International Refereed Journal of Engineering and Science (IRJES)
ISSN (Online) 2319-183X, (Print) 2319-1821
Volume 4, Issue 6 (June 2015), PP.153-156

Frequent Temporal Patterns Mining With Relative Intervals

B.M.G. Prasad[1], P. Naresh[2,] V. Veeresh[3]
[1,2,3](*Computer Science & Engineering, Holy Mary Institute of Technology and Science/JNTUH, Telangana, India)*

Abstract: *In this paper we have focused on the matter of discovering frequent temporal patterns in the database which contains temporal sequences, wherever a temporal sequence could be a set of things with associated dates as well as durations. Since the quantitative temporal info seems to be elementary in several contexts, it's taken under consideration within the mining processes and came back as a part of the extracted data. For this task we've got tailored the classical A Priori framework to propose an effective algorithmic rule supported by a hyper-cube illustration of temporal sequences. The extraction of quantitative temporal info is performed by employing a density estimation of the distribution of event intervals from the temporal sequences. Analyzing information sets shows that the projected algorithmic rule will robustly extract frequent temporal patterns with quantitative temporal extents.*
Keywords: *Temporal Patterns, A priori framework, hyper-cube representation*

I. INTRODUCTION

Generally the mining of frequent patterns considers the large databases which are having large number of sequences. Each and every sequence belongs to the dataset that is possible for duplication. Our mining task is to extract the longest patterns that occur frequently in the sequences which belong to the dataset in the database.
The FSPM is powerfully intended by its application to knowledge extraction from client purchase info DNA sequences, net logs medical statistic etc [1]. Most of those investigations have solely thought-about a linear ordered notion of your time. But, in some contexts, the quantitative temporal dimension of information provides elementary info that ought to be taken under consideration in the mining processes and came as a part of the extracted knowledge. As an example, in blog analysis [4], the browsing patterns will be refined by the time spent to explore the net pages.The comparison between the quantitative time spent on an online page and also the needed time to browse its content may be helpful to determine whether or not some a part of the knowledge has not been scan. In an exceedingly medical context, the temporal information is crucial to diagnose the patients accurately. This is obvious in things starting from diabetes to medical care unit cardiogram observation and artificial ventilation.

Unfortunately, discovering the quantitative temporal extents of events could be a tough drawback. The development of the sequence of things ought to be tangled with the development of representative durations [1][3]. On the one hand, temporal sequence instances are needed to construct the representative duration between 2 events in a very temporal sequence. But, on the opposite hand, the durations are needed to spot the temporal sequence instances. To tackle this drawback, we propose to boost the classical APriori algorithmic rule to handle quantitative temporal extents of the things [1][2].

1. Related Work:

As we considered, the proposed algorithm can efficiently mine the temporal patterns from the databases that is having the temporal sequences. Apart from the present work they can be mined by using the minimum and maximum time gaps [6] between the event datasets while retrieving from the databases. While plotting the graph between the accuracy and the temporal noise gap constraints and maximum gap values are considered.However, those algorithms based on minimum and maximum time gaps will not focus on the constraints of the time intervals. The gaps considered are used to avoid the patterns that are insignificant and also the rules that reduce the number of patterns and the rules that are generated.

II. TEMPORAL SEQUENCES AND TEMPORAL PATTERNS

2.1 Temporal Patterns:

A temporal pattern can be represented as the temporal sequence which can be treated as the representative of the set of equivalent sequences. More generally the patterns are the rules which mean that the term indicating the duration of period between the events [1]. Patterns provides delay period between the events and also the sessions of the particular event.

2.2 Temporal Pattern Extraction
Consider the database as an example which contains the following sequences.

$$\begin{aligned} s_1 &= (A,[1,3]),(B,[4,5]) \\ s_2 &= (A,[1.2,3.3]),(B,[3.8,4.6]),(C,[5.8,6.7]) \\ s_3 &= (A,[1.1,3.4]),(D,[1.5,5]),(B,[4.1,5.3]) \\ s_4 &= (A,[0.9,2.6]),(B,[4.1,5.2]),(C,[6.2,7.3]) \\ s_5 &= (A,[0.1,0.8]),(B,[4.2,5.2]) \end{aligned}$$

By considering the threshold as 0.5 several temporal patterns can be extracted based on the threshold. Some of the patterns may be of the following.

- $((A,[0.1,3.4]),(B,[3.8,5.3]))$ *strictly-covers all the examples, but the mean coverage is low (*≈ 0.36*).*
- $((A,[1.1,3.1]),(B,[4,5]))$ *has a better mean coverage (*≈ 0.71*), but does not strictly-cover* s_5.

Moreover, many temporal extents that strictly-cover the same set of temporal sequences may be computed for a temporal pattern. Our aim is to outline a singular and representative temporal extent for every interval of a temporal pattern [5]. Since, the distributions of temporal intervals will highlight such representative temporal extents, we have a tendency to propose an answer to construct distinctive and representative temporal extents which depends on the estimation of the interval distributions.

2.3 Temporal Sequences:
Temporal sequence is a set of temporal items, where the items are the non empty intervals between timestamps. A sequence can be denoted as [1]

$$\mathcal{S} = \{(s_i,[l_i,u_i])\}_{i\in\mathbb{N}_n}$$

2.4 Hyper Cube Representation:
Example temporal sequence denoted with S is given which is of the dimension n, then the temporal limits of that sequence S define a separate and unique hyper-cube representation in Rn [1][2]. Now temporal limits/bounds of that sequence item sets can be determined by using the orthogonal projection through each and every corresponding timeline.

Sequence:

$$\mathcal{S} = \{(s_i,[l_i,u_i])\}_{i\in\mathbb{N}_n}$$

2.5 Hypercube Representation:

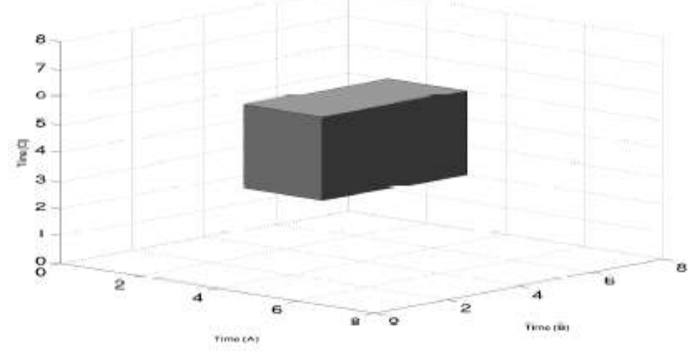

Figure-1: *Hypercube Representation*

III. FRAMEWORK AND ALGORITHMS

Our proposed framework is depended on the basic of APriori algorithm [1][7] in order to construct the lengthy frequent representational signatures. The main thought of this procedure is depended on the following assets: every representational signature which covers the frequent symbolic mark is repeated.These assets are useful in pruning lot of candidates having the size of n knowing the infrequent representational signatures which is of size firmly inferior to the term n because if it's representational signature then it is not treated as a frequent. The proposed algorithm has two stages,
Stage 1: Selecting the patterns that are frequent and having the dimension n.
Stage 2: Generating representational signature of dimension n+1.
The progression discontinues when it is noticed that there are no other potentially lengthy candidate item sets and that are not frequent [1][8].

ALGORITHM:

```
Input:     a: minimal support, S: temporal sequences
        databased: random points density
Output: Ba: the longest frequent temporal patterns (of
          dimension a)
II Construct frequent symbolic
    signatures of dimension 1 in Li
Li=construct_Ll(S,fmin);
a= 1;
while La is not empty do
    B. = 0;
    for p   L. do
        // Selection phase
         f= frequency(E);
        if f > /min then
            Ba = Bu SeleCtiOn(fmin, V.p, f, d);
        end
    end
    if B.   is emptythen
     |  return B._1;
    end
    // Generation phase
    La+i=generateCandidates(Ba);
    n= n + 1;
end
return Ba;
```

IV. EXPERIMENTS AND RESULTS

When plotting a graph between the Accuracy of the patterns mined and the temporal noise the below is the result indicating the accuracy vs temporal noise. The obtained results were ordered as follow: at first, we represent the evaluations of the assumed computation period and the obtained accuracy of the proposed algorithm based on the parameters.

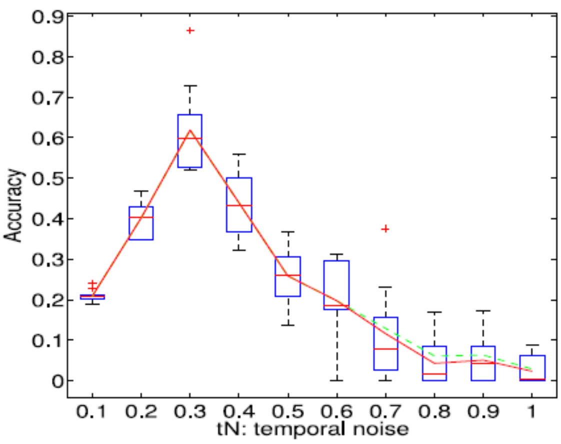

Figure-2: *Accuracy vs Temporal Noise Graph*

V. CONCLUSION

In the present paper we have introduced a temporal expansion of the standard sequential mining of patterns that adds quantitative temporal data to the patterns. By using the hyper-cube representation method for temporal sequences, the temporal extents of the pattern events can also be widely extracted. We have proposed an algorithm which solves the concern extraction and representational signature creation.

REFERENCES

[1]. Mining temporal patterns with quantitative intervals, Thomas Guyet, Ren´e Quiniou
[2]. R. Agrawal and R. Srikant. Mining sequential patterns. In Proceedings of the 11th International Conference on Data Engineering, 1995.
[3]. R. Bellman. Dynamic Programming. Princeton University Press, 1957.
[4]. C. Biernacki, G. Celeux, G. Govaert, and F. Langrognet. Model-based cluster and discriminant analysis with the MIXMOD software. Computational Statistics and Data Analysis, 51(2):587–600, 2006.
[5]. H. Bozdogan. Choosing the number of component clusters in the mixture-model using a new informational complexity criterion of the inverse-Fisher information matrix. In Proceedings 16th Annual Conference of the GfKl, pages 40–54, 1993.
[6]. Y.-L. Chen and T. C.-K. Huang. Discovering fuzzy timeinterval sequential patterns in sequence databases. IEEE Transaction on Systems, Man and Cybernetics, 35(5):959– 971, 2005.
[7]. B. Chiu, E. Keogh, and S. Lonardi. Probabilistic discovery of time series motifs. In Proceedings of the 9th International Conference on Knowledge Discovery and Data Mining,pages 493–498, 2003.
[8]. A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, 39:1–38, 1977.