IADIS International Journal on Computer Science and Information Systems
Vol. 12, No. 1, pp. 45-60
ISSN: 1646-3692

DESIGN AND ANALYSIS OF SIMILARITY MEASURE FOR DISCOVERING SIMILARITY PROFILED TEMPORAL ASSOCIATION PATTERNS

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

ABSTRACT

A wide variety of real time applications generate temporal data. Determining and unearthing similar temporal association patterns is complex and challenging task when considering time stamped temporal databases. Previous works have considered only existing distance measure and did not address retrieval and discovery of similar temporal association patterns using new distance measures. In this paper, we design a new similarity measure which suits the temporal context that can be applied to obtain all valid STAP (similar temporal association patterns). Our approach considers approximating the association pattern support bounds and applying the proposed similarity measure. The results show the proposed approach has better computational complexity compared to other approaches and improved time efficiency.

KEYWORDS

Distance Function, Temporal Pattern, Similarity Degree, Association Rules, Support Sequence

1. INTRODUCTION

A wide number of applications exist in our everyday life which continuously originates temporal data. A temporal dataset may be viewed as a collection of temporal data objects. A database system can be modeled to consist of data objects showing temporal behavior for a wide range of applications. Such a database system may later be used 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 behavior. Some other examples include medical datasets, credit datasets, and co-authorship in DBLP (Manish Gupta, Jing Gao, Yizhou Sun, & Jaiwen Han, 2012). The details of authorship, co-authorship details for an author keep varying and updating over a period of time and shows temporal behavior. 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 behavior 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 (Manish Gupta et al., 2012; S. Mehrabi et al., 2015).

In this paper, we concentrate on discovery of similar temporal association patterns from time stamped temporal data defined over disjoint timeslots. An approach for estimating temporal pattern bounds is studied. This is then used to find and retrieve similar patterns from temporal database. We use similarity measure to find similarly varying temporal patterns.

2. RELATED WORKS

A community may be defined as collection of temporal objects grouped under one category through unsupervised learning process all those which behave similar and whose variations are also temporally similar. The topic of community objects, community outlier object detection and their identification from a given snapshot of temporal dataset is addressed in (Manish Gupta et al., 2012). Such knowledge discovery process for retrieving outliers is called Evolutionary Community Outlier Detection. Many times, the data information of spatial and spatio-temporal objects is stored as the spatio-temporal data. There exist many situations and instances where it is crucial to study the data object collocation behaviors. In many applications such as surveillance, identifying trajectory of similarly moving data objects is necessary such that the similarity of trajectory do not exceed certain threshold is most common operation. Queries fired applying data mining techniques such as, the identification of suspiciously moving people which considers moving data objects pattern, flocks of moving data objects and vehicle convoys. The flock discovery problem is studied in (Marcos R. Vieira, Petko Bakalov, & Vassilis J. Tsotras, 2009) and a generalized pattern discovery framework in stream data is proposed.

Addressing outlier detection and outlier prediction using distance computation approach (Fabrizio Angiulli et al., 2015), is carried out by Fabrizio Angiulli. The approach is an unsupervised machine learning approach which defines solving set, using which an incoming stream data object is judged whether if it is an outlier. Security applications are primarily concerned and require abnormal pattern discovery compared to normal pattern discovery. These abnormal patterns cannot be missed and may harm the applications and also the data associated with such systems. The response time of judging outliers is also important. This detection is still challenging when mining temporal data. A general view of researchers when studying and addressing outlier discovery is that it is treated as a binary decision problem that requires judging, if a given data object is outlier or not.

Markus M. Breunig et al. (2014), considers, “outlier degree” for each data object within a community group and defines LOF (local outlier factor). This factor defines if the considered data object is an outlier which behaves differently compared to neighboring data objects. A huge amount of temporal data is generated from various applications such as news feed data, sensors data, and bank transactions. Understanding such evolving data requires study and understanding of temporal behavior. This is addressed in Manish Gupta et al. (2012). All such objects are called community trend outliers.

Distributed data such as temporal distributed data, spatial sensor data, time series data, stream data, network data, and spatio-temporal data is another form of temporal data. In M. Gupta et al. (2014) the survey addressed is in the context of mining outliers in temporal data and includes discussion on specific challenges such as, Outlier detection from temporal data, Classification of temporal outlier analysis, various prediction models. A general framework is outlined for “community outlier detection” (M. Gupta et al., 2014).

In A. M. Rajeswari, G. V. Aishwarya, V. A. Nachammai, & C. Deisy (2012), TAR (Temporal Association Rule) discovery are used to figure out rare items and unexpected trends. Generally, outliers do not violate semantics but are infrequent. Stock splits represent sudden changes in the stock trend that constitutes outlier behavior. TAR generated are multidimensional, time series and quantitative in nature. Unexpectedness measure is used to prune ARs. Outlier Detection is achieved by finding stock splits.

In D. Birant & A. Kut (2006), outlier detection algorithm considers both non-spatial data objects and spatio-temporal data objects. The approach is a 3-step approach which involves performing unsupervised learning, determining nearest temporal and spatial neighbors.

Another approach for outlier detection uses Minimum spanning tree (MST) and TARs (E. Cipolla & F. Vella, 2014) generated from meteorological data. The importance and advantage of this approach is that damages mat be predicted well ahead usually caused due to abnormal climate conditions. MST is used is to achieve optimization.

Many applications usually consider only a single time instance and do not concentrate on multiple time instances. One application area that involves handling dynamic time is vehicular traffic data. Temporal Outlier discovery from vehicular traffic data is discussed in X. Li, Z. Li, J. Han & J. G. Lee (2009). Normally semantics are not considered in conventional text mining techniques. Association pattern mining (APM) is applied for ontology in (Chao He & Yu-feng Zhang, 2010) which considers semantic property.

Many situations exist where infrequent items can give us important insights into datasets. In W. Ouyang, S. Luo, & Q. Huang (2007) an approach for discovering both indirect and direct association patterns is proposed considering single support values. In W. Ouyang & Q. Huang (2010), this is extended by considering multiple minimum support values for discovering both direct and indirect association rules which also discusses rare item problem.

Deep learning techniques and temporal data mining techniques are applied to find TARs (S. Mehrabi *et al.*, 2015). J. S. Yoo & S. Shekhar (2009), consider temporal data (time stamped temporal data), to retrieve all temporal association patterns that are similar and satisfy given threshold. A dissimilarity measure is proposed in Vangipuram Radhakrishna et al. (2015) for predicting similar patterns. An approach for identifying routine tasks based on their temporal regularities is proposed called as "temporal task foot printing" is proposed (Oliver Brdiczka, Norman Makoto Su, & James Bo Begole, 2010).

Event data is mostly an interval based data. Given an "Event data" (Yi-Cheng Chen, Ji-Chiang Jiang, Wen-Chih Peng, & Suh-Yin Lee, 2010), identifying patterns from such event data is challenging because of complex relationships that exist in event based data. A new approach and representation to handle such complex relationships called the 'incision strategy", which uses the representation termed as the "coincidence representation ", is proposed in Yi-Cheng Chen et al. (2010). A scalable and efficient algorithm is proposed to discover frequent pattern from interval based event data called "CTMiner". An application of temporal pattern mining for Intrusion detection (V. Radhakrishna et al., 2016) discussed temporal support sequences. Itemset is frequent if its support satisfies threshold. There exist some item sets which are frequent only for a certain period and are infrequent for most of other period. TARs are generated organizing time into granules (Tzung-Pei Hong, Guo-Cheng Lan, Ja-Hwung Su, Pei-Shan Wu, & Shyue-Liang Wangl, 2016) which considers a hierarchical granular framework. 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.

Video databases are event based data. Generating ARs from such video databases is challenging which requires complex analysis. TFP-growth tree (Jia Ke, Yongzhao Zhan, Xiaojun Chen, & Manrong Wang, 2013) considers temporal sub events using five temporal relationships and extend FP-Tree algorithm for temporal domain using which ARs are generated from such video databases. For this they use cloud infrastructure (Jia Ke et al., 2013). Every item with reference to a temporal transaction database has its life time and transactions often contain quantitative values. In Chun-Hao Chen, Guo-Cheng Lan, Tzung-Pei Hong, & Shih-Bin Lin (2016), 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 (Zheng Xu, Xiangfeng Luo, Shunxiang Zhang, Xiao Wei, Lin Mei, & Chuanping Hual, 2014) is studied and extended for web search engine. Other recent related works includes Peng Chen, Beth Plale, & Mehmet S. Aktasl (2014); Jianqin Yin et al. (2014); V. Radhakrishna et al. (2015); Vangipuram et al. (2016). 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). Aravind Cheruvu & V. Radhakrishna (2016), Vangipuram Radhakrishna, P. V. Kumar & V. Janaki (2016) and Vangipuram Radhakrishna, P. V. Kumar & V. Janaki (2015) propose similarity measures and an approach for support bound estimation of temporal patterns.

3. SIMILARITY MEASURE

We discuss the proposed similarity measure for finding similar temporal association patterns in this section. The similarity measure uses the support value of temporal and reference pattern at a given time slot and computes dissimilarity value for that timeslot. The overall distance is a function of distance values at every timeslot.

3.1 Temporal Similarity Measure

Given a temporal association pattern, T_s and the reference support time sequence, R_s . We use the similarity measure proposed in this section to find the similarity degree between these two temporal patterns. The temporal pattern is represented as $T_s = \left(T_{s_1} ;\ T_{s_2} ;\ T_{s_3} \ldots\ldots\ldots..T_{s_m}\right)$ and the reference sequence is the vector of support values for all timeslots, $R_s = \ (R_{s_1} ;\ R_{s_2} ;\ R_{s_3} \ldots\ldots\ldots..R_{s_m})$. Then, the fuzzy membership value considering temporal patterns T_s and R_s for k^th^ timeslot is determined using the equation (1)

$$f_k^{T,R} = \ exp\left[-1 * (\frac{T_{s_k} - R_{s_k}}{\sigma_g})^2\right] \tag{1}$$

Considering all time slots, the membership value between temporal and reference pattern is given by the equation (2)

$$f_{T,R}^{avg} = \frac{\sum_{k=1}^{k=m} f_k^{T,\ R}}{|k|} \tag{2}$$

The fuzzy similarity between the two patterns is hence given by equation (3)

$$Sim(T,R) = \ f_{T,R}^{avg} \tag{3}$$

The true dissimilarity between two temporal patterns is hence given by equation (4)

$$dSim(T,R) = 1 - \ f_{T,R}^{avg} \tag{4}$$

3.2 Threshold and Deviation

Let the threshold specified in Euclidean space is denoted using the symbol 'δ '. The value for the deviation (denoted by 'σ_g') is thus given by the equation (5). This value of deviation is actually used in equations (1) and (2) when computing distance value for a particular time slot. In the present context, we estimate the deviation using user defined threshold value.

$$\sigma_g = \frac{\delta}{\sqrt{\ln_e(\frac{1}{1-0.3679*\delta})}} \tag{5}$$

The threshold in transformed domain is given by equation (6) which is used to compare the distance value computed by using equation (4).

$$\delta^g = 1 - exp\left[-1 * (\frac{\delta}{\sigma_g})^2\right] \quad (6)$$

4. DERIVATIONS

Sections 4.1 and 4.2 give the proof for deviation and threshold expressions.

4.1 Standard Deviation

The fuzzy function for estimating similarity between patterns at k[th] timeslot is given by equation (7),

$$f_k^{T,R} = exp\left[-1 * (\frac{T_{s_k} - R_{s_k}}{\sigma_g})^2\right] \quad (7)$$

The values of $T_{s_k} and\ R_{s_k}$ cannot exceed 1. The minimum value is 0. So, we have two cases,

Case-1: Best Case, $T_{s_k} - R_{s_k}$ = 0

In this case, the fuzzy similarity defined by $f_k^{T,R} = exp\left[-1 * (\frac{T_{s_k} - R_{s_k}}{\sigma_g})^2\right]$ is equal to 1. This is the maximum possible value for similarity.

Case-2: Worst Case, $T_{s_k} - R_{s_k}$ = 1

In this case, the fuzzy similarity defined by $f_k^{T,R} = e^{-(\frac{T_{s_k} - R_{s_k}}{\sigma_g})^2} = e^{-(\frac{1}{\sigma_g})^2}$. The value of standard deviation is between 0 and 1. For σ_g = 0, $f_k^{T,R} = 0$ and For σ_g = 1, $f_k^{T,R} = 0.3679$. The values of $f_k^{T,R}$ are between (0, 0.3679) for $T_{s_k} - R_{s_k}$ = 1.

$$\text{Hence, } f_k^{T,R} = \begin{cases} 1\ ;\ T_{s_k} - R_{s_k} = 0 \\ 0;\ T_{s_k} - R_{s_k} = 1\ \ and\ \ \sigma_g = 0 \\ 0.3679;\ T_{s_k} - R_{s_k} = 1\ \ and\ \ \sigma_g = 1 \end{cases} \quad (8)$$

Let, δ is the allowed distance between patterns in the Euclidean space. The difference between distance bounds using Euclidean measure is 1 while using the proposed fuzzy membership function is 0.3679. i.e

$$\delta_{euc}^{bound} = \delta = 1 \quad (9)$$

and

$$\delta_{dSim(T,R)}^{bound} = \delta^g = 0.3679 \quad (10)$$

Dividing equation (9) by (10) on both sides, we get

$$\frac{\delta_{dSim(T,R)}^{bound}}{\delta_{euc}^{bound}} = \frac{0.3679}{1} \xrightarrow{yields} \delta_{dSim(T,R)}^{bound} = 0.3679 * \delta_{euc}^{bound} \tag{11}$$

Using equation (11), and equating distance computations in Euclidean and fuzzy space,

$$1 - exp\left[-1 * (\frac{\delta}{\sigma_g})^2\right] = 0.3679 * \delta \tag{12}$$

This reduces to

$$(\frac{\delta}{\sigma_g})^2 = \ln_e(\frac{1}{1 - 0.3679 * \delta}) \tag{13}$$

Finally, we get the deviation which is defined using equation (14)

$$\sigma_g = \frac{\delta}{\sqrt{\ln_e(\frac{1}{1 - 0.3679 * \delta})}} \tag{14}$$

4.2 Threshold

The dissimilarity value that must be considered for validation of similarity of temporal patterns is given by equation (15) for fuzzy space. Here the notation δ is the allowed distance between patterns in the Euclidean space. For this value, we compute the corresponding dissimilarity limit using equation (15)

$$d_{Sim}^{(T, R)} = \delta^g = 1 - exp\left[-1 * (\frac{\delta}{\sigma_g})^2\right] \tag{15}$$

5. COMPUTING PATTERN AND DISTANCE BOUNDS

Sections 5.1 and 5.2 give the expression for bounds of support values of temporal association pattern combinations generated. Section 5.1 gives the expressions for estimation of maximum and minimum possible support values at k^{th} time slot for association pattern of size, |S|=2 and Section 5.2 gives expressions for estimation of maximum and minimum possible support values of association pattern at k^{th} time slot for size, |S|>2.

5.1 Pattern Size, |S| =2

Assume P_{sk} and Q_{sk} are support values at k^{th} 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 equations (16) and (17)

$$[P_{sk}Q_{sk}]_{max} = P_{sk} - maximum\{(1 - P'_{sk} - Q_{sk}), 0\} \quad (16)$$

$$[P_{sk}Q_{sk}]_{min} = maximum\{(1 - P'_{sk} - Q_{sk}), 0\} \quad (17)$$

5.2 Pattern Size, |S| >2

In this case, P_{sk} denotes |S|-1 pattern size support values at k^th time slot and Q_{sk} are singleton pattern values at k^th 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 (18) and (19). Here, $P_{sk}$and Q_{sk} denotes all possible pattern combinations of size, |S|-1 and 1 respectively.

$$[P_{sk}Q_{sk}]_{max} = P_{sk} - maximum\{(1 - P'_{sk} - Q_{sk}), 0\} \quad (18)$$

$$[P_{sk}Q_{sk}]_{min} = maximum\{(1 - P'_{sk} - Q'_{sk}), 0\} \quad (19)$$

The resultant maximum and minimum values for association pattern, $[P_{sk}Q_{sk}]$ is obtained using equations (20) and (21)

$$[P_{sk}Q_{sk}]_{max} = maximum\left\{[P_{sk}{}^{c1}Q_{sk}]_{max}, [P_{sk}{}^{c2}Q_{sk}]_{max}, \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, [P_{sk}{}^{cl}Q_{sk}]_{max}\right\} \forall k \quad (20)$$

$$[P_{sk}Q_{sk}]_{min} = maximum\left\{[P_{sk}{}^{c1}Q_{sk}]_{min}, [P_{sk}{}^{c2}Q_{sk}]_{min}, \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, [P_{sk}{}^{cl}Q_{sk}]_{min}\right\} \forall k \quad (21)$$

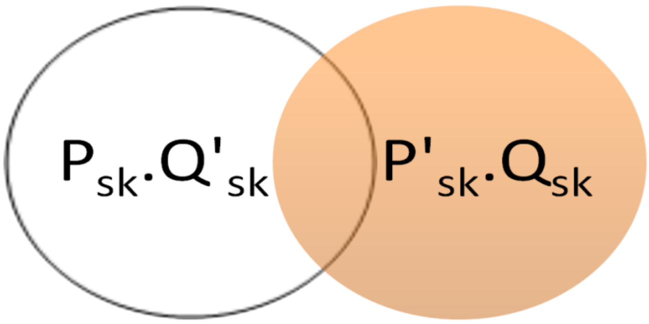

Figure 1. Venn Diagram Representation of Temporal Association Pattern

When estimating pattern with size, |S|>2, we compute true support values at (k-1)^th time slot and use these values to compute support value at k^th time slot. True support values are computed only when required.

5.3 Similarity Bound Computations

5.3.1 Upper-Lower Dissimilarity Bound

Let the maximum support sequence of temporal association pattern is denoted as $U_s = \left(U_{s_1}\ ;\ U_{s_2}\ ;\ U_{s_3}\ \ldots\ldots\ldots..U_{s_m}\right)$ and $R_s = \left(R_{s_1}\ ;\ R_{s_2}\ ;\ R_{s_3}\ \ldots\ldots\ldots..R_{s_m}\right)$ is the reference support sequence. The distance computation at k[th] timeslot is given by (22)

$$\mathrm{Udissim(T,R)}^{\,k} = \begin{cases} 0 & ;\ R_{s_k} \le U_{s_k} \\ 1 - exp\left[-1 * \left(\frac{R_{s_k} - U_{s_k}}{\sigma_g}\right)^2\right] & ;\ R_{s_k} > U_{s_k} \end{cases} \tag{22}$$

The upper-lower bound distance is hence given by

$$\mathrm{Udissim(T,R)} = \frac{\sum_{k=1}^{k=m} \mathrm{USim(T,R)}^{k}}{m} \tag{23}$$

5.3.2 Lower-Lower Dissimilarity Bound

Let the maximum support sequence of temporal association pattern is denoted as $L_s = \left(L_{s_1}\ ;\ L_{s_2}\ ;\ L_{s_3}\ \ldots\ldots\ldots..,L_{s_m}\right)$ and $R_s = \left(R_{s_1}\ ;\ R_{s_2}\ ;\ R_{s_3}\ \ldots\ldots\ldots..,R_{s_m}\right)$ is the reference support sequence. The lower-lower distance computation at k[th] timeslot is given by

$$\mathrm{Ldissim(T,R)}^{k} = \begin{cases} 0 & ;\ R_{s_k} \ge L_{s_k} \\ 1 - exp\left[-1 * \left(\frac{R_{s_k} - U_{s_k}}{\sigma_g}\right)^2\right] & ;\ R_{s_k} < L_{s_k} \end{cases} \tag{24}$$

The lower-lower bound distance is hence given by

$$\mathrm{Ldissim(T,R)} = \frac{\sum_{k=1}^{k=m} \mathrm{Ldissim(T,R)}^{\,k}}{m} \tag{25}$$

5.3.3 Lower Bound

The lower bound distance is

$$\text{Lower bound, LBdist (T, R)} = \mathrm{Ldissim(T,R)} + \mathrm{Udissim(T,R)} \tag{26}$$

6. ALGORITHM FOR TEMPORAL ASSOCIATION PATTERN MINING

6.1 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 (5) and (6).

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 $\mathrm{Udissim(T, R)}$. Retain or prune it based on $\mathrm{Udissim(T, R)}$. 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

If $\mathrm{LBdist\ (T,\ R)}$ exceeds δ^g, then T_s is not similar. If the distance bound, $\mathrm{LBdist\ (T, R)} \leq \delta^g$, the pattern may be similar, so compute its true support sequence. For this true support sequence of temporal pattern, obtain $\mathrm{LBdist\ (T, R)}$. If the distance, $\mathrm{LBdist\ (T, R)} \leq \delta^g$ pattern is similar. Otherwise it is not similar. When pattern is not similar, compare the distance, $\mathrm{Udissim(T, R)}$ 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

6.2 Results

In this section, we compare the results obtained using three approaches naïve, sequential and G-Spamine algorithm of our previous work (Shadi A. Aljawarneh, Vangipuram Radhakrishna, P.V.Kumar, V. Janaki, 2017 and Vangipuram et al., 2016) but using proposed measure. To apply G-Spamine, we use the distance measure proposed in section-3. Sequential approach discussed in J. S. Yoo et al. (2009), is used to compare our approach along with naïve approach. The graph in Figure 2 compares G-Spamine, Sequential and naïve approach for different threshold values.

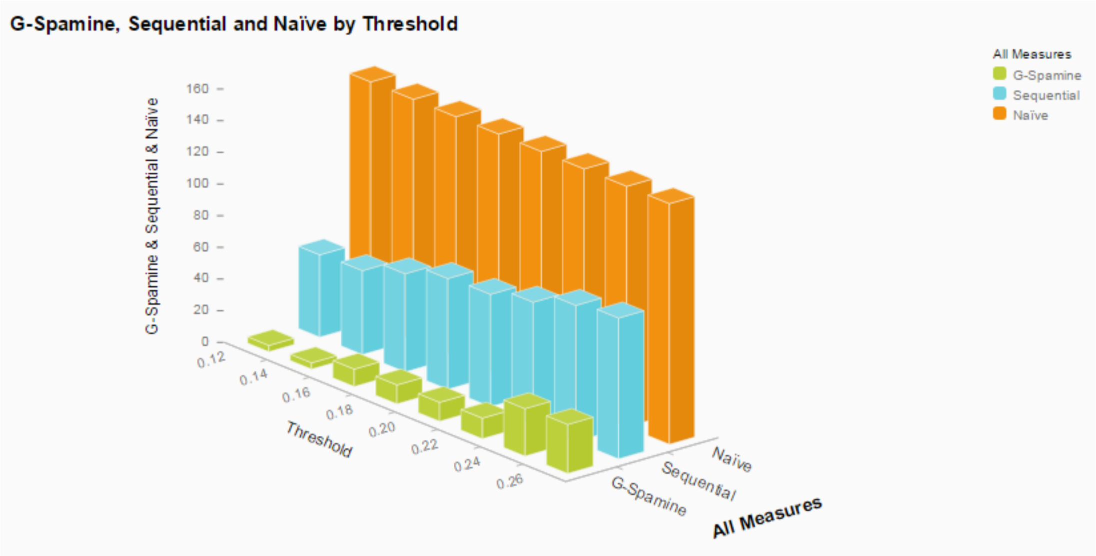

Figure 2. Varying Thresholds

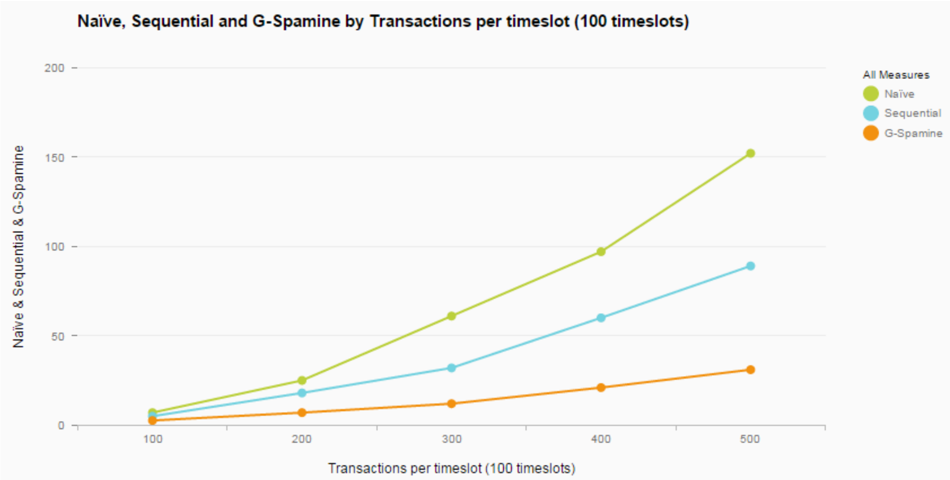

Figure 3. Varying Number of Transactions Per Time Slot

Figure 3 compares G-Spamine, Sequential and naïve approach for varying number of transactions per time slot. Similarly, Figure 4 depicts comparison of G-Spamine, Sequential and naïve approaches for varying number of time slots for a constant threshold equal to 0.26. The time slots are varied considering 100, 200, 300, 400 and 500 time slots each time and each time slot contains 500 transactions per time slot.

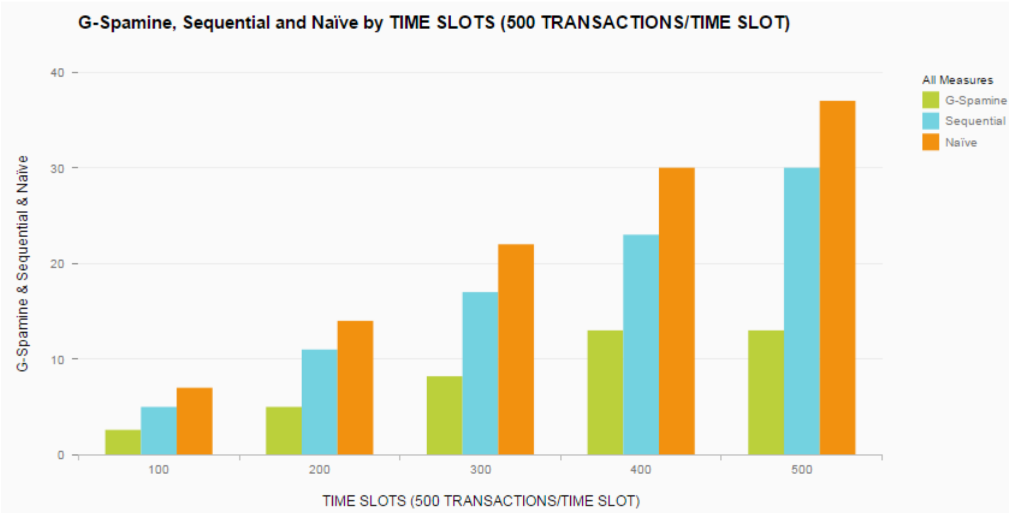

Figure 4. Varying Number of Timeslots (500 T/TS)

Figure 5, shows the graph obtained by considering different thresholds such as 0.12, 0.14, 0.16, 0.18, 0.2, 0.22, 0.24, and 0.26. The number of timeslots is 100 and the number of transactions per time slot is 500 for varying thresholds. The number of items considered is 12. The execution time of our approach is seen to be comparatively more optimal to other two approaches

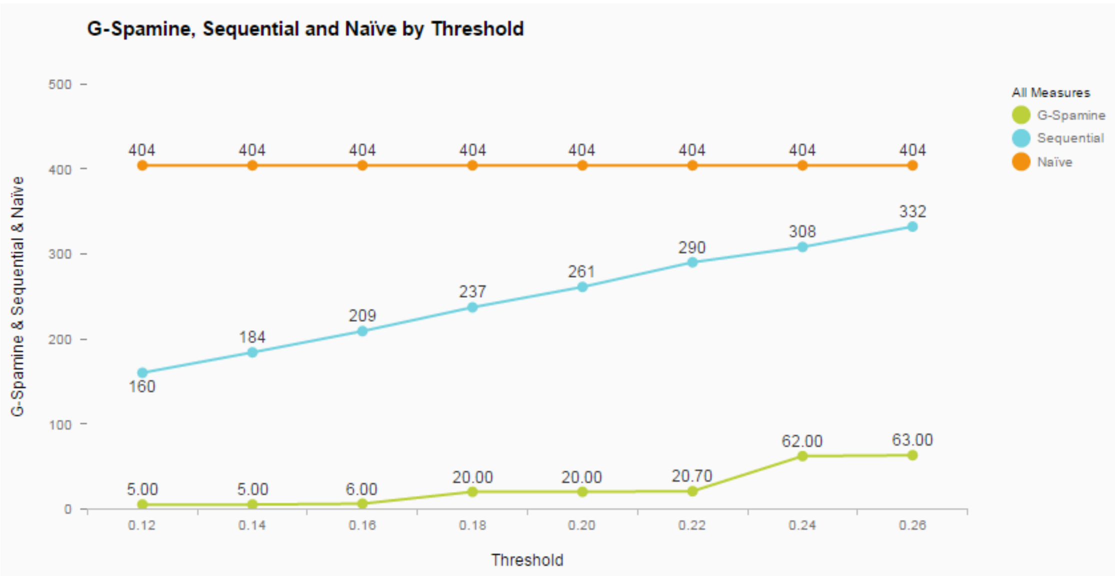

Figure 5. Varying Thresholds (12 Items, 500 T/TS, 100 TS)

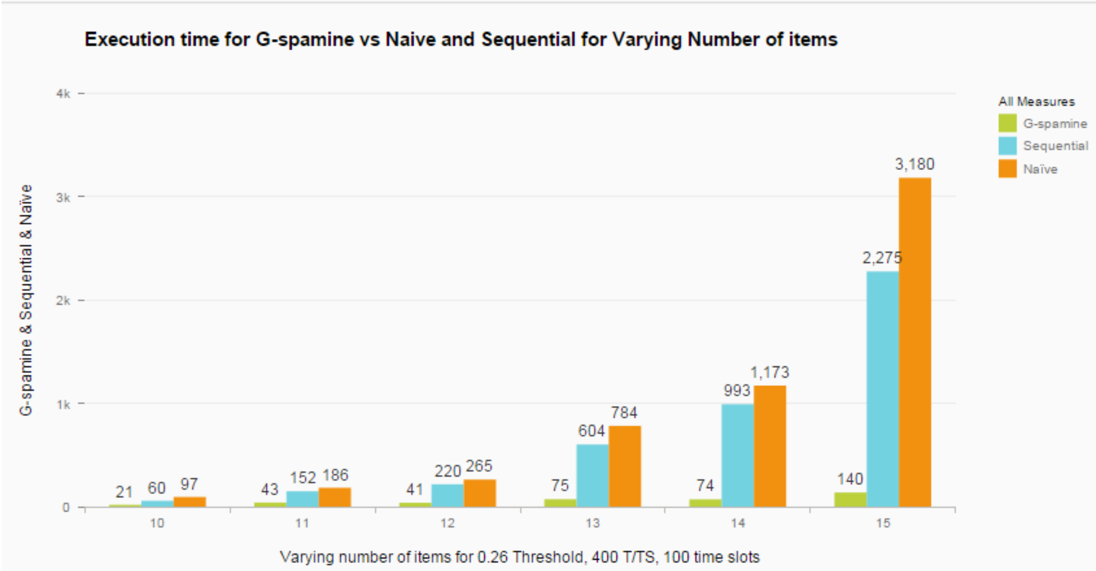

Figure 6. Varying Number of Items (0.26 Threshold, 400 T/TS, 100 TS)

Figure 6, shows the graph obtained by considering the threshold equal to 0.26 with the number of timeslots as 100 and the number of transactions per time slot are 400 for varying number of items. The number of items considered is 10, 11, 12, 13, 14 and 15. The execution time of our approach is comparatively more optimal to other two approaches for finding similarly varying temporal association patterns. The number of true support computations computed using G-Spamine is less when compared to other approaches and this makes the execution time better.

7. CONCLUSIONS

In this paper, we discuss the scope and need for new similarity measure which is followed by the design of proposed measure. Various research contributions in temporal data mining are also discussed. The dissimilarity measure helps in early pruning through upper-lower dissimilarity computation and lower bound distance computation. For this, we obtain the maximum and minimum possible support sequence of temporal association pattern and use the same for obtaining the distance bounds. The measure also satisfies monotonicity. This is an added advantage of the measure as it facilitates to perform early pruning. Results obtained show and prove the improvement achieved compared to other approaches. There is a scope for research to apply the statistical normal distribution to discover temporal patterns. In future, we aim to design temporal similarity measures applying normal distribution concept.

REFERENCES

A. M. Rajeswari, G. V. Aishwarya, V. A. Nachammai, & C. Deisy. (2012). Temporal outlier detection on quantitative data using unexpectedness measure. *Proceedings of the 12th International Conference on Intelligent Systems Design and Applications (ISDA), Kochi*, 420-424. doi: 10.1109/ISDA.2012.6416575

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 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

Chao He & Yu-feng Zhang. (2010). Research on semantic association pattern mining model based on ontology. *Proceedings of the 3rd International Conference on Advanced Computer Theory and Engineering(ICACTE)*, Chengdu, *1*, 497-501. doi: 10.1109/ICACTE.2010.5578967

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.

D. Birant & A. Kut. (2006). Spatio-temporal outlier detection in large databases. *Proceedings of the 28th International Conference on Information Technology in Interfaces*, 179-184.doi: 10.1109/ITI.2006.1708474

E. Cipolla & F. Vella, 2014. Identification of Spatio-Temporal Outliers through Minimum Spanning Tree. *Proceedings of Tenth International Conference On Signal-Image Technology and Internet-Based Systems,*248-255. doi: 10.1109/SITIS.2014.25

Fabrizio Angiulli, Stefano Basta, and Clara Pizzuti. (2005). Detection and prediction of distance-based outliers. *Proceedings of the 2005 ACM symposium on Applied computing (SAC '05), 18*(2), 537-542. doi: 10.1109/TKDE.2006.29

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

Jia Ke, Yongzhao Zhan, Xiaojun Chen, & Manrong Wang. (2013). The retrieval of motion event by associations of temporal frequent pattern growth. *Future Generation Computer Systems, 29*(1), 442-450. http://dx.doi.org/10.1016/j.future.2011.06.004

Jianqin Yin, Guohui Tian, Zhiquan Feng, Jinping Li. (2014). Human activity recognition based on multiple order temporal information. *Computers & Electrical Engineering, 40*(5),1538-1551.

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

Manish Gupta, Jing Gao, Yizhou Sun, & Jiawei Han. (2012). Integrating community matching and outlier detection for mining evolutionary community outliers. *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,*859-867. doi: 10.1145/2339530.2339667

Marcos R. Vieira, Petko Bakalov, & Vassilis J. Tsotras. (2009). On-line discovery of flock patterns in spatio-temporal data. *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS '09),* 286-295. doi:10.1145/1653771.1653812

Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, & Jörg Sander. (2000). LOF: identifying density-based local outliers. *Proceedings of the ACM SIGMOD international conference on Management of data*, 93-104. doi:10.1145/342009.335388

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 Bo Begole. (2010). Temporal task footprinting: identifying routine tasks by their temporal patterns. *Proceedings of the 15th international conference on Intelligent user interfaces*,281-284.

Peng Chen, Beth Plale, Mehmet S. Aktas. (2014).Temporal representation for mining scientific data provenance. *Future generation computer systems, 36*, 363-378.

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, Sunghwan Sohn, Dingheng Li, Joshua J. Pankratz, Terry Therneau, Jennifer, Hong fang Liu, Mathew Palakal. (2015). Temporal pattern and association discovery of diagnosis codes using deep learning. *Proceedings of the International Conference on Healthcare Informatics*, 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,* Retrieved from http://dx.doi.org/10.1016/j.future.2017.01.013

Tzung-Pei Hong, Guo-Cheng Lan, Ja-Hwung Su, Pei-Shan Wu, Shyue-Liang Wang. (2016). Discovery of temporal association rules with hierarchical granular framework. *Applied Computing and Informatics, 12*(2), pp.134-141.

Vangipuram 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 & 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

Vangipuram 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

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 and MIS,*1-6.doi: 10.1109/ICEMIS.2016.7745353.

Vangipuram Radhakrishna, P. V. Kumar, and V. Janaki, 2015, A novel approach for mining similarity profiled temporal association patterns using venn diagrams. *Proceedings of the the International Conference on Engineering & MIS*. 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 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). A survey on temporal databases and data mining. *Proceedings of the International Conference on Engineering & MIS (ICEMIS),* 1-5. doi:10.1145/2832987.2833064

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). 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

X. Li, Z. Li, J. Han & J. G. Lee. (2009). Temporal outlier detection in vehicle traffic data. *Paper presented at the 25th International Conference on Data Engineering,* 1319-1322. Abstract retrieved from http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4812530&isnumber=4812372

Y. Hu, J. Xie, C. Wang and Z. Zhou, 2013, An algorithm for mining top- k influential community based evolutionary outliers in temporal dataset, *Proceedings of the IEEE 25th International Conference on Tools with Artificial Intelligence*, Herndon, VA, pp. 524-531.

Yi-Cheng Chen, Ji-Chiang Jiang, Wen-Chih Peng, & Suh-Yin Lee. (2010). An efficient algorithm for mining time interval-based patterns in large database. *Proceedings of the 19th ACM international conference on Information and knowledge management*,49-58. doi: 10.1145/1871437.1871448

Zheng Xu, Xiangfeng Luo, Shunxiang Zhang, Xiao Wei, Lin Mei, & Chuanping Hu. (2014). Mining temporal explicit and implicit semantic relations between entities using web search engines. *Future Generation Computer Systems, 37*, 468-477.