ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Mining for Events Sequence Using Search Real-Time By GPS

Mr. Atheer Alaa Hammad
Master of Science in Computer Science,
Mahatma Gandhi College,
(Affiliated to Acharya Nagarjuna University)
N.G.O Colony Road, Guntur,
Andhra Pradesh-522006, India.

Vuyyuru Madhavi
Department of Computer Science and Engineering
Mahatma Gandhi College,
(Affiliated to Acharya Nagarjuna University)
N.G.O Colony Road, Guntur,
Andhra Pradesh-522006, India.

ABSTRACT

The actual data mining task is the semi-automatic or automatic analysis of large quantities of data to extract previously unknown, interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection), and dependencies (association rule mining, sequential pattern mining). This usually involves using database techniques such as spatial indices. These patterns can then be seen as a kind of summary of the input data, and may be used in further analysis or, for example, in machine learning and predictive analytics. For example, the data mining step might identify multiple groups in the data, which can then be used to obtain more accurate prediction results by a decision support system. Neither the data collection, data preparation, nor result interpretation and reporting is part of the data mining step, Episode Rule Mining is a popular framework for discovering sequential rules from event sequential data. However, traditional episode rule mining methods only tell that the consequent event is likely to happen within a given time interval after the occurrence of the antecedent events. As a result, they cannot satisfy the requirement of many time sensitive applications, such as program security trading and intelligent transportation management due to the lack of fine-grained response time. In this study, we come up with the concept of fixed gap episode to address this problem. A fixed-gap episode consists of an ordered set of events where the elapsed time between any two consecutive events is a constant.[1] Based on this concept, we formulate the problem of mining precise-positioning episode rules in which the occurrence time of each event in the consequent is clearly specified. In addition, we develop a trie-based data structure to mine such precise-positioning episode rules with several pruning strategies incorporated for improving the performance as well as reducing memory consumption. [2] Experimental results on real datasets show the superiority of our proposed algorithms. the goal for data mining is the extraction of patterns and knowledge from large amounts of data, not the extraction (mining) of data itself. It also is a buzzword and is frequently applied to any form of large-scale data or information processing (collection, extraction, warehousing, analysis, and statistics) with Java In this project we give concept to detect exact time of when next event falls, existing technology only tells when next event will occur but not give exact time. Suppose A, B, C, D, E are the events and after analysis we can say D event will occur only after the completion of A, B, C events and cannot say the exact time. Apart from activity diaries, event-based data is found in a wide range of application. Internet surfing records can be seen as a sequence of events (the web site visits), for example, the same can be said of working careers, historical events, personal and travel histories, medical records, work sampling records, purchase transactions, industrial process and system control records, records of attended university courses, among many others.[3] An important feature in this definition of event-based data is the continuity of the event occurrences which should be preserved, so when handling data where short events occur with long periods of inactivity in between an idle/empty event can be used for describing these periods. Furthermore, if the time and duration constraints are relaxed even other

Cite this article as: Mr. Atheer Alaa Hammad & Vuyyuru Madhavi, "Mining for Events Sequence Using Search Real-Time By GPS", International Journal & Magazine of Engineering, Technology, Management and Research, Volume 6 Issue 3, 2019, Page 14-19.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

data with an explicit ordering can be seen as event-sequences such as protein and DNA sequence data. In all cases a common factor of interest when analysing event-based data is the comparison of event-sequence records and the identification of interesting features as patterns within them.[4]

1. INTRODUCTION

Frequent episode mining (FEM) has emerged as a popular research topic in the data mining community. Given a single event sequence, FEM aims to identify all frequent episodes with the frequency larger than a given threshold.

Here an episode (also known as serial episode [3]) is a totally ordered set of events. One basic problem in FEM is to find episode rules from frequent episodes. Given a frequent episode α, a valid episode rule in the form of lhs −→ rhs can be generated in a straightforward manner: The antecedent lhs is the prefix of α and the consequent rhs is the last event in α, if its confidence is larger than a user-specified threshold. Figure 1 gives a running example for episode rule mining, where capital letters denote events and arabic numbers denote timestamps. In this sequence, we see three occurrences of episode D, A, B when maximum occurrence window size threshold is set to 4. Then, if we take B as the consequent, an episode rule D,A→B can be generated. This rule tells that it is within 2 time intervals after the occurrence of D, A that B will occur (with 100% probability). However, with such a rule discovered by traditional methods, we only know the approximate time range of the occurrence of rhs. In many real world applications, such rules are not practically useful without specifying the exact time of rhs. In this situation, we require the exact time to trigger the responses of the consequent automatically. Data mining is, as the term implies, to mine, to look for the ‘gold’ in vast amounts of data.

It is the process of extracting relationships and useful knowledge from large datasets which would otherwise remain hidden. In this paper, we first propose an enumeration based framework by first mining frequent minimal occurrence episodes and frequent fixed-gap episodes on the whole sequence, and then concatenating each pair of a minimal-occurrence episode and a fix-gap episode to generate a candidate of precise positioning rule. We further observed that for a precise positioning episode rule, the consequent must occur after the antecedent. Hence we can improve the proposed framework by mining fixed-gap episodes only after the occurrences of frequent minimal occurrence episodes. Along this route, we develop a trie-based framework to mine precise positioning episode rules directly based on frequent minimal-occurrence episodes partitioning.[5][6][7]

Objective of the Project

Event-based data are defined as data composed of sequences of ordered events, an event sequence each event or element of an event-sequence has a start time and duration and each begins when the previous one ends.

The types of event present in such a sequence all belong to a set of predefined event types, an event alphabet. An event-based dataset, D, then consists of a set of event-sequences Episode Rule Mining is a popular framework for discovering sequential rules from event sequential data. However, traditional episode rule mining methods only tell that the consequent event is likely to happen within a given time intervals after the occurrence of the antecedent events. As a result, they cannot satisfy the requirement of many time sensitive applications, such as program security trading due to the lack of fine-grained response time. In this study, we come up with the concept of fixed-gap episode to address this problem. A fixed-gap episode consists of an ordered set of events where the elapsed time between any two consecutive events is a constant. Based on this concept, we formulate the problem of mining precise-positioning episode rules in which the occurrence time of each event in the consequent is clearly specified. In addition, we develop a triebased data structure to mine such precise-positioning episode rules with several pruning strategies incorporated for improving the performance as well as reducing memory consumption. Experimental results on real datasets show the superiority of our proposed algorithms.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

2. LITERATURE REVIEW

Online Frequent Episode Mining

In this paper, we formulate the online frequent episode mining problem, which is especially useful to time-critical applications with growing sequences. We propose an efficient algorithm (named MESELO) for this problem. By the concept of last episode occurrence, MESELO can detect the minimal episode occurrences without performing a post-process checking. Also, utilizing the proposed episode trie, MESELO stores all the minimal episode occurrences in a compact way. Experiments on ten real data sets show the efficiency of the proposed algorithm, which is at least one magnitude of order faster than other baseline methods. In our future work we will consider attributes on events to identify more interesting episodes. Additionally, to further increase time efficiency and reduce memory consumption, we will develop the approximate method for online frequent episode mining based on the current exact solution.[8]

Mining Frequent Patterns without Candidate Generation

Our performance study shows that the method mines both short and long patterns efficiently in large databases, outperforming the current candidate pattern generation-based algorithms. The FP-growth method has also been implemented in the new version of DBMiner system and been tested in large industrial databases, such as in London Drugs databases, with satisfactory performance There are a lot of interesting research issues related to FP-tree-based mining, including further study and implementation of SQL-based, highly scalable FP-tree structure, constraint-based mining of frequent patterns using FP-trees, and the extension of the FP-tree-based mining method for mining sequential patterns, max-patterns, partial periodicity, and other interesting frequent patterns.[9]

Discovery of Frequent Episodes in Event Sequences

Sequences of events describing the behavior and actions of users or systems can be collected in several domains. An episode is a collection of events that occur relatively close to each other in a given partial order. We consider the problem of discovering frequently occurring episodes in a sequence. Once such episodes are known, one can produce rules for describing or predicting the behavior of the sequence. We give efficient algorithms for the discovery of all frequent episodes from a given class of episodes, and present detailed experimental results. The methods are in use in telecommunication alarm management [10]

3. SYSTEM ANALYSIS

Existing System

Frequent episode mining (FEM) has emerged as a popular research topic in the data mining community. Given a single event sequence, FEM aims to identify all frequent episodes with the frequency larger than a given threshold. One basic problem in FEM is to find episode rules from frequent episodes. Given a frequent episode α, a valid episode rule in the form of lhs → rhs can be generated in a straightforward manner: The antecedent lhs is the prefix of α and the consequent rhs is the last event in α, if its confidence is larger than a user-specified threshold. However, with such a rule discovered by traditional methods, we only know the approximate time range of the occurrence of rhs. In many real world applications, such rules are not practically useful without specifying the exact time of rhs. In this situation, we require the exact time to trigger the responses of the consequent automatically

Disadvantages of Existing System:

Traditional episode rule mining methods only tell that the consequent event is likely to happen within a given time intervals after the occurrence of the antecedent events. As a result, they cannot satisfy the requirement of many time sensitive applications, such as program security trading due to the lack of fine-grained response time.

Proposed System

In this paper, we formulate the problem of mining precise positioning episode rules (MIPER), which is helpful for applications where automatic responses are needed in a timely manner. We propose one enumeration approach

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

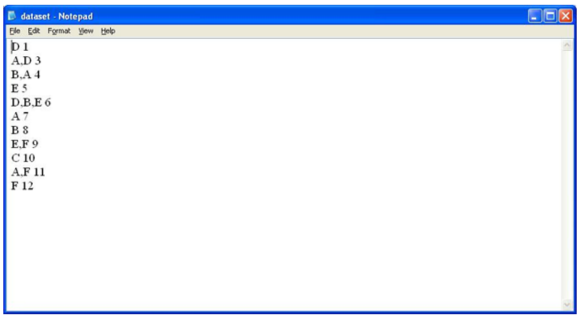
for MIPER and further devise two approaches based on a compact trie structure to enhance the pruning power as well as reduce execution time of the mining process.

Advantages of Proposed System:

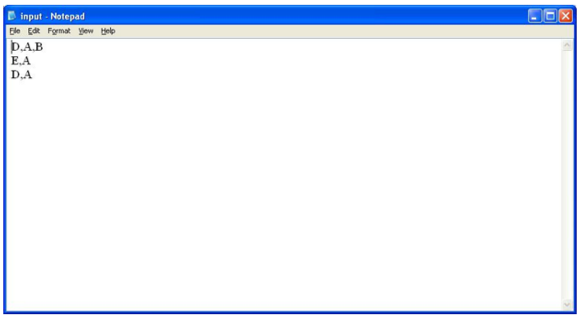
We address the new problem of mining precise positioning episode rules to satisfy the requirement of time-sensitive applications in the real world. We can improve the proposed framework by mining fixed-gap episodes only after the occurrences of frequent minimal-occurrence episodes.

4. OUTPUT SCREENS:

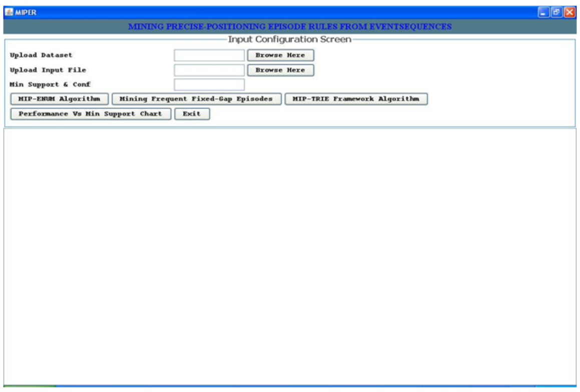
Below is dataset file with events



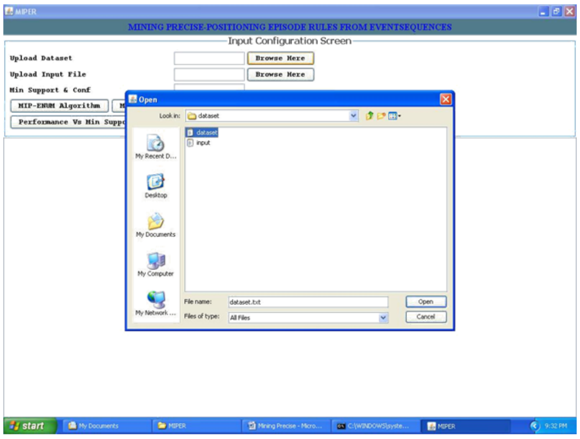
Below is input file which contains query to identify when those events will fall



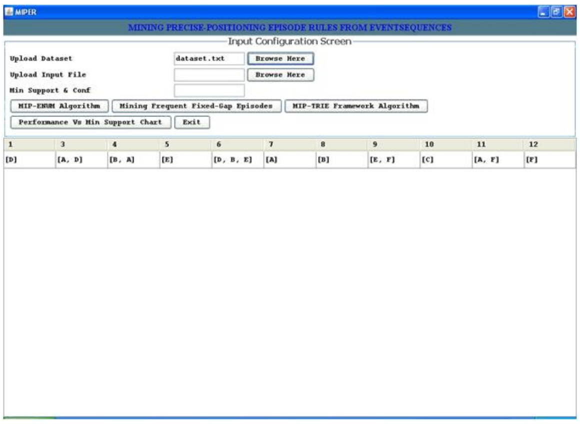
From above input we can detect when EA will occur



Click on 'Browse Here' button and upload dataset file



After dataset upload will get below screen

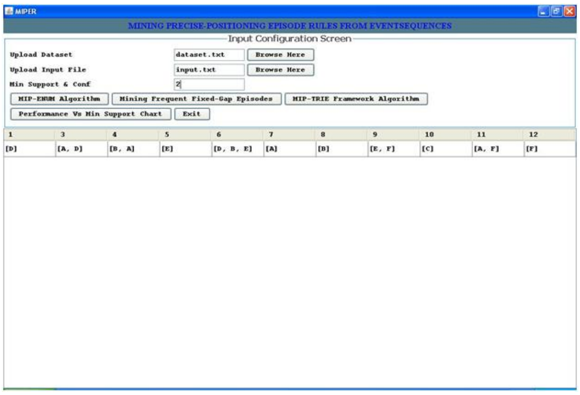


In above table all character values in row refers to events and column header values like 1, 2, 3 etc refers to time period

Now click on second 'Browse Here' button and upload input file

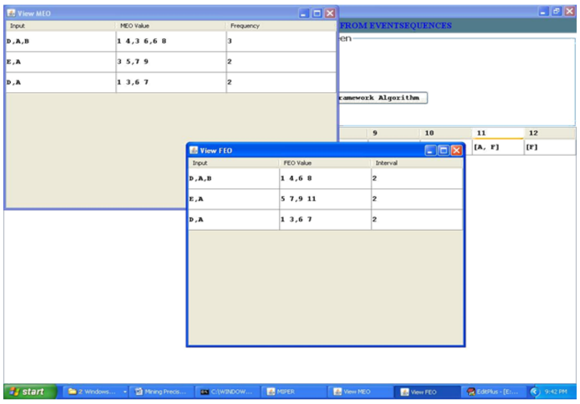
ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

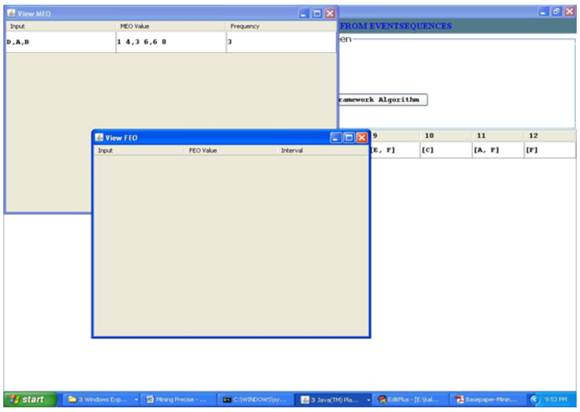
Now enter user define threshold values to detect events which are occurring frequently



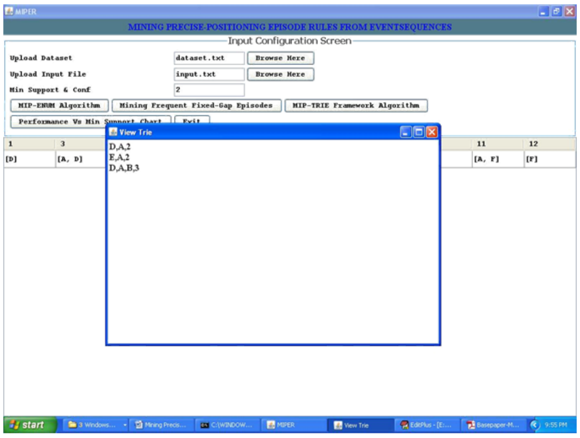
In above screen I enter min support threshold value as 2 which means I want only those events which occur more than or equal to 2 .

Now click on ‘Mip-Enum’ algorithm button to detect time of events given in input file. This algorithm doesn’t have any pruning (removing events which are less frequently occurring) facility due to that lots of pattern will form which causes more memory and execution time.

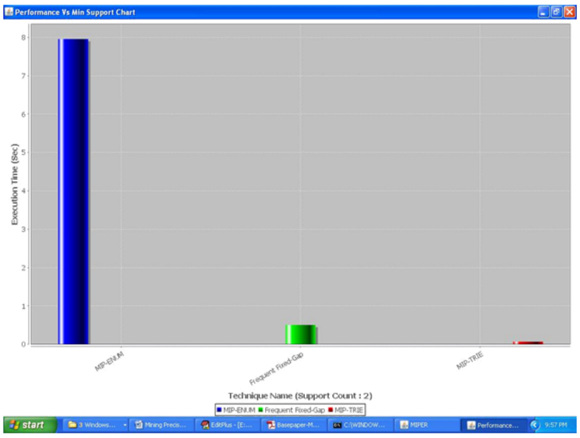




In above screen only one event occur which is greater than or equal to 3. Now click on ‘MIP-TRIE Framework Algorithm’ button to get same output but this algorithm will maintain all frequent patterns in tree instead of generating them which helps in less execution time.



In above screen after D,A and 2 interval later EA will occur, now click on ‘Performance Vs Min Support Chart’ button to get chart of execution time in graph



In above graph x-axis represents algorithm name and y-axis represents execution time

5. CONCLUSION

Event-based data, which are the focus of this project thesis, are a type of sequence data. “A sequence dataset consists of sequences of ordered elements, or events, with or without a concrete notion of time” In this context, a concrete notion of time implies an exact time-stamp indicating the initiation of each event as opposed to a relative notion of time which is implicitly retrieved from the ordering of the events. Sequence data are encountered in a large number of fields, such as shopping/transaction

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

data, internet surfing data, process control data, biological sequences, historical-, biographical-, career- event-sequences, medical records, and, of course, activity diary data, just to mention a few Sequences (or sequence data) are interesting since they enable the understanding of the evolving character of records in a dataset. They can give an overall view of the regular (or irregular) behaviour of the data over time, reveal trends within the data as well as help predict future events.

They allow for comparisons to be made and for the progress, over time, of events incorporated in different data records to be mapped and analysed. When analysing this type of data the patterns that are sought are, most often, sub-sequences whose distribution stands out for some reason. Sub-sequences that appear very often and/or in most of the data records may be interesting to detect, or that exhibit some sort of repetitious pattern, or even that differ from the greater part of the data. Such identification and analysis of sequences or sub-sequences finds applications in many fields. Analysing customer transactions, in terms of the sequence we formulate the problem of mining precise positioning episode rules (MIPER), which is helpful for applications where automatic responses are needed in a timely manner. We propose one enumeration approach for MIPER and further devise two approaches based on a compact trie structure to enhance the pruning power as well as reduce execution time of the mining process.

Experiments evaluate the efficiency of the proposed methods. Also, the effectiveness of precise-positioning episode rules is clearly demonstrated in a case study about China stock market. Sequence mining is a term characterizing a whole field within data mining concerned with the automatic identification of frequently occurring sub-sequences as patterns from large sequence datasets.

Even though frequency is not always the most interesting attribute when studying sequences it is often the one used in automatic searches due to the fact that it can easily be measured.

BILIOGRAPHY:

[1] Xiang Ao, Ping Luo, Chengkai Li, Fuzhen Zhuang, and Qing He. Online frequent episode mining. In ICDE, 2015.

[2] Jiawei Han, Jian Pei, and Yiwen Yin. Mining frequent patterns without candidate generation.In SIGMOD, 2000.

[3] Heikki Mannila, Hannu Toivonen, and A Inkeri Verkamo. Discovery of frequent episodes in event sequences. DMKD, 1997.

[4] Cheng-Wei Wu, Yu-Feng Lin, S Yu Philip, and Vincent S Tseng. Mining high utility episodes in complex event sequences. In KDD, 2013.

[5] Achar, A., Sastry, P., et al. (2013). Pattern-growth based frequent serial episode discovery. Data & Knowledge Engineering, 87:91–108.

[6] Agrawal, R., Imielinski, T., and Swami, A. (1993). ´ Mining association rules between sets of items in large databases. In ACM SIGMOD Record, volume 22, pages 207–216. ACM.

[7] Cho, C.-W., Wu, Y.-H., Yen, S.-J., Zheng, Y., and Chen, A. L. (2011). On-line rule matching for event prediction. The VLDB Journal, 20(3):303– 334.

[8] Gan, M. and Dai, H. (2011). Fast mining of nonderivable episode rules in complex sequences. In Modeling Decision for Artificial Intelligence. Springer.

[9] Huang, K.-Y. and Chang, C.-H. (2008). Efficient mining of frequent episodes from complex sequences.Information Systems, 33(1):96–114.

[10] Laxman, S., Sastry, P., and Unnikrishnan, K. (2007). A fast algorithm for finding frequent episodes in event streams. In 13th ACM SIGKDD. ACM.