ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 9, September 2016**

# A Survey on Incremental Temporal Mining using Incremental TP Miner and Incremental P-TP Miner algorithms

Sonali Rampure[1], Prof. R.V.Argiddi[2]

P.G Student, Department of Computer Science & Engineering, Walchand Institute of Technology ,Solapur University, Solapur, India[1]

Assistant Professor, Department of Computer Science & Engineering, Walchand Institute of Technology,Solapur University, Solapur, India[2]

**ABSTRACT**: Sequential pattern mining is special case of data mining. Most of sequential pattern mining algorithms works on static data which deals with the database should not change. But the databases in real world application do not have static data rather they have incremental database. There are some applications using time interval-based event data have used to discovering patterns from events that persist for some duration. There are two types of novel representations; endpoint and endtime are proposed to simplify the processing of complex relationships among event interval. Based on these proposed representations there are two types of interval-based patterns: temporal pattern and duration-probabilistic temporal pattern, are proposed.This paper attempts to providetwo algorithms Incremental Temporal Pattern Miner (TP-Miner) and Incremental Probabilistic Temporal Pattern Miner (P-TPMiner)to discover two types of interval-based sequential patterns.

**KEYWORDS**: Sequential pattern,Incremental Temporal Pattern Mining, Interval based pattern, Data mining.

## I. INTRODUCTION

In various real-world scenarios, some events which deals with periods of time rather than being instantaneous occurrences cannot be treated as “time points.” In such cases, the data are typically a sequence of events with both start and finish times. For example adopted sensor technology to monitor the electricity usage of all household appliances [10].Applying data mining techniques to many real world applications is a challenging task because the databases are dynamic. Incremental mining algorithms can significantly increase the speed of a task because much of the work that was performed for previous calls tasks can be reused in successive searches. These algorithms are most advantageous when the successive tasks that the incremental algorithm is run on are similar to previous tasks. In this training data is presented one at a time. Incremental mining algorithms for mining frequent patterns that use information collected during earlier mining process to cut down the cost of finding new pattern in whole database. Since mining every time the database grows, it is not efficient and hence the algorithm for incremental mining has to be proposed.

Incremental temporal mining which refers to the maintaining the discovered patterns over time in terms of more items are being added into the database. Because of the most applications append only nature of updating time-series data, incremental mining would be very effective and efficient. Non Incremental temporal mining which deals to those that need to process all items in each iteration of an iterative procedure for refining a final solution. In contrast of incremental algorithm, in non-Incremental mining algorithms all training data are presented simultaneously to the algorithms.

## II. RELATED WORK

In [1] authors used two algorithms (TPMiner and P-TPMiner) to discover temporal pattern, occurrence-probabilistic temporal pattern and duration-probabilistic temporal pattern using two novel representations endtime and endpoint representation.In [2] authors used novel algorithm, namely, Correlation Pattern Miner(CoPMiner), is used to capture

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 9, September 2016**

the usage patterns .With several new optimization techniques, CoPMiner can reduce the search space effectively and efficiently.[3] In this paper, a public data set REDD that is The Reference Energy Disaggregation Data Set which is real time dataset used for many data mining applications..This paper[10] deals with how appliances are used in the home and represents the power use of individual appliances, outperforms the other unsupervised disaggregation methods.In [12],Mining temporal patterns from time interval based data is a difficult problem since processing of complex relations among intervals may require generating and examining large amount of intermediate subsequence.

## III. PROPOSED ALGORITHM

Incremental mining algorithms increase the speed of a task because much of the work that was performed for previous calls tasks can be reused in successive searches.These two algorithms TPMiner and P-TPMiner are in restartable mode. Restartable mode is deals with the storing intermediate results.

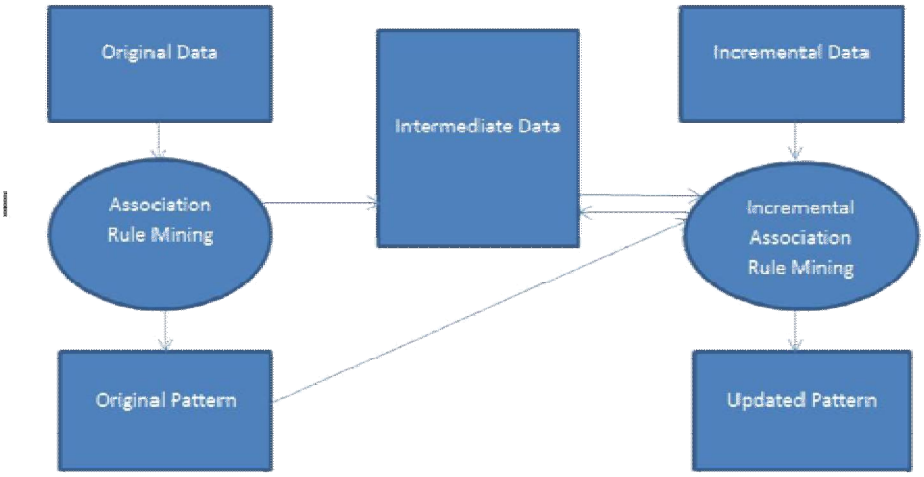


Fig1: Conceptual overview of proposed system

The proposed system fig1. shows that association rule mining process will apply on original data and generated data will stored in intermediate data. This association rule mining process will also produces original pattern.Incremental association rule mining process will apply on incremental data and generated data will stored in intermediate data. This Incremental association rule mining process will produces updated pattern based on available knowledge (obtained from mining of previously Intermediate stored data) and original pattern.The incremental mining algorithm uses incremental mining technique is to re-run the mining algorithm on the only updated database.

The proposed algorithm Incremental Temporal Pattern Miner (TPMiner) first transforms the temporal database into the endpoint representation and then scans the database to calculate the count of each endpoint concurrently. TPMiner removes infrequent endpoints below the given support threshold. For each frequent starting endpoint we build the projected database and call TPSpan recursively to discover sets of all temporal patterns.Save intermediate result to reuse it in incremental mining.In TPSpan, for a prefix it scans its projected database once to discover all local frequent endpoints and remove infrequent ones. Frequent endpoint can be appended to the original prefix to generate a new frequent sequence. A frequent endpoint deals with all endpoint sequence appear in pairs every starting (finishing) endpoint has a corresponding finishing (starting) endpoint,Output this frequent endpoint sequence. Calculated count support to remine previous count support. Finally, we construct the projected database with the frequently extended prefixes and recursively call TPSpan until the prefixes can no longer be extended.

The another proposed algorithm is Incremental Probabilistic Temporal (P-TPMiner) algorithm which uses probabilistic function calculation. For a prefix P-TPSpan first calls the procedure count support. In this to calculate the count of each endpoint concurrently. To generate a new frequent sequence frequent endpoint can be appended to the

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 9, September 2016**

original prefix.We also use the time information of frequent endpoint in DB to estimate the duration probability function. Then, the prefixes are extended by including extended duration probability function. If all endpoints in a frequent endpoint sequence appear in pairs, it deals with every starting (finishing) endpoint deals with corresponding finishing (starting) endpoint,Output this frequent endpoint sequence, including its duration probability function, as the duration-probabilistic temporal pattern. Finally, by constructing, the projected database we can discover all patterns with the frequently extended prefixes and by recursively running P-TPSpan until the prefixes can no longer be extended.

## IV. EXPECTED RESULTS

- Using Incremental Temporal Pattern Miner (TPMiner) generates the sequential pattern
- Using Incremental Probabilistic Temporal Pattern Miner (TPMiner) generates the sequential pattern.

## V. CONCLUSION

To discover sequential pattern both algorithms Incremental Temporal Pattern Miner (TPMiner) and Incremental Probabilistic Temporal Pattern Miner (TPMiner) are deals with incremental database with respect to time. These both algorithms work in restartable mode.We can apply proposed algorithms to real datasets to make the comparison of Incremental temporal mining and non-Incremental temporal mining.

## REFERENCES

1. Yi-Cheng Chen,Wen-ChihPeng and Suh-Yin Lee, "Mining Temporal Patterns in Time Interval Based Data" IEEE Transactions on Knowledge and Data Engineering, 1041-4347 (c) 2015 IEEE.
2. Chen,C.Chen,W.Peng and W. Lee, "Mining Correlation Patterns among Appliances in Smart Home Environment," IEEE 18th Pacific-Asia Conference in Knowledge Discovery and Data Mining, Advances in Knowledge Discovery and Data Mining(PAKDD'14), pp. 210-221, 2014.
3. J.Kolter, and M. Johnson, "REDD: A public data set for energy disaggregation research," KDD workshop on Data Mining Applications in Sustainability (SustKDD'11), pp. 1-6, 2011.
4. S. vanSchaik, D. Olteanu and R. Fink, "ENFrame: A Platform for Processing Probabilistic Data," The 17th International Conference on Extending Database Technology *(*EDBT'14*)*, pp. 355-366,2014.
5. Y. Li, J. Bailey, L. Kulik and J. Pei, "Mining Probabilistic Frequent Spatio-Temporal Sequential Patterns with Gap Constraints from Uncertain Databases," The 13th International Conference on Data Mining *(*ICDM'13*)*, pp. 448-457, 2013.
6. R.Sadasivam and K. Duraiswamy, "Efficient Method to Discover Interval-based Sequential Patterns," Journal of Computer Science, vol. 9, issue 2, pp. 225-234, 2013.
7. A.Wong, D. Zhuang, G. Li, and E. Lee, "Discovery of Closed Patterns and Noninduced Patterns from Sequences," IEEE Transactions on Knowledge and Data Engineering, vol.24, no. 8, pp.1408-1421, 2012.
8. A.Zakour, S. Maabout, M. Mosbah and M. Sistiaga, "Uncertainty Interval Temporal Sequences Extraction," International Conference on Information Systems Technology and Management (ICISTM' 12), pp. 259-270, 2012.
9. Z. Zhao, D. Yan and W. Ng, "Mining Probabilistically Frequent Sequential Patterns in Large Uncertain Databases," The 15th International Conference on Extending Database Technology*(*EDBT'12*)*, pp. 74-85, 2012.
10. H. Kim, M. Marwah, M. Arlitt, G. Lyon and J. Han, "Unsupervised disaggregation of low frequency power measurements," The 11th SIAM International Conference on Data Mining (SDM'11),pp. 747–758, 2011.
11. M. Muzammal and R. Raman, "Mining Sequential Patterns from Probabilistic Databases," The 15th Pacific-Asia Conference in Knowledge Discovery and Data Mining, Advances in Knowledge Discovery and Data Mining, (PAKDD'11), pp. 210-221, 2011.
12. Yi-Cheng Chen, Ji-Chiang Jiang, Wen-ChihPeng and Suh-Yin Lee,"An Efficient Algorithm for Mining Time Interval-based Patterns in Large Databases",(CIKM'10), October 26–30, 2010 2010 ACM.