Journal of Theoretical and Applied Information Technology
10th September 2013. Vol. 55 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN EFFICIENT METHOD FOR MINING PERIODIC TRANSITIONAL PATTERNS IN TRANSACTION DATABASE

[1]S.SRIDEVI, [2]Dr.S.RAJARAM
[1]Department of CSE,Thiagarajar College of Engg, Madurai,Tamilnadu,India
[2]Department of ECE, Thiagarajar College of Engg, Madurai,Tamilnadu,India
E-mail: [1]sridevi@tce.edu, [2]rajaram_siva@tce.edu

ABSTRACT

Transaction database is used to find frequent itemsets from large datasets along with their associated timestamps. The existing frequent pattern mining algorithms such as Apriori, FP-growth, Partition, Pincer Search do not consider the timestamps associated with the transactions. In real time transactions, without timestamps it is difficult to identify the constantly changing behaviour of the frequent patterns in a transaction database. To overcome the above problem, EP-TP (an Efficient Periodic Transitional Patterns) mining method is introduced in this paper. Transitional patterns are used to discover transi-frequent patterns along with their time stamps in a transaction database. The above patterns include both positive and negative transitional patterns. A pattern is said to be a positive or negative transitional pattern if their frequencies dramatically increase or decrease respectively at some point of time in a transaction database. From the above, the business owners can able to analyze the trend of the pattern(s) over the period(s). This trend is used to find out the nature of the pattern and hence the business owner can able to make the necessary step to improve the performance (or analyze the reason) of the negative pattern. The above pattern is tested with Textile Dataset and its performance was compared with existing algorithm. It is also present the experimental study to verify the usefulness and effectiveness of transitional patterns.

Keywords: *Transaction database, Transitional pattern, Frequent pattern, Periodic pattern, Linear pattern*

1. INTRODUCTION

Data mining is the computer-assisted process of digging through and analyzing enormous amount of data and then extracting the meaning of the data. It is concerned with analyzing large volumes of unstructured data to discover interesting regularities or relationships which in turn lead to better understanding of the underlying processes. With recent tremendous technical advances in processing, storage capacity, combination of computer technology and business motive, data mining is seen as an increasingly important tool by modern business to transform unprecedented quantities of digital data into business intelligence. It can be applied in wide range of decision making process, profiling practices such as marketing, analyzing future trends, fraud detection, and Bio-Informatics.

Data mining tools are used to forecast behaviours and trends, allowing business owners to make logistic and knowledge-driven decisions. Traditional Data mining tools can answer business questions that traditionally were too time consuming to resolve because it contains large no of samples. They scour databases for hidden and frequent patterns, finding useful information that experts may miss because it lies outside their expectations. Although data mining or sub-division of data mining is still in its infancy, some wide range of industries – including stock market, retail, finance, health care, and aerospace - are already using data mining tools and techniques to take advantage of historical data. By using machine learning or soft computing technologies, statistical and application of mathematical techniques to sift through warehoused information, data mining helps analysts to recognize association or relationships, trends, frequent patterns, exceptions and anomalies that might otherwise go unnoticed.

Temporal data mining is concerned with such analysis in the case of ordered and unordered data streams with temporal interdependencies. Over the last decade many interesting and novel techniques of temporal data mining were proposed (Ale and Rossi, 2000) and shown to be useful in large applications. Since Spatial or temporal data mining brings together techniques from different fields such as statistics, machine learning, Fuzzy Logic, artificial Intelligence and databases, the literature is sprinkled among many different

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

sources. It is the subdivision of data mining which is defined as extraction of knowledge or information from the data base with respect to the time information. Temporal data mining uses temporal or time series databases.

The focus of this paper is to used to find frequent patterns with respect to time stamp. Most of the existing algorithms like Apriori(Agrawal and Srikant,1994), FP-growth (J.W.Han et al, 2004), Partition, Pincer Search are also used to find frequent itemsets but they didn't consider time stamp. The above algorithms are an influential algorithm for mining frequent item sets for association rules. The user may also interested to find Maximal frequent itemsets(Burdick et al., 2005), closed frequent itemset (Pei, Han and Mao, 2000) and indirect associations(Tan and Kumar, 2002). Mining maximal or closed frequent itemsets which greatly reduces the number of generated patterns by generating only the largest frequent itemsets. Indirect associations are closely related to negative associations in which there is no correlation between item sets generated. They are used to detect "infrequent item pairs that are highly expected to be frequent" without using any domain knowledge.

Pattern mining(Qian Wan and Aijun An, 2009) is a data mining technique that involves finding existing patterns from the transaction database. In this context patterns often means association. The original inspiration for searching association rules came from the desire to analyze supermarket transaction data. It can be used to examine customer behavior in terms of the purchased products along with their time stamps. Transactions in a database environment have two main purposes: To provide reliable units of work that allow correct recovery from failures and keep a database consistent even in cases of system failure, when execution stops (completely or partially) , Power-off and many operations upon a database remain uncompleted, with undefined status. To provide segregation between programs accessing a database concurrently and it must be an atomic, consistent, isolated and durable.

Qian Wan and Aijun An (2009) suggests that finding transitional patterns along with their significant milestones using TP (Transitional Pattern)-Mine algorithm. Transitional patterns are used to discover change over in frequent patterns along with their time stamps. These timestamps are usually called as significant milestones. These milestones are used to find time points at which the frequency of the pattern changes (from increasing to decreasing order or from decreasing to increasing order) most significantly. Consider an example database as shown in Table 1 which has 10 transactions of seven items. Each transaction is provided with a timestamp. These timestamps are arranged either in an ascending order or in descending order. However, interesting differences between these patterns can be found after we consider the time information of each transaction in the database, as shown in the third column of Table 1.

Table 1 An example Transaction Database

TID	LIST OF ITEMS	TIMESTAMP
1	G,E,A,B	Jan 2011
2	A,B,C,D	Feb 2011
3	A,B,D	Mar 2011
4	D,C,B	Apr 2011
5	A,B,C,D,E,G	May 2011
6	G,D,B	Jun 2011
7	A,B,E,G	Jul 2011
8	C,D,G	Aug 2011
9	A,B,D	Sep 2011
10	D,E,G	Oct 2011

Transaction database contains all the transactions from January 2011 to October 2011, one transaction per month. For example the pattern AB occurs every month (and including) before March 2011. But after March 2011, AB only occurs three times in seven transactions, which is equivalent to a frequency of 33 percent. That is to say that the frequency or support of pattern AB decreases significantly after March 2011. Our aim is to find transitional patterns along with their significant milestones. This pattern may be either Linear or Non-Linear Transitional Pattern. The above pattern AB is an example of Non-linear Transitional Pattern because whose occurrence is in non-linear fashion. These patterns are used in enormous applications. For example, in the super market scenario this pattern allows the business owners to identity which product or combinations of products that have recently become more and more popular or not so popular as before. Based on

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

above, the owner can able to know the trend of marketing strategy in retail environments.

The rest of this paper is organized as follows. Section 2 describes related work, section 3 defines proposed method. Research framework and Experimental result of the proposed method are reported in section 4 and section 5. Section 6 covers conclusion and future work.

2. RELATED WORK

The objective of this work is to find the transitional pattern along with the time information. Agrawal and Srikant(1994) proposed frequent pattern algorithm namely Apriori algorithm for mining frequent itemsets which can be used to generate association rules. The advantage in Apriori algorithm is the frequent pattern set is determined correctly and it uses large itemset properly. The drawback is it requires too many database scans. Problem occurs when the database is very large since it involves multiple scans over the database. J.-W. Han et al., (2004) proposed another frequent pattern algorithm namely FP-growth (Frequent Pattern growth) which uses an extended prefix-tree (FP-tree) structure to store the database in a compressed form. FP-growth uses divide-and-conquer approach to decompose both the mining tasks and the databases. The advantage of FP-Growth algorithm is to avoid the costly process of candidate generation and testing used by Apriori.The disadvantage is complexity in building the FP-Growth tree for very large datasets.

Eclat algorithm was proposed by Zaki et al., (2003). It is a depth-first search algorithm which uses set intersection. The basic idea for the Eclat algorithm is it uses transactions set intersections to compute the support of a candidate itemset avoiding the generation of subsets that does not exist in the prefix tree. Fast algorithms for mining emerging patterns by J.Bailey, T.Manoukian and K.Ramamohanarao (2006) said Emerging Patterns are itemsets whose supports change significantly from one dataset to another. It can be useful as a means of discovering distinctions inherently present amongst a collection of datasets and have been shown to be a powerful technique for constructing accurate classifiers.

An approach to discover temporal association rules was proposed by J.M.Ale and G.H.Rossi, (2000). The goal of discovering association rules is to discover all possible associations that accomplish certain restrictions (minimum support and confidence and interesting). However, it is possible to find interesting associations with a high confidence level but with little support. Efficiency is achieved by expanding the notion of association rules incorporating time to the frequent itemsets discovered.

The framework for discovering frequent episodes in sequential data was proposed by Q.Wan and A.J.Wan, (2006).An episode is a collection of events that occur relatively close to each other in a given partial order. An episode can be serial or parallel. In serial episode, events occur in a sequence, and for parallel, there are no constraints are posed on the relative order of the events. The paper also presented an algorithm for producing rules that describe the associations between the discovered frequent episodes. Xin Feng et al., (2009) presents work on mining multiple temporal patterns of complex dynamic data streams which can be used to update the database online and hence find out the multiple temporal patterns. There could be some drawbacks when potential frequent patterns are long.

A general model for mining asynchronous periodic patterns in temporal Databases has been developed by Kuo-Yu et al.,(2005). Their system provides a general model for asynchronous periodic patterns with the help of a four-phase algorithm which is devised to discover periodic patterns from a time series databases presented in vertical format. Here the input of the consecutive algorithm depends on the output of the previous algorithm. So if any one of the algorithm is inefficient it will affect the entire performance.

The discovery of hidden periodic patterns in spatiotemporal data could unveil important information to the data analyst. Existing approaches for finding periodic patterns focus on symbol sequences. However, these methods cannot directly be applied to a spatiotemporal sequence because of the fuzziness of spatial locations in the sequence. Huiping Cao et al.,(2007) define the problem of mining periodic patterns in spatiotemporal data and proposed an effective and efficient algorithm for retrieving maximal periodic patterns. Recent research literature has explored the sequential patterns on transaction data and trajectory analysis on moving objects. However, the above methods cannot be directly applied to mining sequential patterns from a large number of spatio-temporal events.

Qian Wan and Aiijun (2009) proposed an algorithm called TP-Mine algorithm for discovering patterns that dramatically changes over time. These types of pattern are called Transitional Patterns that capture the constantly changing attitude of frequent patterns in a transaction database. These patterns

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

are used to discover frequent patterns along with their time stamps in a transaction database.

3. DESIGN METHODOLOGY

The motivation behind this work is to discover the efficient and periodic transitional patterns so that the business owners can able to analyze the trend of the pattern(s) over the period(s). This trend is used to find out the positive and negative pattern(s) and hence the business owner can able to make the necessary step to improve the performance (or analyze the reason) of the negative pattern. The proposed system is divided into three phases. Phase 1consists of Frequent Pattern generation with or without periodicity. Frequent Pattern generation with periodicity can be achieved by SMCA (The abbreviation was shown in Figure 2) Model and without periodicity can be achieved by FP-Growth.

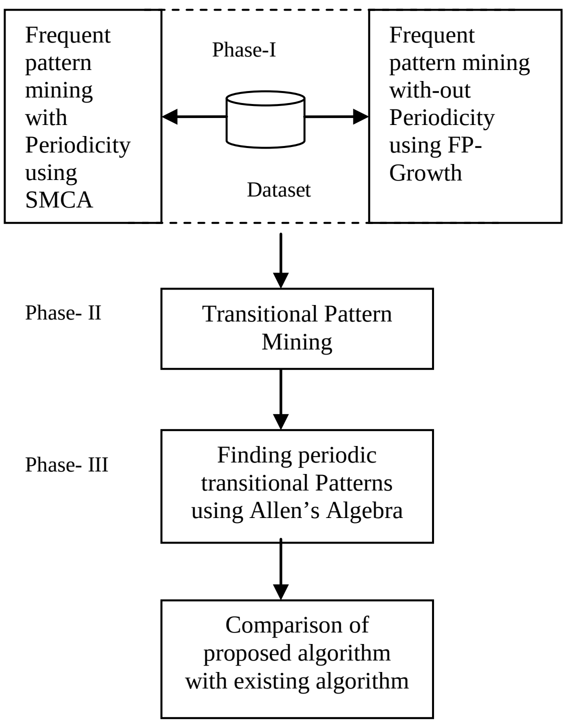

Figure 1: Framework for the proposed system

Phase-2 involves mining of transitional patterns. Phase- 3 covers detection of periodic transitional patterns using Allen's interval. The framework for the proposed method is shown in Figure 1.

3.1 Frequent Pattern Mining

Frequent Pattern can be mined from transaction database with and with-out periodicity. Mining with periodicity can be done by FP-Growth and without periodicity can be done by SMCA Model.

3.1.1 Frequent pattern mining without Periodicity

The transaction database contains millions of instances. This information is arranged either in an ascending order or in descending order with respect to time stamp. As an end user they are not interested in all the items of the dataset. They may interest in frequent or infrequent pattern(s). For that, the support is calculated for the item(s) using table 1. In this paper, we used FP-growth to retrieve the support for the frequent patterns. This algorithm needs only two database scans for calculating support value. This can be applied to find transitional patterns.

3.1.2 Frequent Pattern mining with periodicity using SMCA Model

In this part, a 4-phase algorithm (Kuo-Yu et al.,2005) namely Singular Periodic Pattern Mining (SPMiner), Multievent Periodic Pattern Mining (MPMiner), Complex Periodic Pattern Mining (CPMiner), and Asynchronous Sequence Mining (APMiner) are used to discover periodic patterns from a time series databases. The corresponding diagram was shown in Figure 2.This model uses only vertical dataset. Hence vertical dataset which was shown in Table 2 is constructed from Horizontal Dataset was shown in Table 1. The various notations and their meanings used in this phase are shown in table 3 which was included at the end of the paper.

Table 2 An example Transaction Database presented in Vertical Format

ITEM	TIMESTAMP for year 2011
A	Jan , Feb,Mar, May,Jul,Sep
B	Jan , Feb,Mar, Apr, May, Jun,Jul,Sep
C	Feb,Apr, May, Aug
D	Feb,Mar, Apr, May, Jun,Aug,Sep, Oct
E	Jan , May, Jul Oct
G	Jan , May, Jun,Jul

Singular Periodic pattern mining (SPMiner)

SPMiner discovers valid segments for each single event. The inputs to SPMiner include a vertical format and the interested period interval specified by Lmin(=1) and Lmax, which is

Journal of Theoretical and Applied Information Technology
10th September 2013. Vol. 55 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

specified by the user. Two mining strategies, Potential Cycle Detection (PCD) and Hash-Based Validation (HBV) are used for mining periodic segments for each single event. PCD is designed to recognize possible periods for an event. This can be implemented by scanning the timeliest for an event once. HBV algorithm scans the timelist once and outputs valid segments with period p. Here segments represent synchronous periodic occurrences and it can be overlapped.

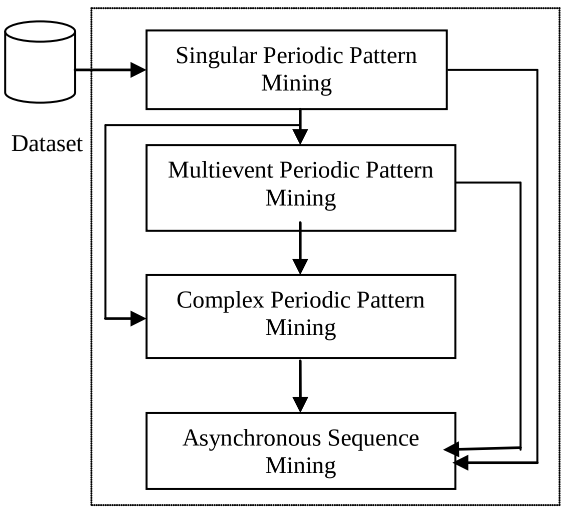

Figure 2: Phase-I Diagram for SMCA Model

Multievent Periodic pattern mining (MPMiner)

To discover valid segments for multievent 1-patterns, two mining methods: Timelist-Based Enumeration (TBE) and Segment-Based Enumeration (SBE) are used. In TBE segments are generated for each combined event set and the timelist is obtained by the timelist intersection from the constituent events. Then, the HBV procedure used in SPMiner is applied to check if valid segments exist for the event set. Enumeration stops whenever no valid segment exists for an event set. In SBE, two segments can be combined if they have the same offsets and the overlapping area has repetitions greater than min rep. The overlapped area is defined by the maximum start position and the minimum end position of the two segments.

Complex Periodic pattern mining (CPMiner)

Discovering complex patterns from singular patterns (both single-event and multievent) has a procedure similar to the segment-based enumeration (SBE) in MPMiner. CPMiner enumerates possible combinations of valid segments of the same period in depth-first order and checks if a combination forms a complex pattern. For two overlapping segments with different offsets, they can form a 2-pattern if the repetition of the overlapping area is greater than min _rep. In other words, an i-pattern is composed of i segments (of 1-patterns) with different offsets.

Asynchronous sequence mining (APMiner)

An asynchronous periodic pattern is defined by the existence of a valid sequence, which is a set of non overlapping valid segments with respect to a pattern. Therefore, a depth-first algorithm is designed to enumerate all combinations of segments with respect to a pattern. Segments are ordered by their start position. If the start position of the new segment is within max_dis of the current subsequence, then the subsequence is extended. This SMCA model generates the frequent pattern(s) along with the periodicity and repetitions.

3.2 Transitional Pattern Mining

Transitional pattern (Qian Wan and Aiijun, 2009) can be mined either by SMCA Model or FP-Growth. The various notations and their meanings used in this phase are shown in Table 3 which was included at the end of the paper. Some basic notations that are used in this mining are as follows: The transitional ratio of pattern X at its ith milestone in D is given by

$$tran^i(X) := \frac{sup_+^i(X) - sup_-^i(X)}{MAX(sup_+^i(X), sup_-^i(X))} \quad (1)$$

$where\ 1 \le i \le ||cov(X)||$(Cov(X) is the no of transactions in D that supports pattern X)

Where $sup_-^i(X)$ and $sup_+^i(X)$ are the support of a pattern X before its ith milestone in D and the support of a pattern X after its *i*th milestone in D.

$$sup_-^i(X) := \frac{i}{\rho(\tau^i(X))} \quad (2)$$

$$sup_+^i(X) := \frac{||cov(X)|| - i}{||D|| - \rho(\tau^i(X))} \quad (3)$$

The support value 1 indicates the pattern X is occurred frequently in the corresponding milestone. This value is less than 0.5 indicates, it is occurred infrequently. The ith milestone of a pattern X is denoted by,

$$\xi^i(X) := \frac{\rho(\tau^i(X))}{||D||} \times 100\%, \quad (4)$$

$where\ 1 \le i \le ||cov(X)||$

The lesser percentage of this value indicates that the corresponding pattern X occurs quickly in Dataset D. The transitional ratio is used to measure the difference of a pattern's frequency before and after its ith milestone. The value of a transitional ratio is between -1 and 1. A transitional pattern is a pattern whose absolute values of transitional ratio are large.

Journal of Theoretical and Applied Information Technology
10th September 2013. Vol. 55 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A pattern X is said to be a transitional pattern if there exists at least one milestone of X, $\xi^i(X) \in T_\xi$ such that:

1. $sup_-^i(X) \geq t_s$
2. $sup_+^i(X) \geq t_s$ and $|tran^i(X)| \geq t_t$ (5)

Where t_s and t_t are support threshold and transitional pattern threshold which are specified by user. The Pattern X is said to be positive transitional pattern when $tran^i(X)$>0 and it is said to negative transitional pattern when $tran^i(X)$ <0. Based on the above Linear and Non-linear transitional pattern are defined.

3.2.1 Linear and Non-Linear Transitional Pattern Mining:

Based on the above definitions, transitional pattern is classified into Linear and Non-Linear pattern which are described as follows:

Linear pattern is a pattern whose values may be either gradually increasing or decreasing by some amount. Transitional ratio at different milestones is found using (1). If all the Transitional ratio values are positive(negative) and whose values are increased(decreased) gradually and which satisfies equation (5) then that pattern is considered as a Linear Positive(negative) Transitional Pattern.

If the values of transitional ratio are combination of positive and negative and which satisfies equation (5) then that pattern is considered as a non-Linear Transitional Pattern. The time point at which, the frequency of the pattern changes can be found by using both Significant Frequency ascending and descending Milestones.

3.2.2 Significant Milestone

Significant milestones are used to find time points at which the frequency of the pattern changes most significantly. These milestones may be frequency ascending or descending milestone.

The Significant Frequency ascending(descending) Milestone of a positive(negative) transitional pattern should satisfy following two conditions:

$a).\, sup_-^M(X) \geq t_s$ (or) $b).\, sup_+^M(X) \geq t_s$ (6)

$\forall \xi^i(X) \in T_\xi, a), tran^M(X) \geq tran^i(X)$

(Or) $b).\, tran^M(X) \leq tran^i(X)$ (7)

Where $tran^M(X)$ is the transitional ratio of particular pattern P and $tran^i(X)$ is the current maximal transitional ratio of pattern P. If the milestone satisfies equation (6a) and (7a), then the tuple $(\xi^M(X), tran^M(X))$ will be added to the ascending Milestone set. If the milestone satisfies equation (6b) and (7b), then the tuple $(\xi^M(X), tran^M(X))$ will be added to the descending Milestone set.

3.3 Periodic Transitional Pattern Mining Using Allen's Interval Algebra

As the user/owner may not interest in the entire transitional pattern, they may be interested in periodic transitional patterns. This can be achieved by using Allen's interval algebra. Allen's described 13 relationships, among these, this paper uses only 3 relationships namely equals, overlaps and contains. Transitional pattern is said to be periodic if the patterns satisfies any one of the relationship which are checked against consecutive years.

3.4 Performance Measurement

The proposed algorithm EP-TP was tested with existing algorithm in terms of minimum support threshold, scalability and running time.

4. RESEARCH FRAMEWORK

4.1 Dataset Selection for the Proposed System

The various real and synthetic dataset that can be used for pattern mining was shown in table 4. The table contains number of transactions and number of items. From which Efficient and periodic Transitional Patterns (EP-TP) was obtained for each dataset. The concept of this paper was tested by using Textile dataset which contains 500 items and 19,550 transactions.

Table 4: Dataset Characteristics

Name of the Dataset	Number of items	Number of Transactions	Number of frequent pattern using Transitional pattern mining
IBM-Artificial	10,000	2,500,435	1658
BMS-POS	1,657	3,367,020	1850
BMS-webview-1	497	1,49,639	479
Retail	16,470	88,163	500
Textile	500	19,550	105

4.2 Input Selection for Transitional Pattern Mining

Phase I: Generation of frequent patterns can be done by FP-growth or SMCA Model. FP-Growth uses horizontal dataset where SMCA Model uses vertical format.

Phase II: The output of FP-growth or SMCA model is given to input of the Transitional Pattern Mining. If the end user is interested only in

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

frequent patterns but not in period then, prior algorithm is used rather than using SMCA model.

4.3 Proposed Algorithm: EP (Efficient and Periodic) Transitional Pattern Mining Algorithm

Steps Involved

1. Frequent pattern generation using SMCA Model or FP –Growth.
2. Generate i^{th} milestone of each frequent pattern X. Each ith milestone much belongs to T_{ξ} ={ lower limit%, upper limit %} that can be specified by the user. If the milestones that don't belong to T_{ξ} , drop that milestone.
3. Find out support of a pattern X before its i^{th} milestone $sup_{+}^{i}(X)$ and after its i^{th} milestone$sup_{-}^{i}(X)$.
4. Check $sup_{+}^{i}(X) \geq t_s$ and $sup_{-}^{i}(X) \geq t_s$, If not so drop the pattern X.
5. Generate Transitional Ratio for all the patterns of step 4.
6. Generate Periodic pattern using Allen's algebra.

5. EXPERIMENTAL RESULTS

In this section, experimental results on textile dataset was summarized. Then comparison between our proposed approach and other related approaches is made. Next, the differences between TP (Transitional Pattern) and EP-TP (Efficient and Periodic Transitional Pattern) are highlighted.

The sample real dataset i.e textile dataset was shown in Figure 3. The support for all the items in the specified range is calculated using frequent pattern mining algorithm (FP-growth or SMCA Model). No of database scan for generating frequent patterns is 2 and no of database scan for generating transitional patterns is one. Thus total no of database scan for generating the entire pattern is three. A TP-Mine algorithm calculates support for all the retrieved items of the specified minimum and maximum limits. It will display only the items whose support value is greater than minimum support threshold. This was shown in Figure 4 and 5.The advantage of FP-Growth algorithm is it is scalable and efficient to generate long and short frequent patterns. One more advantage of this algorithm is it needs only two scan over the database where as Apriori algorithm requires multiple scans over the database and it needs long candidate generation. In FP-Growth algorithm there is no need to generate the candidate.

Figure 3: Sample Dataset

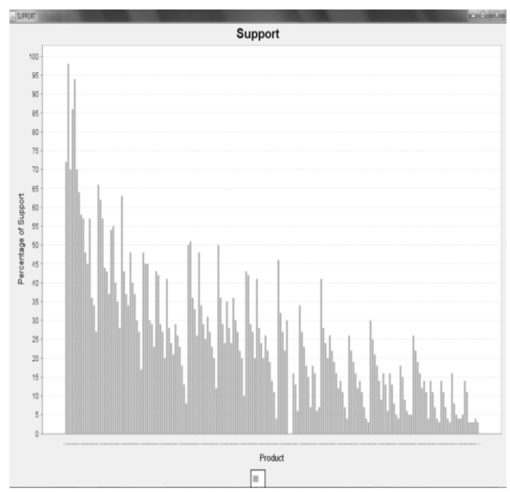

Figure 4: Support of Textile Product

After generating the frequent itemset, the support before the milestone and after the milestone is calculated. This value equals one indicates that the corresponding pattern occurs frequently (mostly the pattern is present in all transactions). This value equals 0.5 indicates which is partial frequent pattern (Half of the transaction database contains the pattern X w.r.to the milestone). If this value is less than 0.5 or equal to zero indicates, the pattern X is less frequent hence it can be rejected or can be for infrequent pattern mining.

Based on support threshold and transitional pattern threshold transitional pattern was found. The pattern may be linear or non-linear pattern. The pattern is said to be negative

Journal of Theoretical and Applied Information Technology
10th September 2013. Vol. 55 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

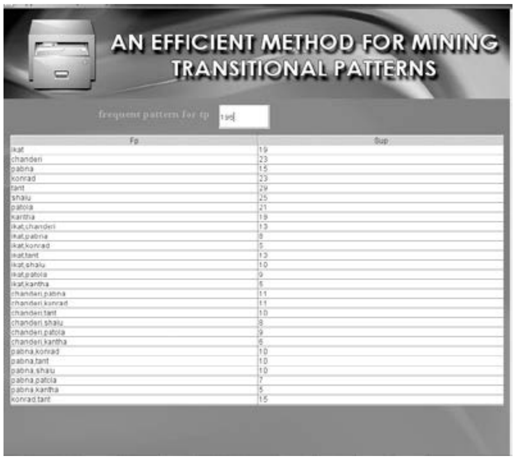

Figure 5: Frequent pattern generation using FP- growth

transitional pattern if the value lies between -1 to 0 and the pattern is said to be positive transitional pattern if the value lies between 0 to 1.Non-linear pattern consists of combination of both. Figure 6 depicts an example of Non-linear pattern. Then it finds the transitional ratio in which it measures the patterns frequency before and after its milestone. Calculation of negative ratio and their milestone was shown in Figure 7. From this we can able to find out the Top 10 negative pattern by arranging them in descending order. The advantage of the proposed algorithm is it eliminates the patterns which are beyond the range in the consecutive steps. The range can be fixed by the end user. For example if the range specified by the user is [35%, 75%]. If any of the pattern (positive or negative) which falls beyond the range is automatically eliminated, since the user is not interested in that range. From the above, one can deduce the execution time of the pattern in consecutive steps, yields reduction in total execution time of the whole process.

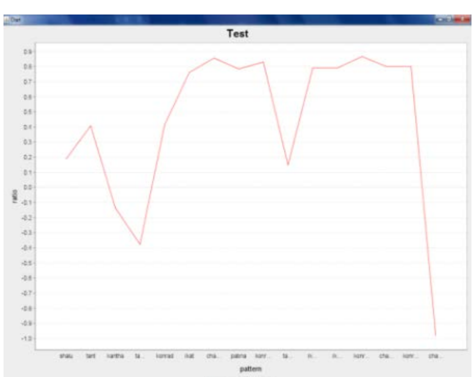

Figure 6: Non-Linear Transitional Pattern

The other benefit of the proposed algorithm is the user can automate their process i.e. import their products in advance if there were positive patterns found in the previous years or analyze the reason for negative patterns found in the previous years. The algorithms were implemented on Java language and tested with Intel core 2 duo processor windows 2007 with 2GB of main memory.

Figure 7: Calculation of negative ratio and their milestone

5.1 Scalability

To study the efficiency and scalability of the proposed work, the set of experiments was conducted on different datasets. Figure 8 shows that running time of textile dataset. If we increase the no of transaction then its running time is also get increased. Scalability on Textile dataset was shown in figure 9. From this figure it is understood that running time is less for large support threshold value. Run time Comparison of the proposed algorithm EP-TP with existing TP algorithm was shown in figure 10. It is proved that the proposed algorithm gives better performance than existing algorithm [17].

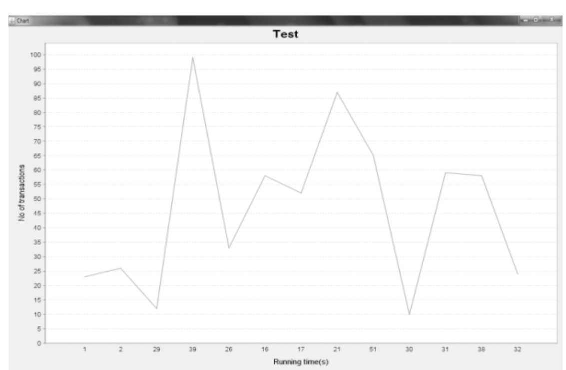

Figure 8: Running time of Textile dataset.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

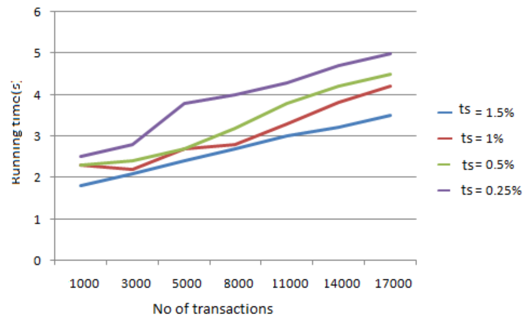

Figure 9: Scalability on Textile Dataset

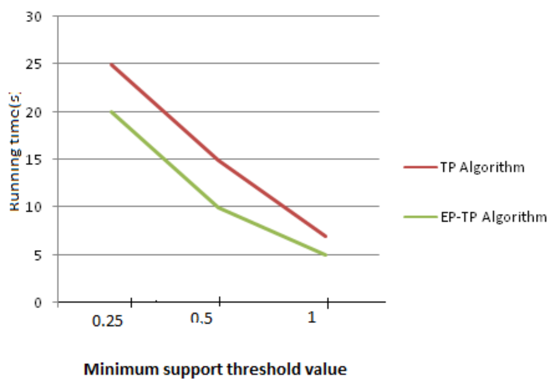

Figure 10: Run time Comparison on Textile Dataset

6. CONCLUSION AND FUTURE WORK

In this paper, an efficient linear and non-linear periodic transitional pattern for transaction databases have been designed and implemented. The major contribution of this work is the provision of facilities to identify the periodic transitional pattern and to find the time at which the frequency of the pattern changes most significantly. This system has been implemented using Java Language and hence it is platform independent. In many real-world problems, it may be more effective to compare transitional pattern mining with other approaches such as synchronous and asynchronous periodic pattern mining. The major contributions of this work are Comparison of EP-TP with TP algorithm along with their milestones in transaction database. It is proved that the proposed algorithm gives better performance than existing algorithm. This application was tested using textile dataset. Experimental results shows that usefulness and effectiveness of transitional patterns.Future work will focus on to compare the performance of transitional pattern with Sequential, synchronous and asynchronous periodic pattern mining and reveal which one is better.

REFERENCES

[1]. Aggarwal, C.C"A Framework for Diagnosing Changes in Evolving Data Streams", *Proc. 2003 ACM SIGMOD Int'l Conf.Management of Data* ,2003, pp. 575-586.

[2]. Agarwal, R.C., Aggarwal, C.C. and Prasad, V.V.V, "Depth First Generation of Long Patterns", *Proc. Sixth ACM SIGKDD Int'l Conf.Knowledge Discovery and Data Mining*,2000, pp. 108-118.

[3]. Agrawal, R., Imieli_nski, T. and Swami, A., "Mining Association Rules between Sets of Items in Large Databases", *Proc. 1993 ACMSIGMOD Int'l Conf. Management of Data (SIGMOD '93*), pp. 207- 216.

[4]. Agrawal, R. and Srikant, R.,"Fast Algorithms for Mining Association Rules", *Proc. 20th Int'l Conf. Very Large Data Bases*,1994,pp. 487-499.

[5]. Agrawal, R. and Srikant, R, 'Mining Sequential Patterns', *Proc. 11th Int'l Conf. Data Eng*.1995,, pp. 3-14.

[6]. Ale, J.M. and Rossi, G.H., "An Approach to Discovering Temporal Association Rules", *Proc. 2000 ACM Symp. Applied Computing*, 2000, pp. 294-300.

[7]. Bailey, J., Manoukian, T. and Ramamohanarao, K.,"Fast Algorithms for Mining Emerging Patterns", *Proc. Sixth European Conf.Principles of Data Mining and Knowledge Discovery*,2002, pp. 39-50.

[8]. Bay, S.D. and Pazzani, M.J, "Detecting Change in Categorical Data: Mining Contrast Sets", *Proc. Fifth ACM SIGKDD Int'l Conf.Knowledge Discovery and Data Mining*,1999, pp. 302-306.

[9]. Bayardo, J.R,"Efficiently Mining Long Patterns from Databases",*Proc. 1998 ACM SIGMOD Int'l Conf. Management of Data 1,998, pp. 85-93.*

[10]. Brijs, T., Swinnen, G., Vanhoof, K. and Wets, G, "Using Association Rules for Product Assortment Decisions: A Case Study", *Proc.Fifth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining*,1999, pp. 254-260.

[11]. Burdick, D., Calimlim, M., Flannick, J., Gehrke, J. and Yiu, T, "Mafia: A Maximal Frequent Itemset Algorithm", *,IEEE Trans.Knowledge and Data Eng., vol. 17, no. 11,2005, pp. 1490-1504.*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[12]. Han J.W, Pei J,Yin Y.W, and Mao R.Y, "Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern TreeApproach", *Data Mining and Knowledge Discovery*, vol. 8, no. 1,2004, pp. 53-87.

[13]. Huiping Cao, Nikos, and David W Cheung, "Discovery of Periodic Patterns in Spatiotemporal Sequences", *IEEE Trans on Knowledge and Data Engg.Vol* 19,2007,pp.453-467.

[14]. Jovonovic et al.,"Foundations of Predictive Data Mining",,*IEEE Transactions on Knowledge and Data Engg*, Vol 1, 2002,pp.439-450.

[15]. Kuo-Yu Huang and Chia-Hui Chang, "SMCA: A General Model for Mining Asynchronous Periodic Patterns in Temporal Databases", *IEEE Trans on Knowledge and Data Engg.*17,2005,pp. 774-785.

[16]. Pei. J, J. Han, and Mao.R.Y. ,"CLOSET: An Efficient Algorithm for Mining Frequent Closed Itemsets", *Proc. ACM SIGMOD Workshop research Issues in Data Mining and Knowledge Discovery*, 2000,pp. 21-30.

[17]. Qian Wan, Aijun An, (2009) "Discovering Transitional Patterns and their Significant Milestones in Transaction Database", *IEEE Transactions on Knowledge and Data Engineering, Vol 21, No 12*,2009,pp.1692-1707.

[18]. Tan, P.N and Kumar, V., "Mining Indirect Associations in Web Data", *Proc. Revised Papers from the Third Int'l Workshop Mining Web Log Data Across All Customers Touch Points*,2002,pp. 145-166.

[19]. Wan, Q. and An, A.J., "An Efficient Approach to Mining Indirect Associations",Journal of *Intelligent Information Systems*, vol. 27, no. 2,2006, pp. 135-158.

[20]. Xin Feng, and Odilon, K.S.,"Mining multiple temporal patterns of complex dynamic data streams", *IEEE Symposium on Computational Intelligence and Data Mining*,2009,pp. 411-417.

[21]. Zaki, M.J. and Gouda, K., "Fast Vertical Mining Using Diffsets", *Proc. Ninth ACM SIGKDD SIGKDD Int'l Conf. Knowledge Discovery and Data Mining)*,2003, pp. 326-335.

Table 3: Notations and Their Meanings

D	Dataset
min_support	minimum support value
FP –Growth	Frequent Pattern Growth Algorithm
SMCA	Singular Multievent Complex and Asynchronous periodic pattern mining
PCD	Potential Cycle Detection method
HBV	Hash-Based Validation
TBE	Time list- Based Enumeration
SBE	Segment-Based Enumeration
X	pattern(s) in Dataset D
$\boldsymbol{cov}(X)$	the Cover of pattern X in D
$\sup(X)$	the Support of pattern X in D
$sup_{-}^{i}(X)$	the Support of pattern X after i^{th} milestone
$sup_{+}^{i}(X)$	the Support of pattern X before i^{th} milestone
$\boldsymbol{\rho}(T)$	the Position of Transaction T in D
τ^{i}	the ith transaction of pattern X in D
$\xi^{i}(X)$	the ith milestone of pattern X in D
T_{ξ}	the Range of $\xi^{i}(X)$ in D
t_s	pattern support threshold
t_t	transitional Pattern threshold
EP-TPM	Efficient and Periodic Transitional Pattern Mining