ISSN No. 0976-5697
Volume 8, No. 3, March – April 2017
International Journal of Advanced Research in Computer Science
RESEARCH PAPER
Available Online at www.ijarcs.info

Discovering Pattern for interval and subinterval based on time for Large Growing Database

Vijay Kumar Verma
Asst. Prof.
Computer Science and Engineering Department
Lord .Krishna .College of Technology
Indore M.P. India

Kanak Saxena
Professor and Head
Computer Application Department
Samrat Ashok Technological Institute
Vidisha M.P. India

Abstract:-Data Mining is a rapidly growing area in research because of its involvement in several disciplines including Statistics, Finance, Health care, Pattern recognition, Temporal databases, Visualization, Classification and prediction, e-commerce, parallel and distributed computing. Time is an important factor for every business organization. Time includes a day, a month or a year. Mining useful pattern with time is an important issue. Traditional pattern mining techniques mainly focus on extracting different patterns. By using time factors we can found regularities and irregularities in different pattern in a particular interval. With the help of time oriented data mining techniques we can predict several important things like sufficient or insufficient quantitative of an item. This paper present an efficient approach to mine instructing pattern based on time. We proposed a new method which reduces number of database scan, execution time and different pattern for interval and subinterval

Keywords: Data mining, Time, Pattern, Regularity, interval

I. INTRODUCTION

Data mining explore and analyze data to represent knowledge for business decision making. Data mining used techniques to extract meaningful facts, relations, and patterns from large databases. Different pattern play an important role for business enterprises to identify unexpected relations in the data items. By Identify these trends and relations we can increase revenue, reduce cost and decide future plan. Mining pattern using time factor from large incremental database usually includes two important things namely, time periods denotes the time duration with respect to the real world, transaction time [1,2,3].
Time based data mining include

1. Prediction
2. Time based trends analysis
3. Classification based on time
4. Time based cluster analysis
5. Time based pattern analysis

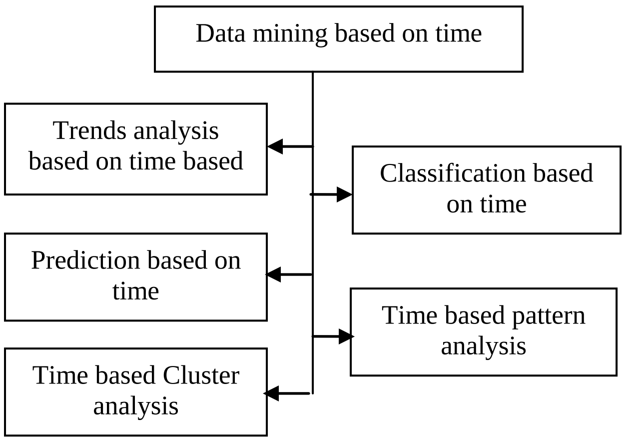

Figure 1 Different data mining dimension related with time

II. BASIC CONCEPTS

Consider a large time oriented database TD. Let n be the total number of interval based on time. $tb^{t,n}$ is a continuous interval from partition I_i to partition I_n.

$$|tb^{t,n}| = \sum_{i=t,n} |I_i|, tb^{t,n} \dot{\subseteq} TD$$

Where is the latest starting partition number?

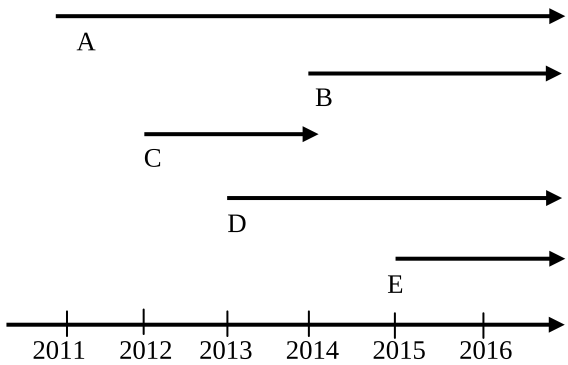

Figure 2 Items with different interval

A. Growing type products-

Item sets appear continually from the prior to last is called progressive candidate set in the previous phase.

B. Intervene type products

Products that dose appear continually in entire interval. Present in certain subinterval or partition which may which may or may not continue.

C. Basic notations

There are three basic notations which are used to maintain time oriented pattern are [8,9,10]

1. S: The partition number of the corresponding starting partition

2. C:The number of occurrences of C

(A) X.C: Present the occurrence in partitions.

(B) X.S: Present Starting partition number of the database when X becomes an element of frequent set.

(C) X.E: Ending partition number of the database when X becomes an element of frequent set

III. FUNDAMENTAL PROBLEMS

Time oriented pattern mining considers an item using its life time in an interval or subinterval. There are two important two important problems that have to be considering

(1) The life time of an item may include a set of intervals based on a given threshold.

(2) Life time of an item not exist for interval but just for subintervals [4,5,17].

IV. LITERATURE REVIEW

In 2010 Tarek F. Gharib et al. "An efficient algorithm for incremental mining of temporal association rules". They presented the concept of temporal association rules. These rules are useful to solve expressions including time. Time based databases are regularly updated so that the discovered pattern is also updated. Running the same algorithm every time is ineffective because ignore old patterns and doing the same things again. Incremental mining techniques cannot deal with these problems. They proposed a new algorithm which store previous rules. This approach provides significant improvement over the conventional methods [6].

In 2011 Sheng Xiang Fan et al. proposed. "Hybrid Temporal Pattern Mining with Time Grain on Stock Index". They used time grain concept and proposed new hybrid models for pattern mining. The proposed approach is based on event length limit. The hybrid sequences are divided into events and differentiate as per threshold value. For the experimental analysis they used Taiwan Stock Exchange data. By the experiments they show the feasibility and effectiveness of proposed approach. The proposed approach is effectively solving the problem of hybrid events. They also showed that in future this method also enhance dynamic hybrid pattern mining [7].

In 2012 Iyad Batal et al. proposed "Mining Recent Temporal Patterns for Event Detection in Multivariate Time Series Data". They proposed a new framework for finding predictive patterns for monitoring and event detection problems. In the proposed framework they convert time series into time-interval. It then constructs temporal patterns using backwards temporal operators. They proposed new temporal constraints for finding temporal patterns (RTPs). They presented an efficient algorithm that mines time-interval patterns starting from patterns related to the most recent observations. They showed that the RTP framework is very useful to efficiently find patterns that are important for predicting various types of disorders associated with diabetes [11].

In 2013 Sílvia Moura Pina proposed "A Constraint-Based Algorithm for Mining Temporal Patterns in Transactional Databases". They introduced temporal constraint with complete cyclic temporal constraints, partial cyclic temporal constraints and timespan constraints. They proposed a new algorithm with these constraints. They main focus is on looking for timespan patterns. The proposed algorithm allows discovery of both lifespans and periodical patterns, according to any time granularity chosen, without any further pre-processing[12].

In 2014 Gurram Sunitha et al "Data Sets: A Survey". They highlight the importance and applications of spatio-temporal pattern mining. They provide survey of mining techniques for discovering three types of spatio-temporal patterns sequential patterns, co-occurrence patterns. They also showed importance of spatio-temporal data mining and its applications with respect to sound. They also proposed survey of the approaches available for mining frequent patterns from spatio-temporal databases specifically from event data sets and trajectory data sets is provided [14].

In 2015 Vangipuram Radhakrishna et.al proposed "A Novel Approach for Mining Similarity Profiled Temporal Association Patterns". In the proposed approach they used two types of supports called positive support and negative support. The advantage of proposed approach is needs only a single scan of temporal database. This approach eliminates the overhead incurred for multiple database scan. The proposed approach also defined for a finite number of time slots, we aim to discover all such temporal association patterns from temporal database which are similar to the reference vector. They also introduced the concept of negative item set support values and negative item set support sequences [15].

In 2015 Yi-Cheng Chen proposed "Mining Temporal Patterns in Time Interval-Based Data". The proposed two novel concepts, endpoint and endtime. They simplify the processing of complex relationships among event intervals. They proposed three types of interval. They also develop two novel methods Temporal Pattern Miner and Probabilistic Temporal Pattern Miner to discover three types of interval-based patterns. The proposed pruning techniques reduce the search space of the mining process. Experimental studies show that both algorithms are able to find three types of patterns efficiently. Furthermore, we apply proposed algorithms to real datasets to demonstrate the effectiveness and validate the practicability of proposed patterns [16].

In 2016 J. Mercy Geraldine1 "Weighted Temporal Pattern Mining With Dimensionality Reduction Using Modified AFCM Technique". They proposed a dimensionality reduction method. This method reduces the quantity of data considered for mining. In the proposed approach the time based data are converted into fuzzy data. Fuzzy data are provided as input to the proposed Modified Adaptive Fuzzy C Means (MoAFCM) algorithm which is a combination of FCM clustering algorithm and Cuckoo search optimization algorithm. Proposed approach performed on these clusters to identify the effective temporal patterns which consider knowledge about the patterns [17].

The fundamental idea behind the proposed approach is to reduce the number of times that data bases need to be scanned. The original database is normally much larger than the incremental database. Therefore, scanning the original database is time-consuming. This work proposed a new method for growing database. The proposed approach is based on the concepts that new records are added in the current year (which is the last partition) of database and remaining partition. The proposed approach scans only current partition. We scanned all previous partition and stored pattern separately

We organized database into two parts current partition and remaining partition. In the current partition is new records are added regularly Whole data base logically to n portions. Portion is dented by (P1, P2, ..., Pn) and Pc denotes current n represent the continuous time interval from partition.

X.count++;
}
}
for all $X \in C^m_k$
{
if $X.count \geq s * |P_m|$

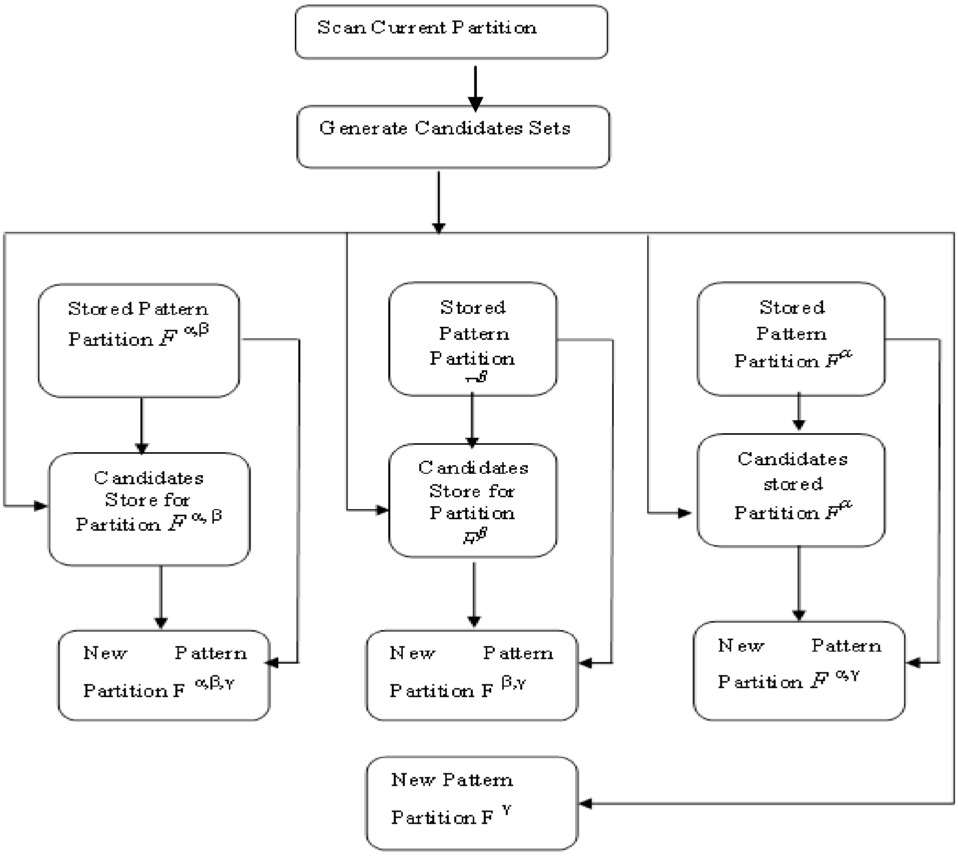

Figure 3 Architecture of the proposed approach

V. PROPOSED ALGORITHM

Input: *DB*: Database (original database) with size with n number of partition form *(db_1 db_2 ……db_n) records is appended into last partition which is current partition. F^n is set of pattern where n is last partition number.*

Output: *pattern with interval and subintervals*

Procedure:

C^m *the candidate in partition m*
F^m *pattern for partition m*
C^α; C^β, $C^{\alpha,\beta}$ *candidate sets in partition α,β and αβ*
F^α; F^β, $F^{\alpha,\beta}$ *pattern for in partition α,β and αβ*
Step1 $F^\alpha = \varphi$; $F^\beta = \varphi$, $F^{\alpha,\beta} = \varphi$
for m=n to 1{
if $C^m_k = \varphi$
{
Break;
}
for all $T \in P_m$
{
for all $X \in C^m_k$
{

$F(X).start = m$; $F(X).count$ += $X.count$;
}
for all $X \in F^\alpha$ *and* $X \in F^\beta$
{
if $X \in F^\alpha$ && $X \in F^\beta$
{
$F^{\alpha,\beta}(X).count = F^\alpha(X).count + F^\beta(X).count;$
}
else if $X \in F\alpha$ && $X \notin F\beta$
{
$F(X).type = \beta;$
}
}

VI EXPERIMENTAL ANALYSIS

For experimental analysis two previous approaches are used and compare the performance of the proposed approach. We used Visual Studio Dot net as front end and SQL server as back end. We have taken real life data set for experimental analysis form a electronic shop and apply the algorithms on 10000 records with 50 different items. We store all records in SQL server. We evaluate the performance of our algorithm and compare it with, PPM and proposed. The experiments were performed on 3rd Generation Core i5 Processor, 2GB Dedicated Graphics and 32GB of SSD Drive with Windows 8 Operating system and 4GB of DDR3 RAM. The 400GB of hard disk drive gives additional storage while the SSD drive helps to boost the starting time.

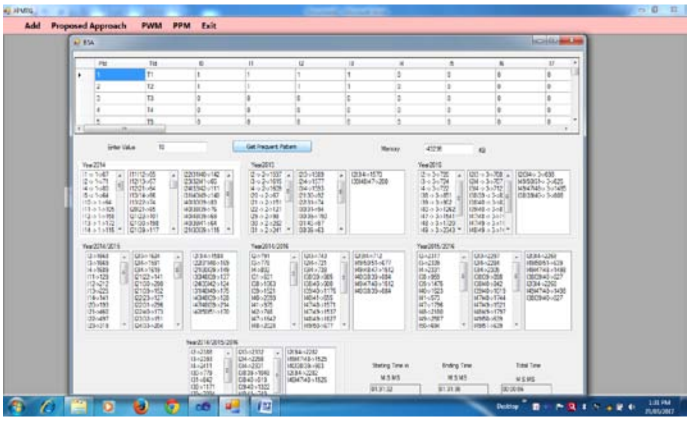

Figure 4 Rear Partition Scanning Approach

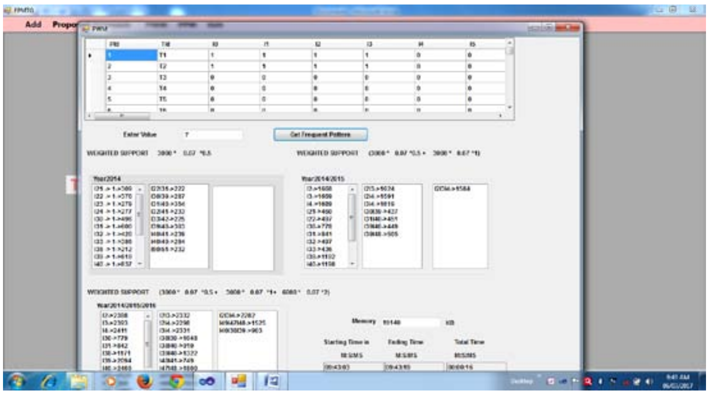

Figure 5 PWM (Progressive Weight Miner) approach

VII. RESULT AND COMPARISONS

Table 1 number of items in 2014 with 5 percent of support value. We compare the performance of the proposed algorithm with PWM (Progressive Weighted Miner). Table 1 show different number of item with one, two and three sets of items

Table 1 Number of items for 2014

Year 2014	PWM (Progressive Weighted Miner).	Proposed Approach
One item set	17	53
Two Item set	9	48
Three Item Set	0	14

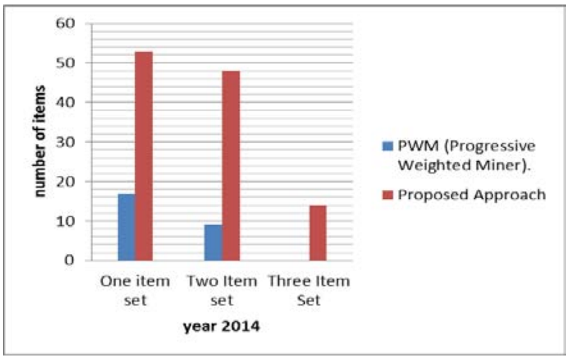

Figure 6 Number of item for 2014

Table 2 show number of item in 2014-15 with same support value. We compare the performance of the algorithm with PWM (Progressive Weighted Miner). Form the output it is clears that some of the item which present in both year and some present only in a particular year.

Table 2 Number of item for 2014-2015

Year 14-15	PWM (Progressive Weighted Miner).	Proposed Approach
One item set	15	28
Two Item set	7	36
Three Item Set	0	3

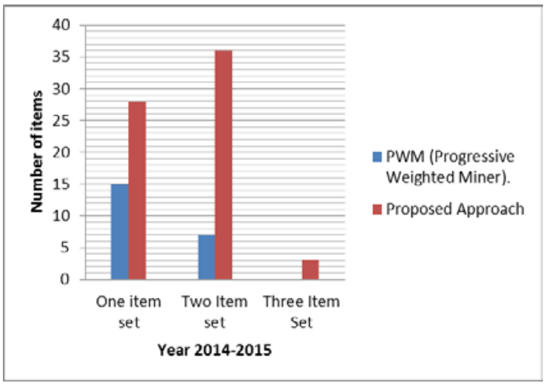

Figure 7 Number of item for 2014-2015

Table 3 show number of item in 2014-15-16 with same support value. Form the output it is clear that some of the item present in all year previous and current year they have the life time in entire database. Some of the item present in sub interval only

Table 3 Number of item for 2014-2015-2016

Year 14-15-16	PWM (Progressive Weighted Miner).	Proposed Approach
One item set	8	27
Two Item set	5	23
Three Item Set	1	5

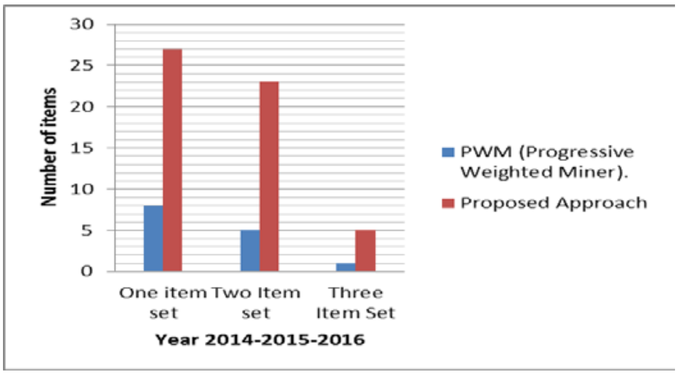

Table 4 show execution time in millisecond and number of records for proposed approach and PPM and PWM. Form the table it is clear that the proposed approach reduce execution time

Table 4 Number of item for 2014-2015-2016

Execution time	RPSA	PPM	PWM
1000	1024	3446	3646
5000	2786	5642	6432
10000	4892	7248	7472

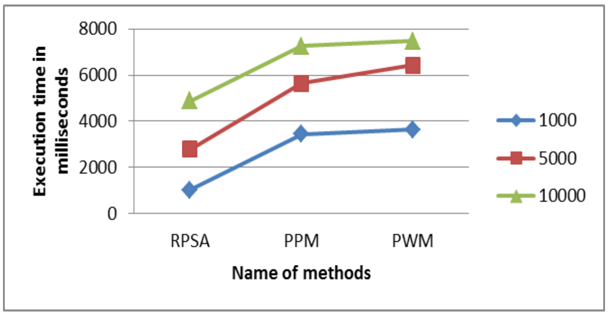

Figure 9 execution time and number of records

VII CONCLUSION AND FUTURE WORKS

After implantation of all three algorithm PPM PWM and RPSA we test these algorithm using different parameter like support count, number of record and memory used on the basis of support count . It clear that proposed algorithm perform well and compared to the PPM and PWM. Form the entire graph it is clear that proposed method outperforms others by avoiding the rescanning of the original databases. This investigation has presented a new method, proposed approach, for temporal frequent pattern mining. Proposed approach does not require the rescanning of the database and can determine new frequent itemsets at the latest time intervals. From the experimental analysis it is also clear that proposed approach reduce execution time greatly as compared to the PPM and PWM. In future we explore this approach for concepts hierarchy using month, and quarterly and weekly

IV. REFERENCES

[1]J. Han, M. Kamber, Data mining, Concepts and techniques, Academic Press, 2003.

[2]Arun K. Pujari, Data mining Techniques, University Press (India) Private Limited, 2006.

[3]B. Nath et al. “Discovering Association Rules from Incremental Datasets, International Journal of Computer Science & Communication Vol. 1, No. 2, July-December 2010.

[4]Tannu Arora et al. “Improved Association Mining Algorithm for Large Dataset”, IJCEM International Journal of Computational Engineering & Management, Vol. 13, July 2011.

[5]Chin-Chen Chang and Yu-Chiang Li, “An Efficient Algorithm for Incremental Mining of Association Rules” Proceedings of the 15th International Workshop on Research Issues in Data Engineering: Stream Data Mining and Applications 2005 IEEE.

[6]Tarek F. Gharib et al. “An efficient algorithm for incremental mining of temporal association rules”, An International Journal Expert System with Application 2010.

[7]Sheng Xiang Fan et al.“ Hybrid Temporal Pattern Mining with Time Grain on Stock Index” This work was supported by the National Science Council, Taiwan under grant NSC99-2632-E-126-001-MY3 and NSC99-2221-E-126-008 2011.

[8] Abhay Mundra et al. “Rapid Update in Frequent Pattern form Large Dynamic Database to Increase Scalability”, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-6, January 2013

[9]Anour F.A. et al. “IMTAR: Incremental Mining of General Temporal Association Rules”, Journal of Information Processing Systems, Vol.6, No.2, June 2010.

[10]N. Pughazendi et al. “Temporal databases and frequent pattern mining techniques”, International Journal of P2P Network Trends and Technology July to Aug Issue 2011

[11]Iyad Batal et al. Mining Recent Temporal Patterns for Event Detection in Multivariate Time Series Data KDD’12, August 12–16, 2012, Beijing, China. Copyright 2012 ACM 978-1-4503-1462-6 /12/08.

[12]Sílvia Moura et al. “(TD)2PaM: A Constraint-Based Algorithm for Mining Temporal Patterns in Transactional Databases Correia, L.P. Reis, and J. Cascalho (Eds.): EPIA 2013, LNAI 8154, pp. 390–407, 2013. © Springer-Verlag Berlin Heidelberg 2013

[13]Siddharth Shah et al. “Incremental Mining of Association Rules: A Survey”, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (3) , 2012,4071-4074

[14]Gurram Sunitha et al. “Mining Frequent Patterns From Spatiotemporal Data Sets: A Survey”. Journal of Theoretical and Applied Information Technology 20th October 2014. Vol. 68 No.2 © 2005 - 2014 JATIT & LLS. All rights reserved.

[15]Vangipuram Radhakrishna et al. “A Novel Approach for Mining Similarity Profiled Temporal Association Patterns” Rev. Téc. Ing. Univ. Zulia. Vol. 38, Nº 3, 80 - 93, 2015

[16]Yi-Cheng Chen et al. “Mining Temporal Patterns in Time Interval-Based Data” IEEE Transactions on Knowledge and Data Engineering · December 2015.

[17]J. Mercy Geraldine et al. “Weighted Temporal Pattern Mining With Dimensionality Reduction Using Modified AFCM Technique” International Journal of Advanced Engineering Technology E-ISSN 0976-3945 2016

