e-ISSN:2319-9873
p-ISSN:2347-2324

RESEARCH AND REVIEWS: JOURNAL OF ENGINEERING AND TECHNOLOGY

TP-Mine: An Approach to Determine the Transitional Patterns and their Significant Milestones.

Radhika Katkum[1,2]*, Harish Kalla[3], Arun Roy Vadde[3], Rama Krishna T[1].

[1]Department of Computer Science and Engineering, Jayamukhi Institute of Technological Sciences, Narsampet, Warangal - 506 332, India.
[2]Department of Physics and Computer Science, Vaagdevi Group of Colleges, Hanamkonda, Warangal - 506001, India.
[3]Department of Electrical & Computer Engineering, School of Engineering and Technology, Wollega University, Post Box No: 395, Nekemte, Ethiopia.

Research Article

Received: 02/05/2014
Revised: 23/05/2014
Accepted: 28/05/2014

***For Correspondence**

Department of Computer Science and Engineering, Jayamukhi Institute of Technological Sciences, Narsampet, Warangal - 506 332, India.

Keywords: Algorithm, Dynamic behavior, Frequent patterns, Mining research, Transaction database.

ABSTRACT

A transaction database usually consists of a set of time-stamped transactions. Mining frequent patterns in transaction databases has been studied extensively in data mining research. However, most of existing frequent pattern mining algorithms does not consider the time stamps associated with transactions. We extended the existing frequent pattern mining framework to take into account the time stamp of each transaction and discover patterns whose frequency dramatically changes over time. We define a new type of patterns, called Transitional Patterns, to capture the dynamic behavior of frequent patterns in a transaction database. Transitional patterns include both positive and negative transitional patterns. Their frequencies increase or decrease dramatically at some time points of a transaction database. We introduced the concept of significant milestones for a transitional pattern, which are time points at which the frequency of the pattern changes most significantly. Moreover, we developed an algorithm to mine the set of transitional patterns along with their significant milestones from the transaction database.

INTRODUCTION

In the present study, the authors extending the traditional frequent pattern mining framework to take into account the timestamp of each transaction, i.e., the time when the transaction occurs. Also tried to define a new type of patterns, called transitional patterns, to represent patterns whose frequency dramatically changes over time[1,2]. Transitional patterns include both positive and negative transitional patterns. The frequency of a positive transitional pattern increases dramatically at some time point of a transaction database, while that of a negative transitional pattern decreases dramatically at some point of time[3,4].

The mining of frequent item sets is to find all the item sets from a transaction database that satisfy a user-specified support threshold. It is one of the fundamental and essential operations in many data mining tasks, such as association rule mining, sequential pattern mining, structured pattern mining, correlation mining and association classification[5]. Since it was first introduced by Agrawal et al in 1993, the problem of frequent item set mining has been studied extensively[6]. As a result, a large number of algorithms have been developed in order to efficiently solve the problem, including the most well-known Apriori, FP-growth and Eclat algorithms. The number of frequent patterns generated from a data set are often excessively large, and most of them or useless or simply redundant. Thus, there has been interest in

discovering new types of patterns including maximal frequent item sets, closed frequent item sets, indirect associations and emerging patterns. Mining for maximal or closed frequent item sets greatly reduces the number of generated patterns by generating only the largest frequent item sets with no frequent superset or no superset of higher frequency. A common characteristic of the above learning methods is that they treat the transactions in a database equally and do not consider the time points at which those transactions occurred. Therefore, the existing algorithms can reveal the static behavior of the frequent patterns. The frequency of the patterns which changes over the time period is called dynamic behavior of the patterns. To capture the dynamic behavior of the patterns, new type of patterns were introduced, called Transitional Patterns[7,8].

Preliminaries

In data mining applications, to extract the interesting patterns from databases, mining frequent patterns is one of the fundamental operations. Let I= {i_1, i_2 … i_n} be a set of items. Let D be a set of database transactions where each transaction T is a set of items and ||D|| be the number of transactions in D. Given X= {i_j … i_k} $\subseteq$ I (j ≤k and 1 ≤ j, k ≤ n) is called a pattern. The support of a pattern X in D is the number of transactions in D that contains X. Pattern X will be called frequent if its support is no less than a user specified minimum support threshold[7].

Definitions

Definition 1(Cover): The cover of an item set X in D, denoted by cov(X,D), is the number of transactions in which the item X appears.

Definition 2(Support): An item set X in a transaction database D has a support, denoted by *sup(X, D)*, which is the ratio between *cov(X, D)* to the number of transactions in D i.e.,||D||.

$$\sup(X, D) = \frac{cov(X, D)}{||D||}$$

Definition 3 (Position of a transaction*):* Assuming that the transactions in a transaction database D are ordered by their time-stamps, the position of a transaction T in D, denoted by ρ(T), is the number of transactions whose time-stamp is less than or equal to that of T. Thus 1≤ ρ(T) ≤ ||D||.

Definition 4: The i^{th} transaction of a pattern X in D, denoted by $T^i(X)$, is the i^{th} transaction in *cov(X)* with transactions ordered by their positions, where 1≤ *i* ≤ *cov(X, D)*.

Definition 5 (i^{th} milestone): The i^{th} milestone of a pattern X in D, denoted by $\xi^i(X)$, is defined as

$$\xi^i(X) = \frac{\rho(Ti(X))}{||D||} \times 100\% \quad \text{where } 1 \le i \le cov(X).$$

Definition 6: The support of the pattern X before its i^{th} milestone in D, denoted by $sup^i_-(X)$ is defined as

$$sup^i_-(X) = \frac{i}{\rho(Ti(X))} \qquad \text{where } 1 \le i \le cov(X).$$

Definition 7: The support of the pattern X after its i^{th} milestone in D, denoted by $sup^i_+(X)$ is defined as

$$sup^i_+(X) = \frac{cov(x)-i}{||D||-\rho(Ti(X))} \quad \text{where } 1 \le i \le cov(X).$$

Definition 8 (Transitional Ratio): Transitional ratio is used to measure the difference of a pattern's frequency before and after its i^{th} mile stone. The Transitional ratio of the pattern X at its i^{th}mile stone in D is defined as

$$tran^i(X) = \frac{sup^i_+(X) - sup^i_-(X)}{Max(\, sup^i_+(X), sup^i_-(X))} \qquad \text{where } 1 \le i \le cov(X).$$

It is easy to see that the higher the absolute transitional ratio of a pattern at its i^{th} milestone, the greater the difference between its supports before and after its i^{th}milestone. A nice feature of this definition

is that the value of a transitional ratio is between _1 and 1. As for transitional patterns, I am interested in patterns whose absolute values of transitional ratio are large, which are defined below.

Definition 9 (Transitional Pattern): A pattern X is a transitional pattern (TP) in D if there exists at least one milestone of X, $\xi^k(X) \in T_\xi$, such that:

$$1.\ sup_{-}^{k}(X) \geq t_s and sup_{+}^{k}(X) \geq t_s and$$
$$2.\ |tran^i(X)| \geq t_{t,}$$

Where T_ξ is a range of $\xi^i(X)$ ($1 \leq i \leq cov(X)$), t_s and t_t are called pattern support threshold and transitional pattern threshold, respectively. X is called a Positive Transitional Pattern (PTP) when $tran^k(X)>0$; and X is called a Negative Transitional Pattern (NTP) when $tran^k(X)<0$.

Definition 10 (Significant Frequency-Ascending Milestone) : The significant frequency-ascending milestone of a positive transitional pattern X with respect to a time period T_ξ is defined as a tuple, $(\xi^M(X), tran^M(X))$, where $\xi^M(X) \in T_\xi$ is the M^{th} milestone of X such that

$$1. sup_{-}^{M}(X) \geq t_s;\ \text{and}$$
$$2.\ \forall \xi^i(X) \in T_\xi, tran^M(X) \geq tran^i(X)$$

Definition 11 (Significant frequency-descending milestone): The significant frequency-descending milestone of a negative transitional pattern X with respect to a time period T_ξ is defined as a tuple, $(\xi^N(X), tran^N(X))$, where $\xi^N(X) \in T_\xi$ is the N^{th} milestone of X such that

$$1.\ sup_{+}^{N}(X) \geq t_s\ ;\ \text{and}$$
$$2.\ \forall \xi^i(X) \in T_\xi, tran^N(X) \leq tran^i(X)$$

Designing

This part of the paper presented the design aspects of system and an algorithm, called TP-mine, for mining the set of positive and negative transitional patterns and their significant milestones with respect to a pattern support threshold and a transitional pattern threshold. The algorithm is given as data flow diagram, which is a graphic tool. It is used to describe and analyze the movement of data through a system-manual or automated. They focus on the data flowing in to the system, between process and in and out of data stores. This is a central tool and the basis from which other components are developed[1, 6]. Figure 1 explains the context level diagram, Figure 2 showed collaboration diagram and the Figure 3 explains the total activity chart.

TP-Mine Algorithm

TP-mine. (Mine the set of Transitional Patterns and their significant milestones)

Input: A transaction database (D), an appropriate milestone range that the user is interested (T_ξ), pattern support threshold (t_s), and transitional pattern threshold (t_t).

Output: The set of transitional patterns (SPTP and SNTP) with their significant milestones[1].

Method

1: Extract frequent patterns, PEN; PENCIL; . . . ; Pn, and their supports using a frequent pattern generation algorithm with min_sup =t_s.
2: Scan the transactions from the first transaction to the last transaction before T_ξ to compute the support counts, c_k($1 \geq k \geq n$), of all the n frequent patterns on this part of the database.
3: S_{PTP} =ø, S_{NTP}=ø
4: for all k = 1 to n do
5: MaxTran (P_k) = 0, MinTran (P_k) = 0
6: S_{FAM} (P_k) =ø,S_{FDM} (P_k) =ø
7: end for
8: for all transactions T_i whose positions satisfying T_ξ do

9: for k = 1 to n do
10: if $T_i \supseteq P_k$ then
11: $c_k = c_k + 1$
12: if $sup_{-}^{c_k}(P_k) \geq t_s$ and $sup_{+}^{c_k}(P_k) \geq t_s$ then
13: if $tran^{c_k}(P_k) \geq t_t$ then
14: if $P_k \notin S_{PTP}$ then
15: Add P_k to S_{PTP}
16: end if
17: if $tran^{c_k}(P_k)$> MaxTran (P_k) then
18: S_{FAM} (P_k) = $\{\xi^{c_k}(P_k), tran^{c_k}(P_k)\}$
19: MaxTran (P_k) =$tran^{c_k}(P_k)$
20: else if $tran^{c_k}(P_k)$ = MaxTran (P_k) then
21: Add $\{\xi^{c_k}(P_k), tran^{c_k}(P_k)\}$ to S_{FAM} (P_k)
22: end if
23: else if $tran^{c_k}(P_k) \leq -t_t$ then
24: if $P_k \notin S_{NTP}$ then
25: Add $P_k$toS_{NTP}
26: end if
27: if $tran^{c_k}(P_k)$<MinTran (P_k) then
28: $S_{FDM}(P_k)$ = $\{\xi^{c_k}(P_k), tran^{c_k}(P_k)\}$
29: MinTran (P_k) =$tran^{c_k}(P_k)$
30: else if $tran^{c_k}(P_k)$ = MinTran (P_k) then
31: Add $\{\xi^{c_k}(P_k), tran^{c_k}(P_k)\}$ to S_{FDM} (P_k)
32: end if
33: end if
34: end if
35: end if
36: end for
37: end for
38: return S_{PTP} and S_{FAM} (P_k) for each $P_k \in S_{PTP}$
39: return S_{NTP} and S_{FDM} (P_k) for each $P_k \in S_{NTP}$

There are two major phases in this algorithm. During the first phase (Step 1), all frequent item sets along with their supports are initially derived using a standard frequent pattern generation algorithm, such as Apriorior FP-growth, with t_s as the minimum support threshold. In the second phase (starting from Step 2 to theend), the algorithm finds all the transitional patterns and their significant milestones based on the set of frequent item sets. As mentioned before, a pattern that is frequent before and after one of its milestones in D with respect to support threshold t_s must be frequent on D with respect to the same threshold. Thus, it is safe for us to first mine the frequent item sets on the entire database using the threshold t_s, and then find the transitional patterns based on the set of frequent item sets.

In Step 2, the support counts of all the frequent patterns on the set from the first transaction to the transaction right before the time period T_ξ are collected. They are used later in computing$sup_{-}^{c_k}(P_k)$, where P_i is a frequent pattern. Step 3initializes the set of positive transitional patterns (S_{PTP}) and the set of negative transitional patterns (S_{NTP}) to empty.

Steps 4-7 initialize the set of significant frequency-ascending milestones for each frequent pattern P_k, $S_{FAM}(P_k)$, and the set of significant frequency-descending milestones for each frequent pattern P_k, S_{FDM} (P_k), to empty. It also initializes the maximal and minimal transitional ratios of P_k, denoted by MaxTran (P_k) and MinTran (P_k), to zero. After the initializations, the algorithm continues to scanthe database D to find the milestones of P_k within the range T_ξ. At each valid milestone $\xi^{ck}(P_k)$ during the scan, it calculates the support of P_k before $\xi^{ck}(P_k)$, i.e., $sup_{-}^{c_k}(P_k)$ and the support of P_k after $\xi^{ck}(P_k)$, i.e., $sup_{+}^{c_k}(P_k)$, If both of them are greater than t_s, the algorithm then checks the transitional ratio of P_k. If the ratio is greater than t_t, then P_k is a positive transitional pattern and is added into the set S_{PTP}. Then, the algorithm checks whether the transitional ratio of P_k is greater than the current maximal transitional ratio of P_k. If yes, the set of significant frequency-ascending milestones of P_k, i.e., $S_{FAM}(P_k)$, is set to contain $\{\xi^{c_k}(P_k), tran^{c_k}(P_k)\}$ as its single element. If not but it is equal to the current maximal transitional ratio of P_k,$\xi^{c_k}(P_k), tran^{c_k}(P_k)$ is added into $S_{FAM}(P_k)$.Similarly, Steps 23-32 are for finding the set of negative transitional patterns and their significant frequency-descending milestones.

e-ISSN:2319-9873
p-ISSN:2347-2324

RESULTS

To demonstrate the efficiency of the TP-Mine algorithm we have done experiments on synthetic data obtained shown in Table 1. The results were shown in Table 2 and explained through Figures 4, 5 and 6. Our experimental results showed that the proposed algorithm is highly efficient and scalable.

Table 1: Example Database

TID	List of Item IDs	Time stamp
001	Pen, Pencil, Eraser, Gum	Nov, 2005
002	Pen, Pencil	Dec, 2005
003	Pen, Pencil, Eraser, Chalk	Jan, 2006
004	Pen, Pencil, Gum	Feb, 2006
005	Pen, Pencil, Books	Mar, 2006
006	Pen,Pencil, Books, Gum, Ink	Apr, 2006
007	Pen,Pencil, Eraser, Books, Ink	May, 2006
008	Pen, Books, Ink	Jun, 2006
009	Books, Gum, Ink	Jul, 2006
010	Pen,Pencil, Eraser,books,Gum,ink	Aug, 2006
011	Pen, Eraser,Books,Ink	Sep, 2006
012	Pen, Eraser, Gum	Oct, 2006
013	Pen,Pencil, Eraser, Ink, Papers	Nov, 2006
014	Pen, Eraser, Books, Gum	Dec, 2006
015	Pen, Eraser, Books	Jan, 2007
016	Pen,Pencil, Eraser, Gum	Feb, 2007

Table 2: Some of the milestones and transitional patterns

i	ξ^i(PEN)$tran^i$ (PEN)		ξ^i(PEN,PENCIL)$tran^i$ (PEN,,PENCIL)		ξ^i(PENERASER)$tran^i$ (PEN,ERASER)		ξ^i(BOOKS,INK) $tran^i$ (BOOKS,INK)	
1							37.50%	+66.67
2							43.75%	+35.71
3					43.75%	+44.90	50.00%	0
4	25.00%	-8.33	25.00%	-50	62.50%	+60	56.25%	-35.71
5	31.25%	-9.09	31.25%	-54.55	68.75%	+54.55	62.50%	-66.67
6	37.50%	10	37.50%	-60	75.00%	+50	68.75%	-100
7	43.75%	-11.11	43.75%	-66.67				
8	50.00%	-12.5	62.50%	-44.90				
9	62.50%	+11.11						
10	68.75%	+10						
11	75.00%	+9.09						

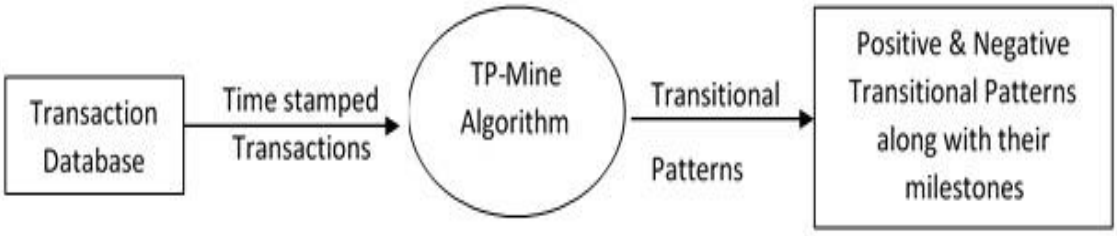

Figure 1: Context level diagram

e-ISSN:2319-9873
p-ISSN:2347-2324

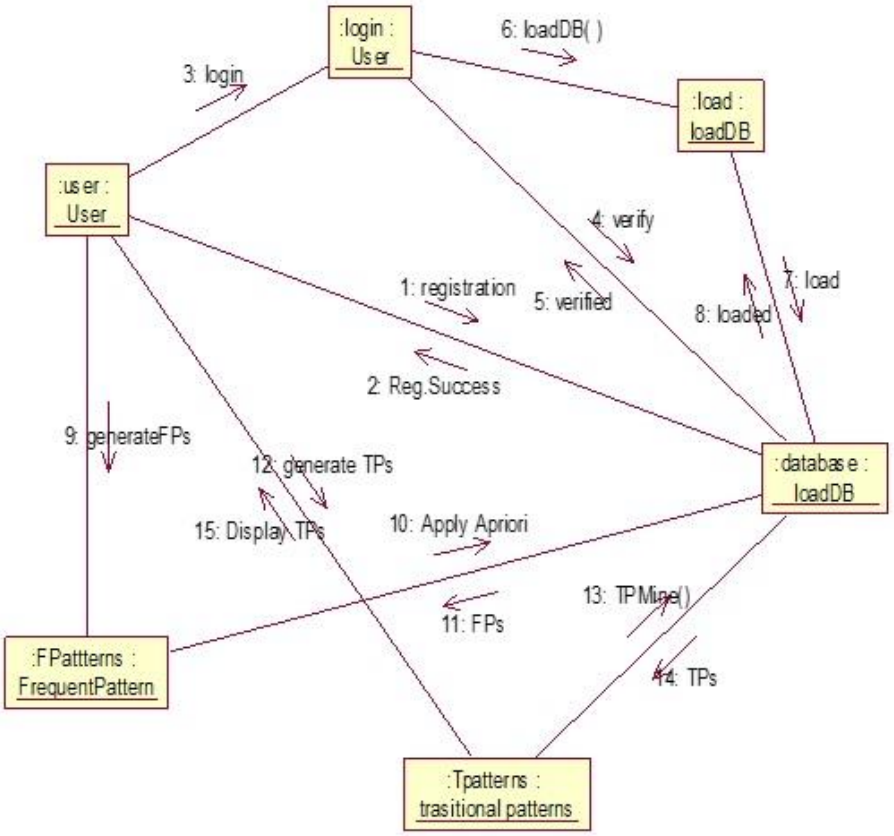

Figure 2: Collaboration diagram

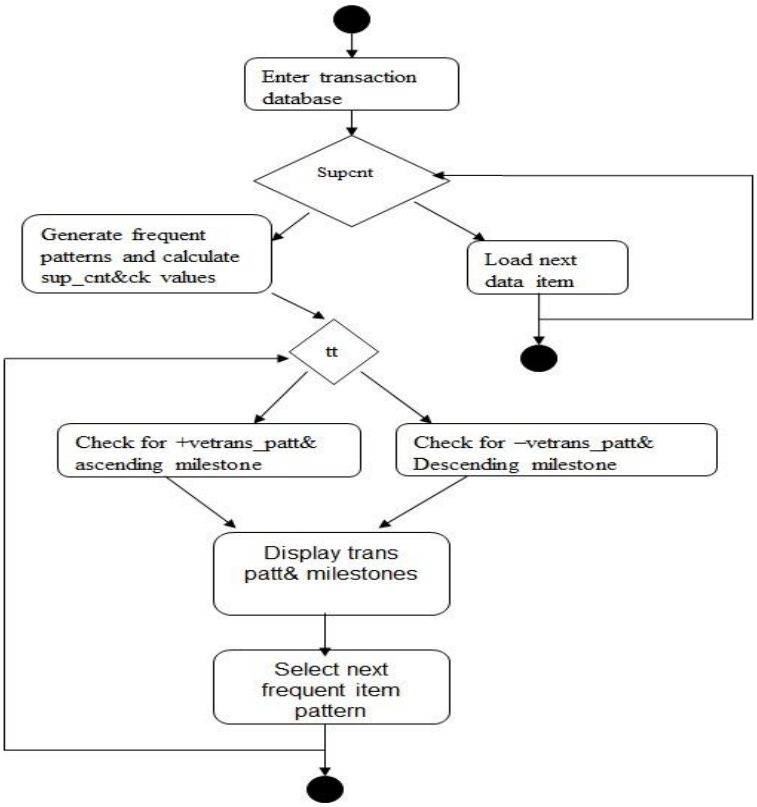

Figure 3: Activity diagram

e-ISSN:2319-9873
p-ISSN:2347-2324

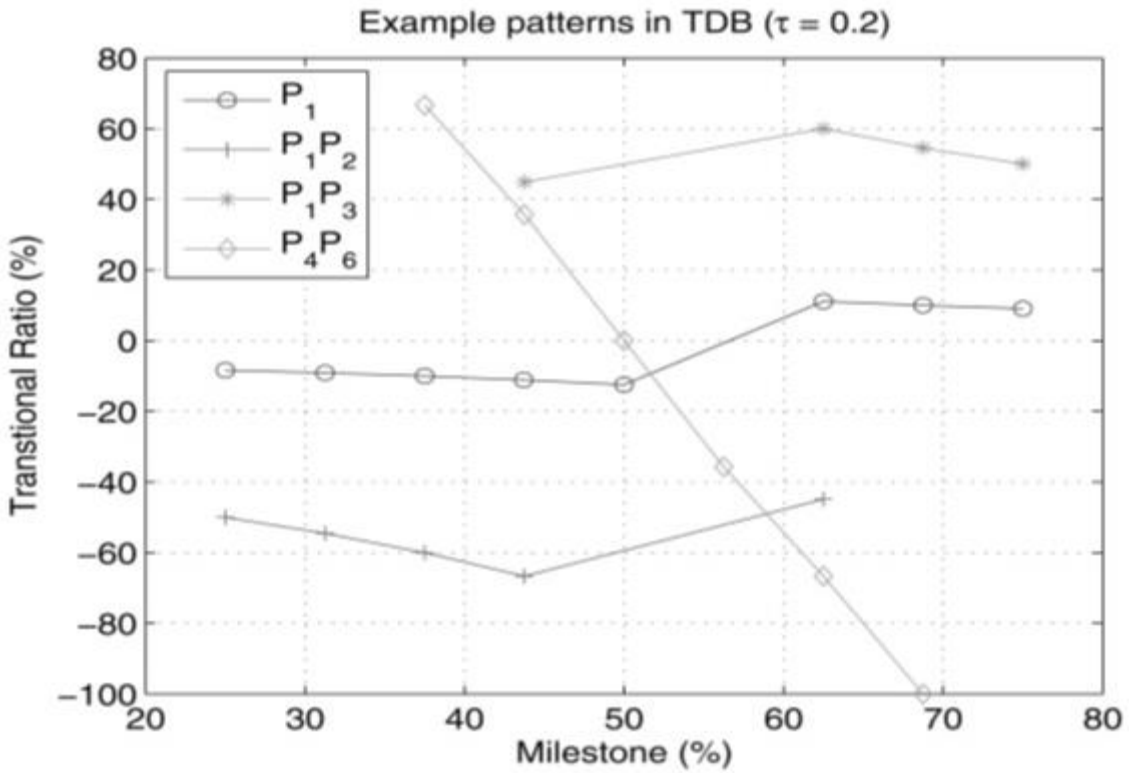

Figure 4: Transitional ratios in TDB

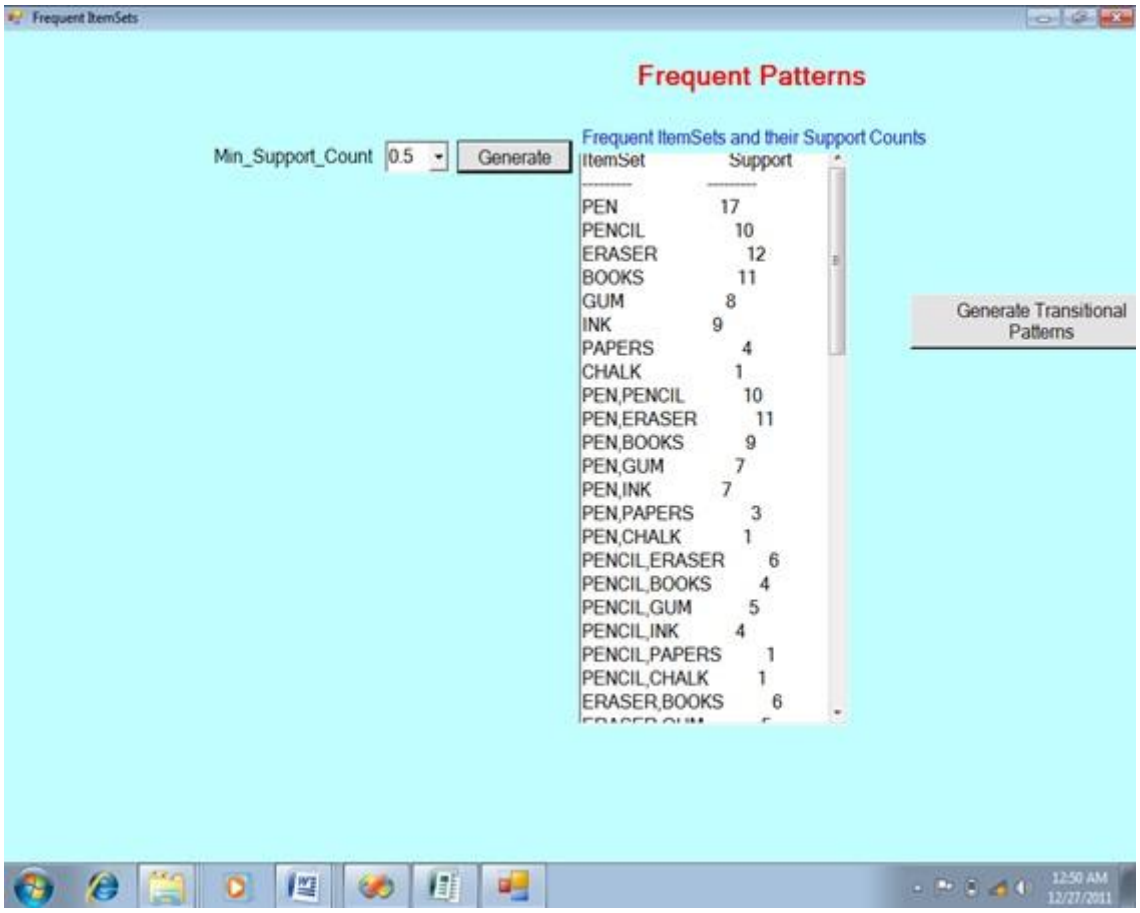

Figure 5: Frequent patterns with their supports

e-ISSN:2319-9873
p-ISSN:2347-2324

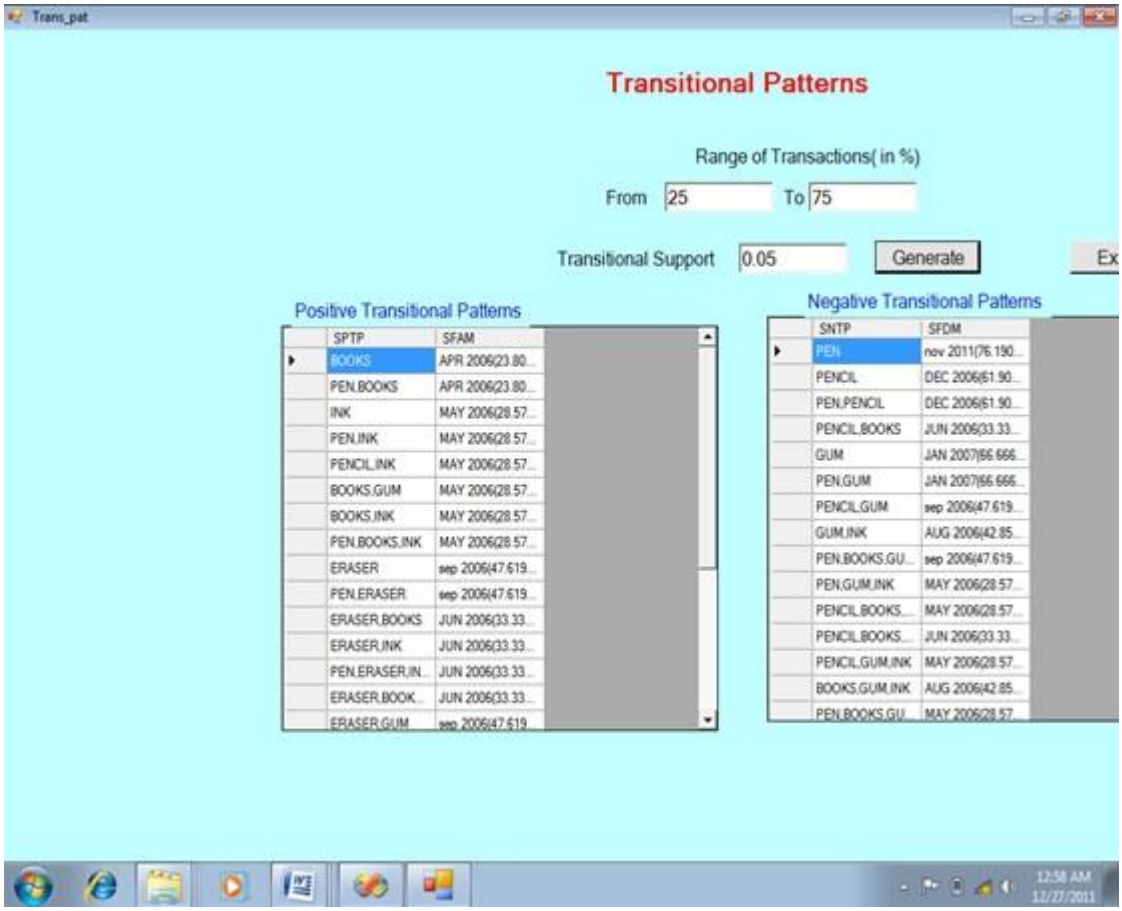

Figure 6: Transitional patterns and their milestones

CONCLUSION

A limitation of existing frequent item set mining framework is that it does not consider the time stamps associated with the transactions in the database. As a result, dynamic behavior of frequent item sets cannot be discovered. In this study, we introduced a novel type of patterns, positive and negative transitional patterns, to represent frequent patterns whose frequency of occurrences changes significantly at some points of time in a transaction database. And also defined the concepts of significant frequency-ascending mile stones and significant frequency-descending milestones to capture the time points at which the frequency of patterns changes most significantly. To discover transitional patterns, proposed the TP-mine algorithm to mine the set of positive and negative transitional patterns with respect to a pattern support threshold and a transitional pattern threshold. This algorithm takes one database scan after mining frequent patterns to find the transitional patterns and their significant milestones. Experimental results showed that the proposed algorithm is highly scalable. The experimental study demonstrated the use fullness of transitional patterns in two real-world domains and showed that what is revealed by the transitional patterns and their significant milestones would not be found by the standard frequent pattern mining framework.

REFERENCES

1. Wan Q, An A. Discovering transitional patterns and their significant milestones in transaction databases. IEEE Trans Know Data Eng. 2009;21(12):1692-07.
2. Srikant R, Agarwal R. Mining generalized association rules. Fut Gen Comp Sys. 1997;13(2):161-80.
3. Sujatha D, Shyamala P. PTP-Mine: range based mining of transitional patterns in transaction databases. Glob J Com Sci Tech. 2012;12(2):21-8.
4. Toivonen H. Sampling large databases for association rules. Proceedings of Int Conference on Very Large Databases. VLDB, 1996: 134-45.

e-ISSN:2319-9873
p-ISSN:2347-2324

5. Agarwal RC, Agarwal CC, Prasad VVV. Depth first generations of long patterns. Proceedings of Sixth ACM SIGKDD Int Conference. KDD, 2000:108-18.
6. Agarwal R, Imielinski T, Swami A. Mining association rules between sets of items and large databases. Proceedings of ACM SIGMOID Int Conference. Mang Data, 1993:207-16.
7. Kumar KB, Bhaskar A. ETP-Mine: An Efficient Method for Mining Transitional Patterns. Int J Database Management Sys. 2010;2(3):1-11.
8. Wan Q, An A. Transitional patterns and their significant milestones. Data Mining, ICDM 2007. Seventh IEEE International Conference on. IEEE, 2007.