovery of Direct and Indirect Association Patterns in Large Transaction Databases," *International Conference on Computational Intelligence and Security, 2007* pp. 167-170.

[14] W. Ouyang and Q. Huang, "Mining Direct and Indirect Association Patterns with Multiple Minimum Supports," *International Conference on Computational Intelligence and Software Engineering,* 2010, pp. 1-4.

[15] S. Mehrabi *et al*., "Temporal Pattern and Association Discovery of Diagnosis Codes Using Deep Learning," *Healthcare Informatics (ICHI), 2015 International Conference on*, Dallas, TX, 2015, pp. 408-416.
doi: 10.1109/ICHI.2015.58.

[16] J. S. Yoo and S. Shekhar, "Similarity-Profiled Temporal Association Mining," in *IEEE Transactions on Knowledge and Data Engineering*, vol.21,no.8,pp.147-1161,Aug.2009.

[17]Vangipuram Radhakrishna, P. V. Kumar, and V. Janaki, " 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), Article 57 , 6 pages. DOI: http://dx.doi.org/10.1145/2832987.2833069.

[18] Oliver Brdiczka, Norman Makoto Su, and James Bo Begole, " Temporal task footprinting: identifying routine tasks by their temporal patterns", In *Proceedings of the 15th international conference on Intelligent user interfaces* , 2010, pp281-284.

[19] Yi-Cheng Chen, Ji-Chiang Jiang, Wen-Chih Peng, and Suh-Yin Lee, "An efficient algorithm for mining time interval-based patterns in large database", In *Proceedings of the 19th ACM international conference on Information and knowledge management* (CIKM '10). ACM, 49-58. DOI=http://dx.doi.org/10.1145/1871437.1871448

[20] V.Radhakrishna, P.V.Kumar, V.Janaki, "A Novel Similar Temporal System Call Pattern Mining for Efficient Intrusion Detection", Journal of Universal Computer Science, Volume 22, Issue 4, 2016, pp. 475-493.

[21]Tzung-Pei Hong, Guo-Cheng Lan, Ja-Hwung Su, Pei-Shan Wu, Shyue-Liang Wang, "Discovery of temporal association rules with hierarchical granular framework", Applied Computing and Informatics, Volume 12, Issue2, 2016, Pages 134-141.

[22]Jia Ke, Yongzhao Zhan, Xiaojun Chen, Manrong Wang, "The retrieval of motion event by associations of temporal frequent pattern growth", Future Generation Computer Systems, Volume 29, Issue 1, January 2013, pp. 442-450.

[23]Chun-Hao Chen, Guo-Cheng Lan, Tzung-Pei Hong, Shih-Bin Lin, "Mining fuzzy temporal association rules by item lifespans, Applied Soft Computing", Volume 41, April 2016, Pages 265-274, ISSN 1568-4946, http://dx.doi.org/10.1016/j.asoc.2016.01.008.

[24] Zheng Xu, Xiangfeng Luo, Shunxiang Zhang, Xiao Wei, Lin Mei, Chuanping Hu, "Mining temporal explicit and implicit semantic relations between entities using web search engines", Future Generation Computer Systems, Volume 37, July 2014, Pages 468-477, ISSN 0167-739X, http://dx.doi.org/10.1016/j.future.2013.09.027.

[25] Peng Chen, Beth Plale, Mehmet S. Aktas, "Temporal representation for mining scientific data provenance", Future Generation Computer Systems, Volume 36, July 2014, Pages 363-378, ISSN 0167-739X, http://dx.doi.org/10.1016/j.future.2013.09.032.

[26]Jianqin Yin, Guohui Tian, Zhiquan Feng, Jinping Li, "Human activity recognition based on multiple order temporal information", Computers & Electrical Engineering, Volume 40, Issue 5, July 2014, pp. 1538-1551.

[27]V.Radhakrishna, P.V.Kumar, V.Janaki, "A Novel Approach for Mining Similarity Profiled Temporal Association Patterns", REVISTA TECNICA DE LA FACULTAD DE INGENIERIA UNIVERSIDAD DEL ZULIA, Volume 38, Issue 3, Dec 2015, pp. 80-93.

[28]V.Radhakrishna, P.V.Kumar, V.Janaki, "An Efficient Approach to find Similar Temporal Association Patterns Performing Only Single Database Scan", REVISTA TECNICA DE LA FACULTAD DE INGENIERIA UNIVERSIDAD DEL ZULIA, Volume 39, Issue 1, Feb 2016, pp. 241-255.

[29]Vangipuram Radhakrishna, P. V. Kumar, and V. Janaki, " Mining Outlier Temporal Association Patterns", In *Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies* (ICTCS '16). ACM, New York, NY, USA, , Article 105 , 6 pages.

[30]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

[31]Vangipuram Radhakrishna, P. V. Kumar, and V. Janaki. 2015. 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 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 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.

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 served in the capacity of s 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 have been successfully awarded with Ph.D.

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 served in various positions as Associate professor, Professor at Kakatiya Institute of Technology and Science, Warangal and Principal at Ramappa Engineering College, Warangal, Telangana, India. 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 more than ten research scholars towards their research at various universities as supervisor and co-supervisor and also guides research scholars towards award of their Ph.D.