INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS
ISSN 2320-7345

RETRIEVAL OF SEQUENCE PATTERN FROM SEQUENTIAL DATABASE USING MULTIPLE MINIMUM SUPPORTS

S.Priyanka[1], Ms. S.Leela[2]

[1]PG Full Time Student, CSE Department, Karunya University, Coimbatore

[2]Assistant Professor, CSE Department, Karunya University, Coimbatore

[1]pinkysomu24@gmail.com, [2]leela@karunya.edu

Abstract

The technique for determining time-related behaviour in sequence databases is sequential pattern mining (SPM). Discovery of items from sequential database is done based on frequencies of an item. The frequencies for various items in a sequence database are not exactly equal, in real-life applications. The rare item problem may occur If all items are set with the same minimum support it affect effective retrieve of interesting patterns irrespective whether minsup is set too high or too low. Introducing the concept of multiple minimum supports (MMSs) to SPM. It allows users to specify the minimum item support (MIS) for each item according to its natural frequency.MIS to store the crucial information about frequent patterns. In this study, an efficient method is proposed to discover the complete set of sequential patterns based on MMSs.

.

Keywords: Data mining, Sequential patterns, clustering, Multiple minimum supports

1. INTRODUCTION

Data mining is the process of searching large store of data to discover the patterns and likely outcomes will be predicted. Data mining can produce results that cannot be retrieved through simple query. It is the powerful tool in finding the patterns and relationship within the data. Sequential pattern mining (SPM) is a powerful technique in data mining for mining frequent time-related behaviour from a sequence database.

A given set of sequences will be considered data sequence and that data sequence as the input data. Individual data-sequence will be taken as a list of transactions, and each transaction holds a set of literals, called items. Provided user-specified minimum support known as minsup, SPM goal is to find all sub sequences that satisfy's minsup in a sequence database. The key factor that makes SPM realistic is the minsup. It is used to condense the search space and to limit the number of rules generated. Using single minsup unreservedly assumes all items in the database are of have like frequencies, which is over and over again not the case in real-life applications. That is, some items which appear very frequently and in same time rarely in the database. In such status, if minsup too high, finding those rules that engage infrequent items or rare stuff in the database. Setting minsup too low,

generating very large number of patterns and most of them are pointless. To separate out all of them would be an inconvenience for assessment makers.

For example in departmental store consider the items which is regularly bought by customer and uncommon item purchased that is rare items. The item which is purchased often is considered as frequent items. Based on the occurrence of an item minimum item support (MIS) will be given. Due to various merchandize uniqueness, each item have its own customer crowd, such as video games for adolescents, Grains and health drinks for elders or milk for everyone. Since the populations of the customer groups are not the same, it is impractical to discover interesting sequential patterns for all kinds of merchandises with a single minsup.

2. MULTIPLE MINIMUM SUPPORTS

The problem in single mix-up which is also known as the rare item problem is addressed by the concept of multiple minimum supports (MMSs) to extend the problem of SPM (Liu., 2006)[1]. To reproduce their only one of its kind nature it allows users to specify diverse minsups for different items and generate different sequential patterns; that is, there are different thresholds for various sequential patterns, depending on which items or events are enclosed in the patterns., The definition of a pattern is modified to find a pattern connecting rare items; that is, a pattern is considered as frequent if its frequency satisfies the minimum MIS value among all items. In short span of time mining the frequent pattern is done by using downward closure property If an item set does not satisfy minsup the supports of all its supersets will not satisfy minsup either, and the mining procedure will stop searching all of its supersets. However, frequent pattern mining with the MMSs, the downward closure property no longer holds. It is because a superset of an infrequent pattern may become a frequent one if a rare item is appended to it. Adding a rare item to an infrequent pattern will lower its threshold and allow the superset to have the possibility of becoming frequent which is explained in (Ya-Han Hu., 2012)[2].

3. RELATED WORKS

3.1 Web Access Pattern Tree

Web access pattern tree (WAP-tree) (C.I. Ezeife., 2004) [3] mining is a technique in sequential pattern mining for web log access sequences, which initially stores the unique web access sequence database on a prefix tree, alike to the frequent pattern tree (FP-tree) (Han et al.,2004)[4] for storing unordered data. WAP-tree algorithm then, mining the frequent sequences from the WAP-tree by means of recursively re-constructing intermediate trees, initial with suffix sequences and finale with prefix sequences.

This paper proposes a well-organized approach to mine frequent sequences by means of the WAP-tree, which utterly eliminates the need to engage in abundant re-constructions of intermediate WAP-trees during mining. The projected algorithm builds the frequent header node links of the original WAP-tree in a pre-order style and uses the position code of each node to identify the ancestor and descendant relationships between nodes of the tree. It then, identifies frequent sequential pattern, all the way through progressive prefix sequence search, starting with its first prefix subsequence event.

3.2 Temporal High Utility Item sets

Value of this item set is considered as the utility of an item set, and utility mining aims at identifying the item sets with its high value. Compare to pre-specified threshold of the data stream, the temporal high utility item set support will be larger. Discovery of temporal high utility item sets is an important process for mining interesting patterns like association rules from data streams. In this paper, (Chun-Jung Chu., Vincent S.Tseng., 2008) [5] propose a method, namely THUI (Temporal High Utility Item sets)-Mine, for mining temporal high utility item sets efficiently and effectively from data streams. Regarding mining temporal high utility item sets from data streams this is the first work. The contribution of THUI-Mine is by generating fewer candidate item sets such that the execution time can be reduced significantly in mining all high utility item sets in data streams that it can effectively identify the temporal high utility item sets. In this fashion, with less memory space and execution time the process of

discovering all temporal high utility item sets under all time windows of data streams can be achieved successfully. This meets the grave requirements on time and space efficiency for mining data streams.

3.3 A genetic-fuzzy mining approach

The common technique seen among mining technique is mining association rules in transaction of data. The previous mining approaches set a single minimum support value for all the items and identify the interaction among transactions using binary values. (Chun-Hao Chen Tzung-Pei Hong., 2008)[6] proposed a genetic-fuzzy data-mining algorithm for extracting association rules and membership functions from quantitative transactions using a single minimum support. In real applications, different items may have different criteria to judge their frequency. Proposing an algorithm which combine clustering, fuzzy and genetic concepts for extracting sensible multiple minimum support values, fuzzy association rules and membership functions from quantitative transactions. It initially uses the k-means clustering approach to gather alike items into groups. All items in the same cluster are considered to have similar character and are assigned similar values for initializing a better grouping. Individual chromosome is then evaluated by the criteria of requirement satisfaction and appropriateness of membership functions to estimate its fitness value.

4. CONCLUSION

Sequential pattern mining is one of the important domains of data mining. Since items in the database do not always have alike frequencies, in view of the concept of MMSs will produce solution of the rare item problem. To retrieve efficient mining sequential patterns with MMSs, a compact data structure, to store and compress entire sequence database called a PLMS-tree, is projected and an algorithm, called MSCP-growth, was developed to discover the inclusive set of patterns from the PLMS-tree.

REFERENCES

[1] Liu, B., 2006. Web Data Mining: Exploring Hyperlinks, Contents and Usage Data, 1st ed. Springer

[2] Ya-Han Hu□, Fan Wu, Yi-Jiun Liao An efficient tree-based algorithm for mining sequential patterns with multiple minimum supports National Chung Cheng University, Chiayi 62102, Taiwan, ROC

[3] C.I.EZEIFE., YI LU.,Mining Web Log Sequential Patterns with Position Coded Pre-Order Linked WAP-Tree, University of Windsor, Windsor, Ontario, Canada, N9B 3P4

[4] Han, J., Pei, J., Yin, Y., Mao, R., 2004. Mining frequent patterns without candidate generation: a frequent-pattern tree approach. Data Mining and Knowledge Discovery 8, 53–87.

[5] Chu, C.-J., Tseng, V.S., Liang, T., 2008. An efficient algorithm for mining temporal high utility item sets from data streams. Journal of Systems and Software 81 (7), 1105–1117.

[6] Chun-Hao Chen., Tzung-Pei Hong Vincent S. Tseng Chang-Shing Lee A genetic-fuzzy mining approach for items with multiple minimum supports: Springer-Verlag 2008