VOL. 7, NO. 10, OCTOBER 2012 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

A NOVEL METHOD FOR MINING VIDEO ASSOCIATION RULES USING WEIGHTED TEMPORAL TREE

V. Vijayakumar[1] and R. Nedunchezhian[2]
[1]Research and Development Centre, Bharathiar University, Coimbatore, India
[2]Sri Ramakrishna Engineering College, Coimbatore, Tamil Nadu, India
E-Mail: veluvijay20@gmail.com

ABSTRACT

With the ever-growing digital libraries and video databases, it is increasingly important to understand and mine the knowledge from video database automatically. Video association mining is a relatively new and emerging research trend used to discover and describe interesting patterns in video. The traditional classical association rule mining algorithms can not apply directly to the video. It differs in two ways such as, spatial and temporal properties of the video and significance of the video sequence items. Most of the video association rules mining algorithms discover frequent item sets considers only temporal properties that do not consider the quantity in which items have been appeared in the video sequence. This paper discusses an efficient method for discovering a weighted temporal association rules from a large volumes of video sequence data in a single scan of the database using Weighted Temporal Tree structure. Video association rule consists two key phases are (i) Video pre-processing and (ii) Video association rule mining. The pre-processing phase converts the original video sequence into a temporal video transaction format. The mining phase consists three tasks namely, weighted temporal tree construction, frequent pattern extraction and rule mining. The proposed weighted temporal tree based association rule mining did not require multiple scans. The mined association rules have more practical significance and identifies the valuable rules comparing with Weighted Tree based algorithm. We also presented results of applying these algorithms to a synthetic data set, which show the effectiveness of our algorithm.

Keywords: weighted temporal tree, frequent temporal patterns, video association rule mining.

1. INTRODUCTION

Data mining is a process for extracting non-trivial, implicit, previously un-known and potentially useful information from in the large databases [1]. Multimedia databases are being acquired at an increasing rate due to technological advances in sensors, computing power and storage. Mining the hidden relationship among semantic concepts in video is important for effective content-based video retrieval and has gained great attention recently. There are two kinds of videos in our daily life namely, videos with some content structure such as the movie and news and videos without any content structure such as the surveillance and sports videos [2]. Many video mining approaches and techniques have been proposed for extracting useful knowledge from these video databases [3, 4, 5, 6, 7]. Video association rule mining is still in its infancy and an under-explored field. Only limited work was developed.

An association rule is defined as an expression X=> Y, where X and Y are sets of items and XUY = Φ. The rule implies that the transactions of the database which contain X tend to contain Y. The strength of the association rules can be measured in terms of its support, confidence and interest. The minimum support threshold indicates a user-specified minimum number of transactions that must contain the itemsets X and Y before the rule are considered relevant. The minimum confidence threshold is the minimum percentage of transactions that must contain both X and Y. A strong association rule satisfies both minimum support and confidence [8]. Association rule mining technique is broken into two steps. First step, it finds all frequent item sets. This step is computationally I/O intensive. Given 'm' items, there can be potentially 2^m itemsets. Efficient methods are needed to traverse this exponential growth of itemsets search space to enumerate all the frequent itemsets. Second step, it generates confident rules. This step is easier, but the overall performance of a mining algorithm is determined by the first step [9], [1]. Agrawal first introduced Apriori algorithm in 1994 for mining associations in market basket data [10]. Due to complex candidate generation in the data set Jiewai Han [10] invented a new technique of FP-growth method for mining frequent pattern without candidate generation. The efficiency of this mining technique is better than all most all algorithms like Apriori, aprioriTid, Apriori Hybrid because of a large dataset is compressed into a condensed, smaller data structure which avoids costly and repeated data scan.

Video Association Mining is referred to as the process of discovering association between the video items [4, 5, 6]. Two types of association are identified in video. First, Intra associations are those in which all items involved in the association are the same such as visually similar shots of the same object taken from different view points. For example in movie database the shot consist of similar objects in different view to impress the viewers. Second, Inter associations are those which consist of items of different types, which are scenes that consist of visually distinct shots of different objects. For example, the surveillance video database consists of different objects in the different shots.

Specification of spatial and temporal information in a video sequence plays an important role in conveying video content. A temporal aspect of association rule was proposed by Juan [11, 12]. To evaluate video associations, the temporal information integrates with traditional association measures (support and confidence). With a temporal video database, relationships between items discover with satisfy certain temporal constraints. Existing algorithms for mining association rules cannot be applied to temporal databases directly. This is because, in the existing algorithms, if an itemset is supported by a tuple, the tuple must contain all the items in the itemset. For temporal databases, an itemset, e.g. {A, B}, is supported as long as all the items in {A, B} are contained in a set of tuples which satisfy certain temporal constraint. Reference [13] represents various types of temporal data models such as univariate symbolic time series, symbolic time sequences, multivariate symbolic interval series and itemset sequence.

Min Chen, Shu-Ching Chen and Mei-Ling Shyu proposed a hierarchical temporal association mining approach to systematically capture the characteristic temporal patterns with respect to the events of interest [3]. B. Sivaselvan and N.P Gopalan proposed a technique for frequent set construction during the association mining phase for generating video associations [14]. Juan M. Ale and Gustavo H. Rossi expanded association rules with incorporating time to the frequent itemsets discovered. The algorithm A priori is modified to incorporate the temporal notions such as temporal support [15]. Naqvi, M., Hussain, K., Asghar, S. and Fong, S. proposed technique called Incremental Standing method for Segment Progressive Filter that modifies the frequent patterns in pace with changes to the database over time. The algorithm is used for supporting the temporal association rule mining in transaction database with different exhibition periods [16].

Recent researches in the field of temporal association rule mining are using Apriori based approach. It may still encounter some difficulties for different datsets. The limitations in the approaches are (i) Repeated scan is required to perform the mining in Apriori based temporal association rule. (ii) Generates a huge number of candidates in case of a dataset, which is large and/or sparse. The existing Apriori based temporal association rule, may easily cause thrashing when dataset become large and sparse. F. Guil, A. Bosch and R. Marín proposed Temporal Set- Enumeration Tree algorithm for frequent temporal pattern sequence mining from datasets. The algorithm uses a unique tree-based structure for storing all frequent patterns discovered in the mining process [17]. Keshri Verma and O. P. Vyas proposed an algorithm gives time sensitive approach for mining frequent item in the dataset. Temporal H-mine algorithm takes advantage of H-struct data structure and dynamically adjust link in the mining process [18].

Lots of algorithms for mining temporal association rules have been proposed at present and most of them treat each item uniformly but decision-makers are more inclined to items whose profits are higher than others. The classical model of association rule mining employs the support measure, which treats every transaction equally. In contrast, different transactions have different weights in real-life data sets. For example, in the market basket data, each transaction is recorded with some profit. Different items always have different importance. To reflect them, the way of draw weight into items and use weight association rules can solve the problem. There are many interesting algorithms for finding frequent itemsets based on user defined minimum support, and a few algorithms are based on weighted concept [7] [19]. Ke Sun and Fengshan Bai proposed a weighted association rule mining with a new measure w-support, which does not require pre-assigned weights. It takes the quality of transactions into consideration using link-based models. First, the HITS model and algorithm are used to derive the weights of transactions from a database. Based on these weights, w-support is defined to give the significance of item sets. It differs from the traditional support in taking the quality of transactions into consideration. An Apriori-like algorithm is proposed to extract association rules whose w-support and w-confidence are above some given thresholds [20]. Kanak Saxena proposed temporal weighted miner algorithm, which reflects the importance of each transaction period then, the algorithm partitions the time-variant database in light of weighted periods of transactions and performs weighted mining [21].

Lin Lin and Mei-Ling Shyu proposed video semantic concept detection framework via weighted association rule mining. First, they utilized the functionality of multiple correspondence analysis to measure the correlation between different 1-feature value pairs and the classes to infer the high-level concepts from the extracted low-level audio-visual features. Next, the association rules are generated by using the weighted 1-feature-value pairs, where the correlation information and the percentage information are integrated as the final weight measurement. Classification is performed directly by using those weighted 1-feature-value pair rules [7].

The semantics of weight is a measure of the importance of an item set. If an item set is very important, for example, it is under promotion, or it is highly profitable, then even if not many customers have bought it, it is still an interesting item set to the user. Preetham Kumar, Ananthanarayana V S proposed a method for discovering a weighted association rules from a large volumes of data in a single scan of the database using Weighted Tree data structure [10].

The remainder of the paper is organized as follows. Section 2 introduces the motivation of the work. We detail the proposed framework in Section 3. Section 4 shows the experimental results and Section 5 concludes this paper.

2. MOTIVATION

From literature review on association of data items based on weights and temporal concepts in video domain, it is evident that researchers do not confirm which algorithm has the best performance. Moreover, less number of the algorithms is based on the quantity in which the items have been appeared in the video sequence. In a

VOL. 7, NO. 10, OCTOBER 2012 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

large video database it is possible that, even if the item set appears in very few scene clusters, it may be appeared in a large quantity for every scene cluster in which it is present, and may lead to very high semantic information. Consider for example a sample video temporal transaction database given in Table-1, and Table-2 in which every element of each video transaction represents quantity of the respective item.

Table-1. Sample video temporal transaction database.

Trans ID	Video items					
TID001	A	B	A	A	A	
TID002	C	A	C	C		
TID003	A	C	A	B	B	B
TID004	A	C	C	B		
TID005	B	D	B	D	B	D

Table-2. Sample video database with item count.

Trans ID	Video items			
	A	B	C	D
TID001	4	1	0	0
TID002	1	0	3	0
TID003	2	3	1	0
TID004	1	1	2	0
TID005	0	3	0	3

In this table, if user defined minimum support are two transactions, then an item D is not frequent and will not appear in the set of frequent item sets, even though it is appearance in large quantity, and leads to more semantic information than other frequent items. Therefore the quantities of items are appeared is the most important component.

The weight of frequent item sets appeared in the video sequence is different from the traditional weighted measure and is based on the quantity in which the items have been appeared in the sequence. The weighted temporal support of an item set is calculated by the product of the total weight of items in the item set and the support of the item set.

$$Weight(I) = \frac{\sum_{j}\sum_{i} qij}{n}$$

where i = 1, 2,….k; and j = 1, 2, .. n and qij represents a quantity of an item iεI , in a jth transaction.

In a video transaction database, if some items appear in a small quantity, when the number of items in an item set is large, the total quantity may be large and the weight of that set is greater than the minimum support. It may be attractive if the user considers a total quantity, which leads to profitability, as interesting. It may not be popular, if an item set with many light weighted items should not be considered interesting.

For example, consider the sample database given in Table-1. Assume w_min_sup = 2 and applying weight definition given in equation, then the items A, B, C and D are frequent 1-itemsets.

Since, Weight (A) = 8/4= 2.0, Weight (B) = 8/4 = 2.0, Weight(C) = 6/3 = 2.0, Weight (D) = 3/1 = 3.0. Further, 2-itemset {A, B} is also frequent as Weight ({A, B}) =12/3 = 4. The item B appears in smaller quantity in the transaction sequence {1, 4} and A appears in the small quantity in the transaction sequence {4}, but an item set {A, B} is frequent. This issue has motivated us to propose a method to discover frequent item sets based on weight in a single scan of the database.

3. PROPOSED METHOD

We proposed a weighted temporal tree for association rule mining as shown in Figure-1. Our system framework consists of two steps namely, video transformation and weighted temporal tree association rule mining. The video transformation step contains three stages, (i) Shot detection such as detection of key frames from video sequence; (ii) Shot clustering such as the extracted key frames clustered and (iii) Shot labeling such as class labels are assigned to each key frame and finally the video temporal transaction database is constructed. In the next step, we mined associations from the temporal transaction dataset to find strong correlations using the weighted temporal tree.

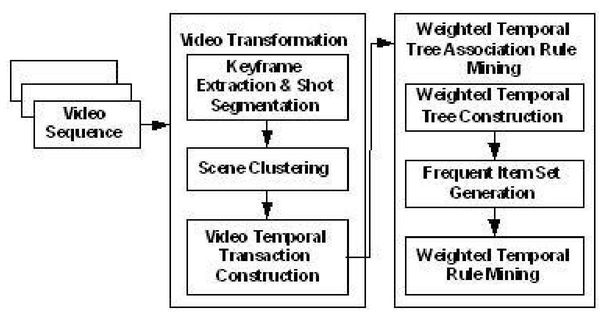

Figure-1. Proposed method.

3.1. Video transformation

To mine video data, one of the most important tasks is to transform the original video sequence into a temporal sequence dataset. To facilitate this goal, we adopt the following video processing techniques.

(i). Shot segmentation: The video shots are segmented and key frames were extracted using histogram techniques.
(ii). Shot clustering: The shot clustering mechanism is adopted to cluster the visually similar shot (Key frames) using k-means clustering algorithm to explore the relationships among the shots.
(iii). Shot labeling: Then each shot will receive a class label, we sequentially aggregate the class information of each shot by its original temporal order to form a shot

ISSN 1819-6608

cluster sequence. As demonstrated in Figure-2(a), where each icon (key) image denotes one shot and the letter below it indicates its class label, the acquired shot cluster sequence is given in Figure-2(b). Finally, the video temporal sequence database is constructed.

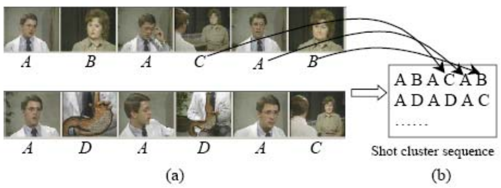

Figure-2. Transforming a video into a relational dataset: (a) video shots in the original temporal order (left to right, top to bottom); (b) a shot cluster sequence.

3.2. Weighted temporal tree association rule mining

After video pro-processing, video V would be transformed into a temporal transaction dataset *D*. The correlations among the items of *D* would reflect the associations among video shots and clusters. To mine association from video transaction database *D*, we develop a weighted temporal tree ARM algorithm by integrating the basic mechanism of weighted tree and temporal concepts. It can be decomposed into three phases, namely weighted temporal tree construction; temporal frequent patterns construction and finding temporal rules. The proposed algorithm generates association rules in video temporal dataset whose temporal support and temporal confidence are larger than min-weight-supp and min-conf specified by users, respectively.

Algorithm for discovery of weighted temporal association rules

Input: Video temporal transaction database D

Output: Weighted temporal association rules

Steps

a) Construct Weighted Temporal Tree (WTTree) with every elements of the Video temporal transaction database in D.
b) Discover all temporal frequent itemsets based on temporal weighted minimum support.
c) Apply Association Rules Mining Algorithm to discover weighted temporal association rules based on user defined minimum confidence. This step is a straight forward step. The most important step is discovery of frequent itemsets based on user defined temporal weighted minimum support.

Structure of the weighted temporal tree (WTTree)

Weighted Temporal Tree has three different nodes, namely head node, transaction node and temporal node.

(i) The head node shown in Figure-3(a) contains two data fields namely an item name, temporal count of the item. And also it contains two pointers, one pointing to the transaction node, and the other is pointing to the next branch head node.
(ii) The transaction identifier node shown in Figure-3(b) has two parts. The first part represents a transaction id and the second part indicates temporal weight in that transaction. This node has two pointers, one pointer pointing to the next transaction id having this particular item and another pointing to the temporal location of the item in the transaction.
(iii) ii) The temporal node shown in Figure-3(c) has one data part. It represents a temporal location. This node has only one pointer pointing to the next temporal location having this particular item.

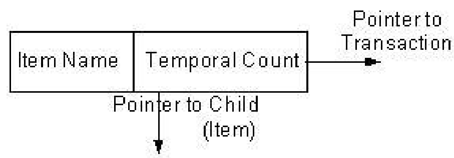

Figure-3(a). Head Node.

Figure-3(b). TID Node.

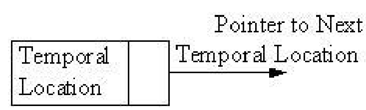

Fig.3(c). Temporal Node.

(i) Construction of weighted temporal tree

Input: Video temporal transaction database D

Output: Weighted Temporal Tree

Method

For each item with quantity 'q' in a temporal database T □ D do begin

Create a head node labeled with 'q' and add the transaction identifier nodes to the respective item head node along with temporal location node.

end

(ii) Algorithm to determine all temporal frequent itemsets based on temporal weight

F is 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 {AB} and

{BA} treated as different set of item sets and is an element of P.
Fw is the set of all frequent one item sets.

Input: A Weighted Temporal Tree

w_min_sup = user specified weighted minimum support

Output: Set of all frequent item sets, Fw.

Method
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_min_sup
then flag=1;
end
if (flag==1) then
if (sum (quantities of elements of T w.r.t to f)) /|T| ≥ w_min_sup then
Fw = Fw U f
end

(iii) Discovery of weighted temporal association rules

Input: Fw, a set of temporal frequent itemsets obtained based on weight.
c, a user defined minimum confidence.

Output: Set of all Weighted Temporal Association rules.

Method
for every items set f in Fw
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)
end
end

Illustration

Consider for an example, a sample video temporal database given in Table-1. A Weighted Temporal Tree for this database is shown in Figure-4.

The patterns obtained from the video association mining differ with patterns derived by the conventional frequent pattern mining due to the temporal property. Since changing the position of frequent items in a pattern leads to another pattern which is different. For instance, in video association mining the pattern ABCD differs from ABDC because in the second pattern D occurs before C.

If w_min_sup = 2 then applying above algorithm, we get Fw = {A, B, C, D}. If we applied step-2 of the algorithm, the total number of itemsets are grows extremely. Consider a set {A, B}, which appears in transaction 4. i.e., T= {1}. Also, sum of quantities/|T|=5/1=5. Even though {A, B} satisfies w_min_sup, weight of B corresponding to the number of transactions in T is less than w_min_sup. i.e., weight (B) in T = 1/1 =1 < w_min_sup. Hence {A, B} is not frequent. By similar arguments we can prove that {B, C} is also not frequent.

Consider a set {B, D}, which appears in transaction 5 i.e., T = {5}. Also, sum of quantities/|T|=6/1=6. Hence it is frequent. Similarly, the set {D, B}, with appears in the transaction 5 also frequent.

Finally, we found that the sets Fw = {{A}, {B}, {C}, {D}, {B, D}, {D, B}} are frequent sets.

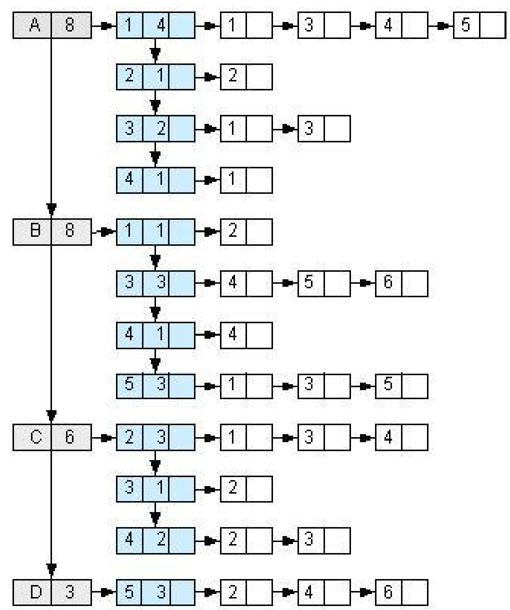

Figure-4. Weighted temporal tree for Table-1.

4. EXPERIMENTAL ANALYSIS

For the experimental analysis, simulation of video temporal transaction database is generated with various kinds of sizes and data distributions using the dataset generator [22], and in the data set which we have used, every element of the transaction is considered as quantity of the corresponding items in the database. The synthetic data set was generated with five predicting attributes and five domain attributes constructed randomly. The experiments were conducted on a 2.10 GHz Intel Dual core system with 4GB RAM running on Microsoft Vista. The algorithm was implemented using Java. The data sets used here contain transactions 10, 20, 50, 100, 300, 500 and 1000 with 6 items. The weighted minimum support used is equal to 2.

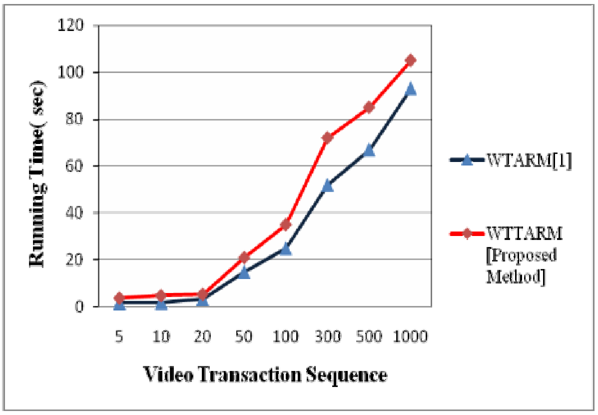

Figure-5. Running time.

Even though the Weighted Temporal Tree algorithm [Proposed Method] requires more time than WTARM [10] algorithm as shown Figure-5, it is efficient in the temporal rule generation. It requires more space than WTARM [10]'s algorithm. This may be due to check its temporal property in the itemset.

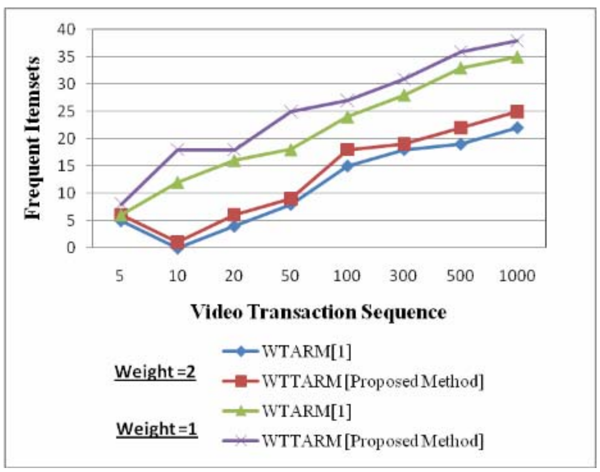

Figure-6. Generated frequent itemsets.

As Figure-6. shows, along with weighted minimum support increasing, the number of frequent items is descending. The reason is that the greater the value of wighted minimum support, the more items are filtered. The number of rules generated by the WTTARM is greater than WTARM [10] because the temporal properties produces worth items.

5. CONCLUSIONS

To facilitate video database management, we have explored a new research area of video association rule mining. In this paper, we proposed a novel method for discovering frequent itemsets based on weights temporal tree is discussed. The temporal concept is integrated with the weighted tree. The quantities of items appeared in the sequence is given the significance for the weight computation. The proposed algorithm overcomes the repeated scans. Though the algorithm reduced the number of scans considerably, we fell there is still some overhead a result of the limited scans of the original input and traversal of the tree. Case specific applications of generated frequent patterns such as classification, indexing and summarization and dedicated algorithms for the same is a candidate for further research.

REFERENCES

[1] Han J. and Kamber M. 2000. Data Mining: Concepts and Techniques. Morgan Kauffmann.

[2] Marcela X. Ribeiro, Agma J. M. Traina, Caetano Traina and Paulo M. Azevedo-Marques. 2008. An Association Rule-Based Method to Support, Medical Image Diagnosis with Efficiency. IEEE Transactions on Multimedia. 10(2): 277-285.

[3] Min Chen, Shu-Ching Chen and Mei-Ling Shyu. 2007. Hierarchical Temporal Association Mining for Video Event Detection in Video Databases. Proceedings of the 2nd IEEE International Workshop on multimedia databases and data management, in conjunction with IEEE International conference on data engineering, Turkey.

[4] Zaiane O.R., Han J., Li Z. and Chiang J. 1998. Multimedia Miner-A System Prototype. In: Proceedings of ACM SIGMOD International Conference on Management of Data. pp. 581-583.

[5] Wijesekara. D. and Barbara D. 2000. Mining Cinematic Knowledge-Work in Progress. In: Proceedings of International Workshop on MDM-KDD. pp. 98-103.

[6] Zaiane. O.R., Z. Li, J. Han and J. Hou. 1998. Mining Multimedia Data. In: Proceedings of CASCON: Meeting of Minds Conference. pp. 83-96.

[7] Lin Lin and Mei-Ling Shyu. 2010. Weighted association rule mining for Video Semantic Detection. International Journal of Multimedia Data Engineering and Management. 1(1): 37-54.

[8] R. Agrawal and R. Srikant. 1994. Fast algorithm for mining association rules. In: Proceedings of VLDB. pp. 487-499.

[9] Agrawal R. and Srikant R. 1994. Fast Algorithms for Mining Association Rules. IBM Research Report J9839, IBM Almaden Research Center.

[10] Preetham Kumar and Ananthanarayana V. S. 2010. Discovery of weighted association rules mining. Proceedings of IEEE. 5: 718-723.

[11]Juan M. Ale and Gustavo H. Rossi. 2002. An Approach to Discovering Temporal Association Rules, ACM SIGDD.

[12]Keshri Verma and O. P. Vyas. 2005. Temporal Association Rule Using Without Candidate Generation. Proceedings of 3rd International Conference, Caliber, Cochin, Kerala, India.

[13]Fabian Morchen. 2007. Unsupervised Pattern Mining from Symbolic Temporal Data, ACM SIGKDD Explorations Newsletter - Special issue on data mining for health informatics. 9(1): 41-55.

[14]B. sivaselvan and N.P. Gopalan. 2006. An efficient frequent temporal pattern mining algorithm. Information technology journal. 5(6).

[15]Juan M. Ale and Gustavo H. Rossi. 2000. An approach to discovering temporal association rules. Proceedings of the ACM symposium on applied computing. 1: 294-300.

[16]Naqvi M., Hussain K., Asghar S. and Fong S. 2011. Mining Temporal Association Rules with Incremental Standing for Segment Progressive Filter. Communications in Computer and Information Science. 136, Part 7, pp. 373-382.

[17]F. Guil, A. Bosch and R. Marín. 2004. TSET: An algorithm for mining frequent temporal patterns. In: Proceedings of the 1st International Workshop on Knowledge Discovery in Data Streams, in conjunction with ECML/PKDD. pp. 65-74.

[18]Keshri Verma and O. P. Vyas. 2005. Efficient calendar based temporal association rule. SIGMOD Record. 34(3): 63-70.

[19]Dai Yueming, Li Yanwei and Wang Jinxin. 2011. Mining algorithm of weighted temporal association rules. Computer Engineering and Applications. 47(26): 156-158.

[20]Ke Sun and Fengshan Bai. 2008. Mining Weighted Association Rules without Preassigned Weights. IEEE transactions on knowledge and data engineering. 20(4): 489-495.

[21]Kanak Saxena. 2008. Efficient Mining of Weighted Temporal Association Rules. Proceedings of IEEE. pp. 421-425.

[22]Dataset generator: www.datasetgenerator.com/.