International Journal of Information & Computation Technology.
ISSN 0974-2239 Volume 7, Number 1 (2017), pp. 1-12

http://www. irphouse.com

Weighted Pre-Large-Tree Algorithm Generating Temporal Association Rule for Non-Frequent Items in Video Dataset

K. Kumar
Research Scholar, PRIST University, Chennai, Tamil Nadu, India.

Dr. P. Sudhakar
Annamalai University, Annamalai Nagar, Tamil Nadu, India.

Abstract

Internet posts heavy load of videos and images today. Data mining is capable of producing various patterns, correlations and associations. A number of efficient algorithms are proposed to do such mining processes. Anyway the recent advancement of huge video databases open doors for Association Rule mining (ARM) which is one of the current data mining techniques designed to group objects together from large video databases in extracting the interesting correlation and relation among video data. Most algorithms that are proposed on video association mining consider the frequent item sets and ignore the role of non-frequent items. But the reality is those non-frequent items are to be inspected for their contribution in any of the major shot sequences in the dataset. We attempt to propose such an algorithm that will look into the role of non-frequent items while mining the temporal sequences in video associations.

Keywords: Pre-Large-Trees, itemsets, association

INTRODUCTION

Video mining is an unsupervised discovery of patterns in an audio-visual content [1]. The temporal (motion) and spatial (color, texture, shapes and text regions) features of the video are utilized for mining task.

Existing algorithms on association rules assume that items have equal weights, and very few algorithms propose a weighted concept. Researchers were defining a weighted support which is calculated by multiplying a support of pattern with a weight. The weight is according to particular items which is more profitable.

Cai et al. presented two new algorithms in the year 1998 to find weighted frequent itemsets in a given database. The items are marked with weights to reflect their importance to the user. The weights may correspond to special promotions on specific products, or the profitability among different items. Hence it emerged mandatory to attach weight field to every item in the database. The first step is to search for the maximum size of the large itemsets. This requires a scan of the database. Further, these algorithms are based on candidate generation and pruning techniques, in addition to the application of ksupport bound property. This result in multiple scans of the database to find all weighted frequent itemsets.

Upon reading the recent works on the association of data items based on weights and temporal concepts[2][3] in video domain, it can be stated that researchers do not confirm which algorithm has the best performance. This is due to varying temporal attributes in videos.

RELATED LITERATURE AND MOTIVATION

Association rule problem is first introduced by R. Agrawal [2] to find frequent patterns, associations, correlations, or casual structures among sets of items or objects in provided data sources or other information repositories. Association relationship is useful in predicting marketing, decision analysis and market basket analysis fields [1]. Several techniques have been developed for mining association rules [2] such as FP-Growth algorithm [5], mining of generalized and multi-level rules, constraint based rule mining and mining multi- dimensional rules.

Temporal association rule mining was introduced by Wang, Yang and Muntz in 1999 together with the introduction of the TAR (Temporal Association Rule) algorithm [3]. Temporal association rule mining attempted to solve the problem on handling time-series by including time expression into association rules [4].

Temporal association finds possible relationships among the different item sets in temporal database. Temporal association rules are different from traditional association rules by the fact that temporal association rules attempt to model temporal relationships in the video database. Few types of temporal association rules defined in the literature are inter-transaction rules, episode rules, trend dependencies, sequence association rules and calendric association rules. On reading [4, 5, 6] most of existing techniques are developed based on temporal content analysis. Recent TAR algorithms that have been presented for general temporal association rule mining in database are PPCI algorithm, SPF and ITRAM. These rules are categorized as Calendric Association rule, Cyclic Association rule, Association rule based on partition, progressive weighted miner, incremental temporal association rule and periodic temporal association rule. Temporal databases are known to be continually being updated or appended. Videos are of temporal nature. Temporal association rule mining must synchronize with these updated transactions without any loss of time granularity and motion based attributes might not be ignored. Existing rule mining techniques cannot deal with the upcoming transactions of database as they might operate in batches.

Lu et al, proposed an algorithm in the year 2001, called Mixed Weighted Association rules which uses the concept of weighted support and with this algorithm it is possible to find vertical and horizontal association rules.

Zhang et al in the year 2003, assigned a weight by studying the novelty of data, based on the concept of weighted support, and Yun (2007) proposed a weighted confidence on mining interesting patterns.

Liewean Cheng in 2009, proposed two algorithms based on the concept that the greater the difference among items in an association rule, the higher the weights. The purpose is to discover perfect cross section relationship among items and then extract the unknown patterns.

C. H. Lee, suggested progressive partition miner (PPM) [6]. In PPM the database is first partitioned by the size of time granularity. The PPM algorithm apply a filtering threshold mechanism on every partition of the database to prune out cumulatively infrequent 2-itemsets. PPM also employs database scanning reduction technique. However, the limitation of this technique is its ability to deal with problems of incremental mining.

Cheng. Y. Chang et al gave an algorithm called segmented progressive filter (SPF) [7] that is based on the Segmentation and progressive filtering. The basic idea is to first divide the database into certain imposed time granularity. Then exhibition period of each item is investigated, and further segment the database based on their common start and end timings. For each part of the database it finds the 2-candidate item set with a cumulative filtering threshold. SPF applies best scanning reduction technique for generating candidate K-item set. In addition to generate all candidates it generates the sub-candidate and counts for the value of support. Temporal databases are continuously updated or appended. But SPF does not perform many incremental mining technique on refreshing database.

J. M. Ale et al extends the ideas of association rule incorporation of the time to their frequent item sets [6]. It tries to extend the existing non-temporal mining model by introducing the concept of temporal support. Such discovered association rule is done in a two-phase process; at first it find the frequent item set according to the lifespan of the item set and secondly it uses that frequent item sets to generate the rules. These rules are checked based on the applied confidence. In this proposed technique it however does not consider the updates of the database.

M. Chen et al presented a temporal association rule model to be used in video database for video event detection [5]. In this approach it captures the characteristics of temporal patterns with respect to the event of interest based on shots. M. Chen et al proposed their framework based on feature extraction, hierarchical temporal association mining and multimodal data mining. Traditional association rule mining approaches use a manually assigned threshold. The advantage of Chen's model is to use an adaptive mechanism for determining the essential threshold.

Byon et al presented an Exponential Smoothing (ES) filter for temporal association rule mining [2]. ES filter takes two steps; one will partition the database and then feed them into a Progressive Weighted Miner (PWM). PWM will use a weight function that gives greater weights to recent data than old data; each weight is divided by equal period [4].

Ru Miao et al suggested the idea of Apriori-extended mining periodic temporal association rules (MPTAR) [15]. Prior techniques of TAR did not consider the individual item exhibition period. MPTAR solved this problem, by including the exhibition period of individual item. Again MPTAR is a two-step periodic rule mining mechanism. The first step is mining the continous attribute through cycle curve and the second step is calculating the period of the attribute repeatedly. MPTAR did not define the cumulative threshold, and it is short of upcoming transaction entries in the association rule mining.

Non frequent items or patterns that occur in a temporal datasets are to be given due importance because of their roles in the shot sequences or repetition of shots in a video. Vijayakumar.V & Nedunchezhian.R(2012) proposed a novel method for mining temporal association rules using a weighted temporal tree[1]. This motivated us to propose a Pre-Large-Tree that can generate association rules[1][7] for large databases, which minimize the number of scans.

A pre-large concept is focused on the non frequent items specially and defined with a *w_low_thold* and *w_up_thold*. Both thresholds will be set by user. *w_low_thold* is based on the number of customized records permitted. If the numbers of customized records go beyond the permitted number, the algorithm will scan the database again to get the ending results. But if the number of customized records[5] does not go beyond the permitted number, the execution time is better saved. An itemset should be bigger than the *w_up_thold*, so that it will be thought as a large item set. If the ratio of support of an item set is below the *w_low_thold*, then it is considered as small item set. Pre-Large itemsets stores the item one by one in the growing mining process and minimizes the movements of itemsets from large to small items and vice-versa. *w_low_thold* is based on the number of updated records permitted in the database.

PROPOSED METHOD

Our System consists of two steps: ***Video transformation*** and ***Weighted Pre-Large-Tree (WPLT) algorithm*** with association rules mined.

<u>Video transformation</u>

The video transformation[4][6] contains three stages,

(i) **Shot segmentation -** Using histogram techniques[7] we can extract the key frames.
(ii) **Shot clustering -** This is done to explore the relationships among the shots.
(iii) **Shot labeling -** A class label[1] is assigned for each shot by its original temporal order as per sequence.

<u>Weighted Pre-Large-Tree(WPLT) algorithm</u>

(i) Construct Weighted-Pre-Large-Tree (WPLTree) with every elements of the Video temporal transaction database in D.
(ii) Discover all temporal frequent itemsets and non-frequent itemsets based on temporal weighted minimum support.
(iii) Discover association rules for frequent and non-frequent items.

(i) Algorithm to construct Weighted-Pre-Large-Tree

Input: Video temporal transaction database D

For each item with quantity ’q’ in a database D

do begin

Create Header_Table and Pre_Header_Table

Output Weighted-Pre-Large-Tree

end

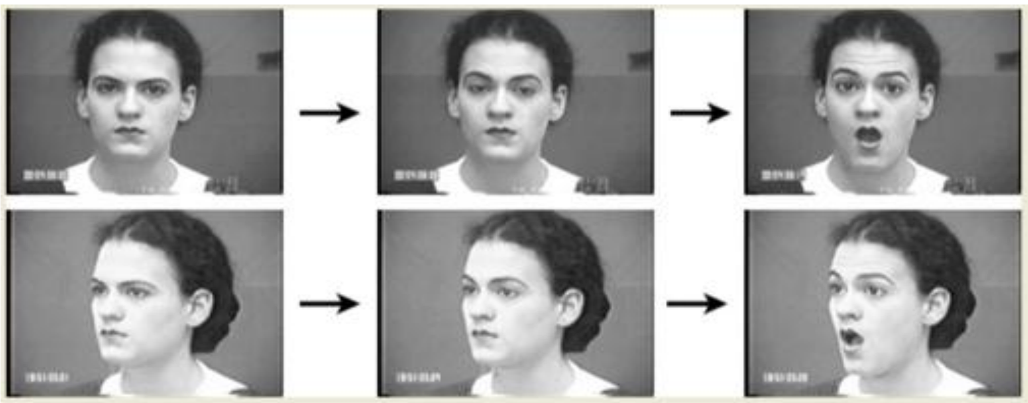


Figure 1: Video shot sequences (a to f)

Based on Fig 1, the database will contain 6 transactions and 6 itemsets denoted as {a} to {f}.

Table 1. Transaction itemsets

Tid	**Items**
1	a, b, d, e
2	a, a, b, c, d
3	a, b, b, c, d, e, f
4	a, a, c, d, e, e, f
5	d, e, f, f
6	a, b, f

Step 1: Finding frequency and priority of large Items

In the given database, ‘a’ has occurred 7 times, so ‘a’ has the highest priority and ‘c’ has occurred only three times, so it has the lowest priority.

Table 2. Frequency and priority of large items

Item	**Frequency**	**Priority**
a	7	1
b	4	5
c	3	6
d	5	2
e	5	3
f	5	4

Step 2: Ordering of items based on priority in descending order.

Table 3. Ordering and prioritizing of items

Tid	**Items**	**Ordering**
1	a, b, d, e	a, d, e, b
2	a, a, b, c, d	a, a, d, b, c
3	a, b, b, c, d, e, f	a, d, e, f, b, b, c
4	a, a, c, d, e, e, f	a, a, d, e, e, f, c
5	d, e, f, f	d, e, f, f
6	a, b, f	a, f, b

Step 3: Header_Table construction

Header_Table contains the frequency of occurrences of large items in the given database.

Table 4. Header_Table

Item	Frequency
a	7
d	5
e	5
f	5

Step 4: Pre_Header_Table construction
Pre_Header_Table contains the frequency of occurrences of pre-large items (Non frequent items, Nfw) in the given database.

Table 5. Pre_Header_Table

Item	Frequency
b	4
c	3

Step 5: Constructing Pre-Large-Tree, based on the values of Header_Table and Pre_Header_Table

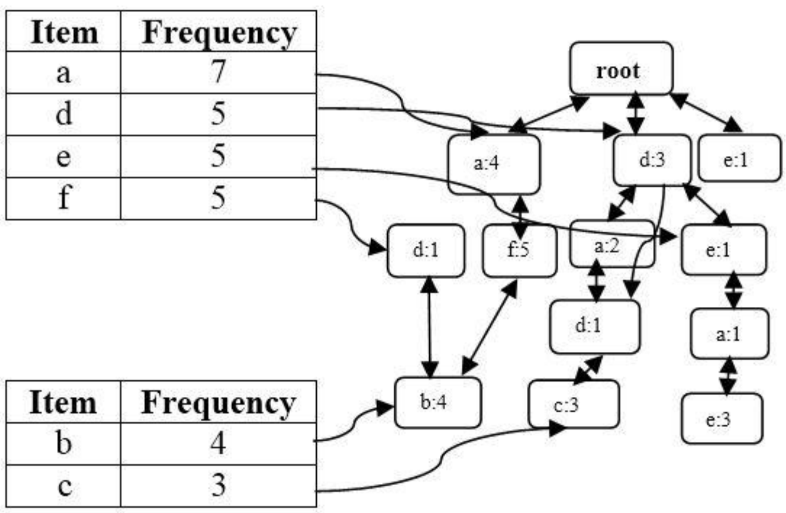


Figure 2. Weighted Pre-Large-Tree

(ii) Algorithm to determine Frequent and Non-frequent itemsets(Nfw)

The definition of weight defined in the discovery of frequent itemsets using weighted Tree approach is different, and is based on the quantity in which the items have been bought. If an itemset satisfies user defined weighted minimum support, then it is considered as a frequent itemset. But, even if some items appear in a small quantity, when the number of items in an itemset is large, the total quantity may be large, and the weight of that set is greater than the weighted minimum support. Depending on the application requirements, this may or may not be desirable. It may be desirable if the user considers a total quantity, which leads to profitability as interesting. It may not be desirable, if an itemset with many light weighted items should not be considered interesting. Further, in the weighted case, the downward closure property does not hold good. It is not necessary that all subset of a frequent itemset are frequent, necessary that all the subsets of a frequent itemset are frequent, since the meaning of frequent itemset is modified to handle weighted support. Hence downward closure property need not be true in a weighted case.

It is based on temporal weight F, the set of all frequent 1-itemsets, P, is the set of all non-empty subsets of F excluding the sets containing one item, and f is set of temporal items such as {ad} and {da} treated as different set of item sets and is an element of P. Fw is the set of all frequent one item sets. Vice- versa of the above process gives NFw, the set of all non-frequent item sets.

For example, consider the sample database given in Table 1. Assume *w_low_thold* = 4 and applying weight definition, then the items a,d,e,f are frequent

1-itemsets. Since, weight (a) = 7/4= 1.7, weight (d) = 5/3 = 1.66, weight(e) = 5/2 = 2.5, weight(f) = 5/1 = 5.0. Further, 2-itemset {a,d},weight ({a,d}) =12/3 = 4. An itemset {a,d} is frequent compared to other sets.

Input: A Weighted-Pre-Large-Tree
Output: Set of all frequent item sets, Fw and Set of all non-frequent item sets, NFw.

for each f in P

do begin T = {TIDs of first item in f}

for each m in f other than first item

do

if (transid.templocation of the ‘i’th item in the item set<transid.templocation of the ‘j’ th item in the item set)

begin

T = T ∩ {TIDs in which m is present}

end

if T is non empty then for each item ‘a’ in f

```
begin
if (sum (quantities of ‘a’ in every transaction t in T)) /|T| ≥ w_low_thold then
   flag=1;
else
   flag=0;
end
if(flag==0) and sum (quantities of elements of T w.r.t to f) /|T| ≥ w_up_thold then
         NFw = NFw U f
if ( flag==1) and sum (quantities of elements of T w.r.t to f) /|T| ≥ w_low_thold then
         Fw = Fw U f
end
```

(iii) Algorithm to discover association rules for frequent and non-frequent items

Input: Fw, a set of temporal frequent itemsets and NFw, non-frequent itemsets obtained based on weight. c, is *w_low_thold*.

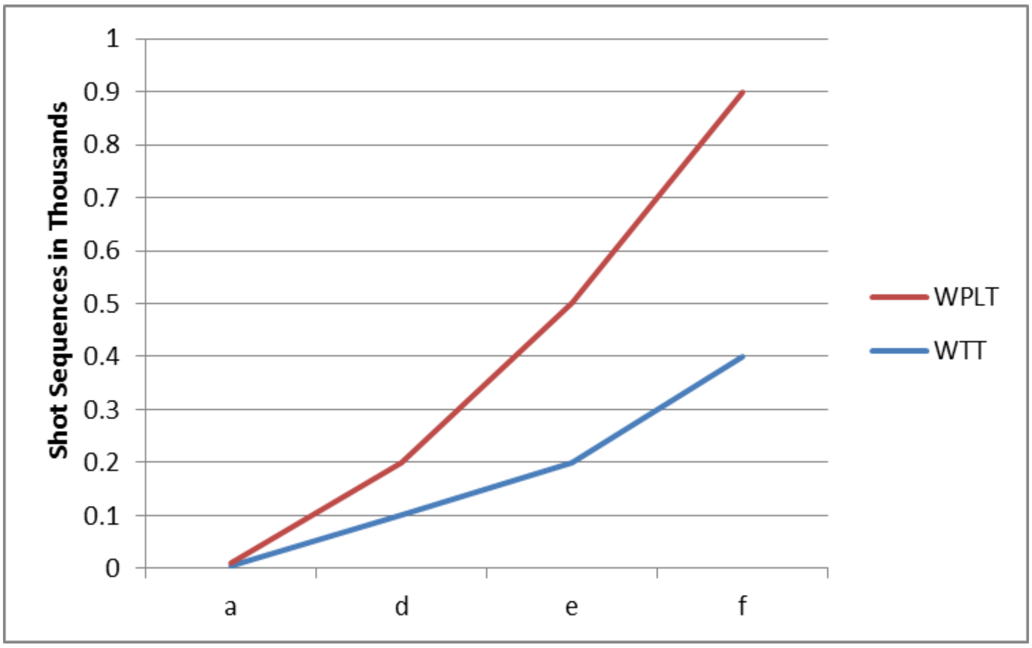
Output: Set of all Weighted Temporal Association rules.

```
for every items set f in Fw and NFw
begin
   for every subset s in f
 begin
   if(temporal weight(f)/ temporal weight(s)) ≥ c output
   a rule of the form s =>(f-s)
   if(temporal weight(f)/ temporal weight(s)) ≤c output
   a rule of the form s =>(f+s)R (repeat)
 end
end
```

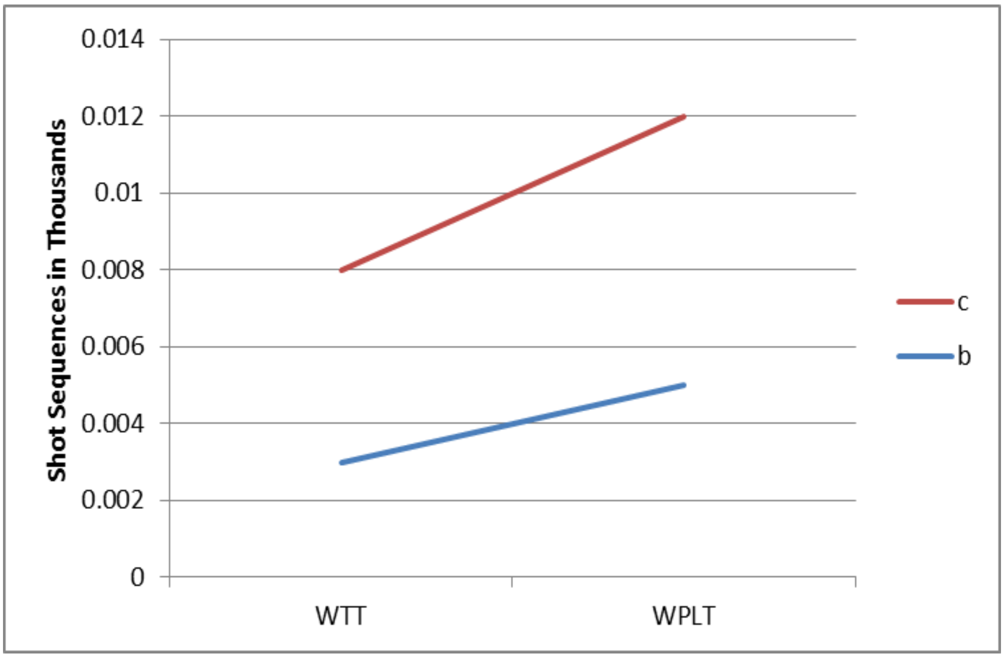
EXPERIMENTAL RESULTS

The proposed WPLT algorithm is implemented in Java, runnable with JCreator v2.5. The Video dataset taken as sample in this paper is assumed to be generating thousands of records in its shot sequences repetitive. The *w_up_thold* is set to 4 and *w_low_thold* to 2. The WPLT(proposed) algorithm is tested against the existing WTT algorithm in the following parameters.

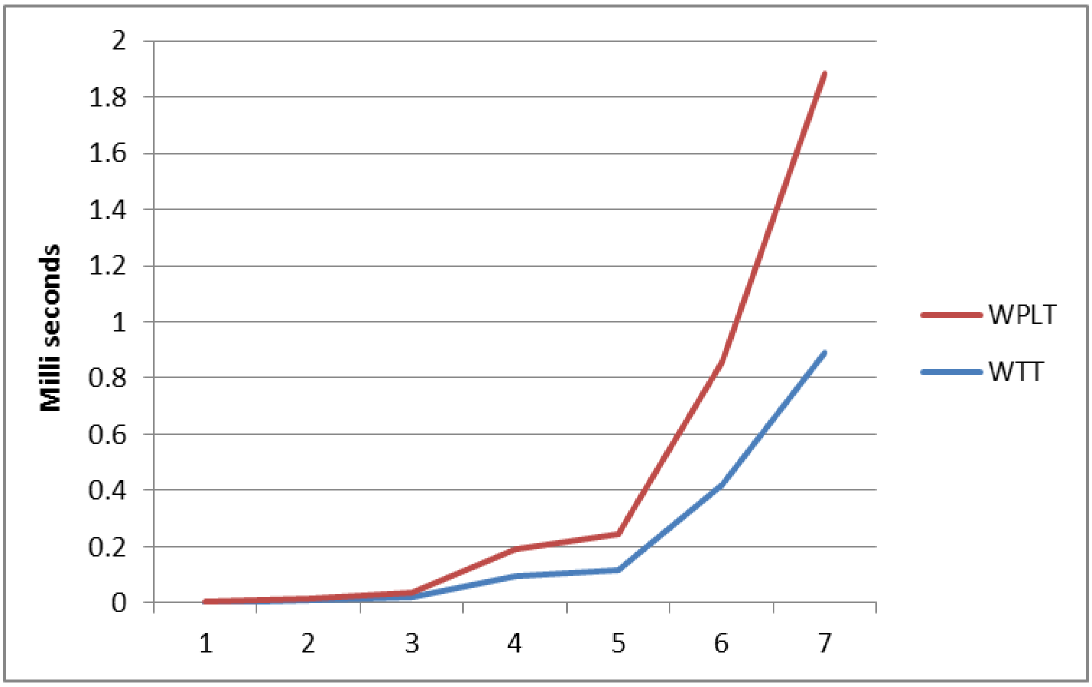
Generated Frequent itemsets with WPLT(proposed) and WTT algorithms



Generated Non Frequent itemsets with WPLT(proposed) and WTT algorithms



Comparing Execution time of WPLT(proposed) and WTT algorithms



CONCLUSION

The 1-itemset {b} {c} and 2-itemsets {b,c} or {c,b} which are Pre-Large-items are the non-frequent items Nfw in our transaction database. But if we observe the participatory nature of Nfw in the shot sequences, we can confirm that the association rule (f+s)R says that Nfw occupies the repeated scenes for the given dataset. Hence they cannot not be ignored by any algorithm and need to be identified for more shot sequences in the sample video dataset used. Rather, Pre-Large-Tree algorithm will be faster than batch FP-tree and FUFP-tree maintenance algorithms in handling customized records.

REFERENCES

[1] Vijayakumar.V & Nedunchezhian.R, 2012, 'A novel method for mining video association rules using weighted temporal tree', ARPN Journal of Engineering and Applied Sciences, vol. 7, no. 10, pp.1335-1340.

[2] Lin Lin & Mei-Ling Shyu, 2010,'Weighted association rule mining for video semantic detection', International Journal of Multimedia Data Engineering and Management, vol. 1, no.1, pp. 37-54.

[3] Preetham Kumar & Ananthanarayana V. S, 2010, 'Discovery of weighted association rules mining': Proceedings of IEEE, pp. 718-723.

[4] Naqvi M., Hussain K., Asghar S. & Fong S, 2011, 'Mining temporal association rules with incremental standing for segment progressive filter', Communications in Computer and Information Science, vol. 136, pp. 373-382.

[5] Dai Yueming, Li Yanwei & Wang Jinxin, 2011, 'Mining algorithm of weighted temporal association rules', Computer Engineering and Applications. vol. 47, no.26, pp.156-158.

[6] Lin C & Hong T, 2014,' Maintenance of pre-large trees for data mining with updated records', Information Sciences, vol.3, no.23,pp. 1-16

[7] Ding J, & Yau, 2009, 'An innovative data structure for mining association rules among infrequent items', Computers and Mathematics with Applications, vol. 57 no.2,pp.290–301.