PTP-Mine Range Based Mining of Transitional Patterns in Transaction databases

P.Shyamala [1]

[1] Auroras Technological and Research Institute, JNTU

Received: 15 April 2012 Accepted: 30 April 2012 Published: 15 May 2012

Abstract

Transaction database is a collection of transactions along with the related time stamps. These transactions are defined using some prototypes. They are called as the Transitional patterns that denote the vibrant nature of the frequent patterns in the database. The considerable high points for the transaction database are the timestamps also called as time durations. They are the points that have the alteration in the recurrence of the prototypes. There is majorly couple of stages in existing TP-Mine algorithm. The initial stage is to find out the frequent patterns and the second stage is to discover Transitional patterns or the styles and the significant milestones. These patterns consist of the two kinds of the styles likely the positive one and the negative one. In the previous time cases the effort that was made on the research was to build up the algorithms by planning the total series of the transitional patterns. In our paper we consider that the alterations made in the consequent period regarding the total concept of database is not noteworthy. So for this reason we have put forward an entirely latest transitional patterns methodology called periodical transitional pattern mining. The experimental outputs are appealing and apparent those were produced by this periodical transitional pattern mining and has high importance that when evaluated utilizing the present patterns.

Index terms— Data mining, Association Rules, Frequent Pattern, Transitional Pattern, significant milestones

1 INTRODUCTION

n the recent time the expansively considered field in the data mining is the frequent pattern mining. The advantage of this technique is that it can be widely utilized in several methodologies such as Association rule mining [2] [3] [12], sequential pattern mining [4], structured pattern mining [7], correlation mining [5], and associative classification [9]. Even to determine these frequent patterns that are present in the transaction databases, plenty of procedures were put forward specifically Apriori [3], FP-growth [8], and Eclat [15]. The drawback of these procedures is that they can produce huge amount of the patterns on the condition of low support threshold value. As this drawback being one of the reasons, maximum procedures are not utilized for the data mining job. So, as a case of evading the above mentioned ineffectual and outmoded patterns, the latest patterns such as frequent itemsets One of the common characteristic of these procedures is that they don't determine the time stamps that are related to the actions in the database. This leads to the loss of exposure of the vibrant nature of the patterns. For an illustration let us consider an electronic showroom's database. The amount of retails of refrigerator during the peak summer season is very much more when compared to that of the retails in the winter. But when we determine all the transactions in whole on an average the retails are frequent, while in actual they are frequent only in the peak season. So in order to find out such a vibrant characteristics, latest patterns have been established like Transitional patterns [13] [14]. There are two kinds in them such as the positive patterns and the negative patterns. Positive transitional patterns have the characteristic of incrementing the recurrence of the pattern's time stamp where as the negative patterns decrement the recurrence of the pattern's time stamp.

The periodic positions where the recurrence of the transitional patterns alters often are termed as the significant milestones of the transitional pattern. So to determine these transitional patterns and the respective significant milestones TP-Mine algorithm [13][14] [17] has been projected.

The two most important and chief stages of the TP-Mine [14] algorithm are as follows:

1. In the Initial stage each and every frequent pattern is produced from the present transaction database. 2. The consumer have some of the preferences, In the consequent stage based on those preferences the frequent patterns are utilized for the production of the transitional patterns and significant milestones, By considering all the above mentioned problems we in our article are introducing an adaption to the present TP-Mine [14] algorithm with a view of decrementing the patterns that were derived in the initial stage. As a result the count of the calculations required for the production of transitional patterns gets decremented. By this there is a feasibility to develop the importance of transitional patterns by keeping in account the periodicity of the mentioned series as the milestones.

The structure of the remaining paper is designed as below: The second part of the paper illustrates the preface and defining of the terms that were being utilized in the TP-Mine [14] algorithm. The later part that is the 3 rd one illustrates the PTP-Mine Algorithm and the 4 th one represents the tentative outputs of the approach. And as a final part, the 5 th one terminates the approach.

2 II.

3 PREFACE AND DEFINITIONS

The following are the definitions taken from [14]. Apart from these, new definition defined in definition 3.2. The major elementary function of data mining approaches in digging out the valuable patterns that are present in the databases is excavation of the frequent patterns. Assuming a group of items is I= {i1, i2 ? in} with a group of database transactions D that consists of individual transaction T. T is termed as the group of items and the count of the transactions in D is derived by ||D||.

The specified equation{ ... } (1 ,) j k X i i I j k and j k n = ? ? ?

? is termed as a pattern. The pattern's X in an action D has a support value, which is termed as the count of the transactions in the D that consists of the respective pattern X. This X is defined as recurrent or simply frequent if and only if the corresponding support value is greater to the consumer mentioned the least support threshold value.

4 Definition 2.1:

The representation of the cover for the individual itemset X in D. This is represented using cov(X, D). This is the count of the transactions that consists of the item X.

5 Definition 2.2:

The individual itemset X for the transaction database D consists of a support value. The representation of it is sup(X, D). This is termed as the proportionate value involving the cov(X, D) and the count of transactions in D which is represented by||D||. ( ) i T X is defined

as the th i transaction for the cover of X i.e., cov( ) X that has the transactions arranged based on their positions.

Here 1 cov(,) i X D ? ? . Definition 2.5: For a pattern X in D, the th i milestone which is represented by ( ) i X ? , is termed as(( )) ( ) 100% || || i i T X X D ? ? = ×

, where 1 cov( ) i X ? ? Definition 2.6: For a pattern X in D, the support value preceding its th i milestone in D which is represented by sup _( ) i X is termed as follows:sup ( ) (( )) i i i X T X ? ? = , where 1 cov( ) i X ? ? Definition 2.7 For a pattern X in D, the support value following its th i milestone in D which is represented by sup ( ) i X + is termed as follows: cov( ) sup ( ) || || (( )) i i X i X D T X ? + ? = ?

, where 1 cov( ) i X ? ? Definition 2.8: For a pattern X in D at the th i milestone the Transitional ratio is termed as follows: sup ( ) sup ( ) ( ) (sup ( ),sup ( )) i i i i i X X tran X Max X X + ? + ? ? = , where 1 cov( ) i X ? ? Definition 2.9: For a pattern X in D is considered as transitional pattern (TP) in D if on minimum, single milestone of , ( ) k X X T ? ? ? must be present, so that: sup ( ) s up ( ) | ( ) | k k s s i t

6 X t and X t and tran X t

? + ? ? ? here T ? is the series of ? i (X) (1? i ? cov(X)) t

7 TP-MINE[14] ALGORITHM

The group of positive and negative transitional patterns including their significant milestones is produced using TP-Mine algorithm [14].

Input There is a clear understanding from the outcome in the tables that the count for the transitional patterns that were produced in the transaction data group of magnitude 16 that come in the lower region of support and threshold value is just 26. And it is also found out that those transactions that were produced from the standard data groups that has more than ten transactions directs towards the large value complication and maximum patterns are also not significant patterns. Due to this reason the patterns must be produced in the periodical series which will help us to determine the large significant patterns. In order to reduce the complication an algorithm called Bide [16] algorithm is preferred for determining the frequent patterns and decrementing the count of the pattern. A transaction database (D), an appropriate milestone range that the user is interested (T ?), pattern support threshold (t s), and transitional pattern threshold (t t). transition range(tdr i)

8 Output:

The group of transitional patterns (positive(S PTP) and negative(S NTP)) including their significant milestones.

Algorithm:

1. Determine frequent patterns a. In order to preserve the individual patterns, snip out the patterns that are as subset for the superset frequent patterns and has the equal support values(definition 3.1). As a result the count of the patterns gets reduced. Many tests have been organized on artificial data groups having 8 values and with transactions greater than 200. JAVA 1.6_20th build was utilized for execution of the PTP-Mine including TP-Mine [14] for examination. A computer unit prepared with core2duo processor, 2GB RAM and Windows XP loaded was utilized for inquiring the algorithms.

9 a) Dataset Characteristics

The dataset prepared is a very opaque dataset, which assists in excavating enormous quantity of recurring clogged series with a profitably high threshold somewhere close to 90%. It also has a distinct element of being enclosed with more than 200 transactional series and 8 divergent objects. Reviewing of serviceable legacy's consistency has been made use of by this dataset.

Comparative study: In assessment with TP-Mine has made its mark as a most preferable, superior and sealed example of transitional pattern mining Table **??** : contrast account of patterns derived below diverse supports by PTP-Mine and TP-Mine [14] Fig1 : Count of the patterns derived by TP-Mine [14] and PTP-Mine The fig- **??** represents the count of the patterns that were reduced on contrasting the PTP-Mine with the TP-Mine [14] and equally we can distinguish that the functioning of PTP-Mine in determining the +ve and -ve transitional patterns is sound. This serves as the proof for defining that the PTP-Mine subsequently works better on contrasting with the TP-Mine [14]. The PTP-Mine decreases the count of patterns in order to decrement the calculation complications and generates a large amount of transitional patterns to progress the construction of constraint dependent resolutions impact (fig 2 and fig 3). [1] [2]

[1] PTP-Mine: Range Based Mining of Transitional Patterns in Transaction databases Global Journal of Computer Science and Technology Volume XII Issue II Version I

[2] Global Journal of Computer Science and Technology Volume XII Issue II Version I

RESEARCH | DIVERSITY | ETHICS

2324

Figure 1: Definition 2 . 3 :Definition 2 . 4 :

3132

Figure 2: Definition 3 . 1 : 3 . 2 :TD

2

Figure 3: 2 .

3

Figure 4: 3 .

2

Figure 5: Fig 2 :

3

Figure 6: Fig 3 :

2

t
),
and

transitional pattern threshold(t t).
D is assumed as a group of transactions
scheduled in Table 1, T ? = {25%, 75%}, s t =0.05, t t =0.5.
As mentioned above the algorithm consists of a couple
of important stages. Coming to the initial stage the
production of every recurrent itemsets including the
support values is done utilizing Apriori [3] or FP-growth
[8] considering s t as minimum support threshold. This
stage involves the generation of n count of frequent
patterns by the algorithm for which support ? s t in the
transaction database D. The consequential frequent
patterns (n=87) have been represented in Table 2.
Subsequently in the second stage the support
count (k c) for each and every frequent patterns in the
group of transactions starting from the initial action to
latest one with the time stamp T ? is derived by the
algorithm. Later the milestones of

k (1)
P ?
?
in
the
k
n

series T ? are also derived. It checks whether the

[Note: : A transaction database (D), milestone range (T ?), pattern support threshold (s]

Figure 7: Table 2 :

3

Pattern	Incrementing Milestone	Transitional Ratio
P3	Aug2006(62.5%)	60.0%
P4	Mar2006(31.25%)	72.50%
P6	Apr2006(37.5%)	72.0%
P1, P3	Aug2006(62.5%)	60.0%
P1, P4	Mar2006(31.25%)	68.57%
P1, P6	Apr2006(37.5%)	66.66%
P3, P4	May2006(43.75%)	67.85%
P3, P5	Aug2006(62.5%)	60%
P3, P6	May2006(43.75%)	57.14%
P4, P6	Apr2006(37.5%)	66.66%
P1,P3, P4	May2006(43.75%)	67.85%
P1,P3, P5	Aug2006(62.5%)	60%
P1,P3, P6	May2006(43.75%)	57.14%
P1,P4, P6	Apr2006(37.5%)	58.33%

Table 4 : A group of Negative transitional patterns and the corresponding decrementing milestones.

Pattern	Decrementing Milestone	Transitional Ratio
P2	May2006(43.75%)	-66.66%
P6	Sep2006(68.75%)	-63.33%
P1, P2	May2006(43.75%)	-66.66%
P1, P6	Sep2006(68.75%)	-56.0%
P2, P4	May2006(43.75%)	-74.07%
P2, P5	apr2006(37.5%)	-60.0%
P4, P6	aug2006(62.5%)	-66.66%
P1,P2, P4	may2006(43.75%)	-74.07%
P1,P2, P5	apr2006(37.5%)	-60.0%
P1,P4, P6	aug2006(62.5%)	-58.33%
P2,P4, P6	may2006(43.75%)	-61.11%
P1,P2,P4, P6	may2006(43.75%)	-61.11%

Figure 8: Table 3 :

1

			January 2012
13.	if	and	then
14.	if	then	
15.	if P k ? S PTP then		
16.	Add P k to S PTP		
17.	end if		
18.	if		then
19.			
20.			
21.	else if		then
22.	Add		
23.	end if		
24.	else if		then
25.	if P k ? S NTP then		
26.	Add P k to S NTP		
27.	end if		
28.	If		then
29.			
30.			
31.	else if		then
32.		Add	

Figure 9: Table 1 ,

5

Figure 10: Table 5 :

6

January 2012
26

Figure 11: Table 6 :

7

Patterns and the corresponding Decrementing Milestones.		
Pattern	Decrementing Milestone	Transitional ratio
P4,P2,P1,P6	May2006(43.75%)	-61.11%
P3,P2,P1,P6	Aug2006(62.5%)	-16.66%
P4,P3,P1,P6	Aug2006(62.5%)	-16.66%
P3,P2,P1,P5	Aug2006(62.5%)	-16.66%
P4,P1,P6	Aug2006(62.5%)	-58.33%
P4,P6,P5	July2006(56.25%)	-35.71%
P3,P2,P1	May2006(43.75%)	-22.22%
P4,P2,P1	May2006(43.75%)	-74.07%
P2,P1,P5	Apr2006(37.5%)	-60.0%
P2,P1,P6	Aug2006(62.5%)	-44.44%
P4,P1,P5	Aug2006(62.5%)	-16.66%
P3,P1,P6	Sep2006(68.75%)	-26.66%
P4,P5	Aug2006(62.5%)	-44.44%
P1,P5	Apr2006(37.5%)	-19.99%
P1,P6	Sep2006(68.75%)	-56.0%
P4,P1	Sep2006(68.75%)	-26.66%
P2,P1	May2006(43.75%)	-66.66%
P4,P6	Aug2006(62.5%)	-66.66%
P4	Sep2006(68.75%)	-37.14%
P1	June2006(50.0%)	-12.50%
P4,P2,P1,P6,P5	Apr2006(37.5%)	-40.00%
P4,P3,P2,P1,P6	May2006(43.75%)	22.22%
P6	Sep2006(68.75%)	-63.33%

Figure 12: Table 7 :

9

Support	TP-Mine[14]	PTP-Mine
20%	80	142
30%	63	127
40%	57	83
50%	41	69
60%	23	45
70%	7	19
80%	3	14

Figure 13: Table 9 :

10

20%	234	434
30%	217	449
40%	111	493
50%	105	507
60%	223	531
70%	139	557
80%	143	562

Figure 14: Table 10 :

V.

.1 CONCLUSION

We observe that Compare to TPMine [14], PTP-Mine produces less no. of frequent patterns. TP-Mine [14] algorithm produces less number of positive and negative transitional patterns because of transitional pattern threshold that is assumed to be 0.5, in which case only the patterns which change dramatically are displayed. In PTP-Mine the transition pattern threshold is considered as 0.1 so that, even the positive and negative patterns which have less change are also exhibited. The tentative outcome delivered by our methodological algorithm that was put forward is decidedly resourceful and proficient.

[Issue (2011)] , *Issue* July -Aug 2011. 3.

[Brin et al. ()] 'Beyond Market Baskets: Generalizing Association Rules to Correlations'. S Brin , R Motwani , C Silverstein . *Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '97)*, (ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '97)) 1997. p. .

[Wang and Han ()] 'BIDE: Efficient Mining of Frequent Closed Sequences'. Jianyong Wang , Jiawei Han . *Proceedings of the 20th International Conference on Data Engineering (ICDE '04)*, (the 20th International Conference on Data Engineering (ICDE '04)Washington, DC, USA) 2004. IEEE Computer Society. p. 79.

[Agarwal et al. ()] 'Depth First Generation of Long Patterns'. R C Agarwal , C C Aggarwal , V V V Prasad . *Proc. Sixth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '00)*, (Sixth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '00)) 2000. p. .

[Pasquier et al. ()] 'Discovering Frequent Closed Itemsets for Association Rules'. N Pasquier , Y Bastide , R Taouil , L J Lakhal ; R , Jr Bayardo . *Proc. 1998 ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '98*, (1998 ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '98) 1999. 1998. p. . (Proc. Seventh Int'l Conf. Database Theory (ICDT '99))

[Wan and An (2009)] 'Discovering Transitional Patterns and Their Significant Milestones in Transaction Databases'. Q Wan , A An . *IEEE Trans. on Knowledge and Data Engg* Dec 2009. 21 (12) p. .

[Agrawal and Srikant ()] 'Fast Algorithms for Mining Association Rules'. R Agrawal , R Srikant . *Proc. 20th Int'l Conf. Very Large Data Bases*, (20th Int'l Conf. Very Large Data Bases) 1994. p. .

[Zaki and Gouda ()] 'Fast Vertical Mining Using Diffsets'. M J Zaki , K Gouda . *Proc. Ninth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '03)*, (Ninth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '03)) 2003. p. .

[Han et al. ()] 'From Sequential Pattern Mining to Structured Pattern Mining: A Pattern-Growth Approach'. J W Han , J Pei , X F Yan . *J. Computer Science and Technology* 2004. 19 (3) p. .

[Liu et al. ()] 'Integrating Classification and Association Rule Mining'. B Liu , W Hsu , Y.-M Ma . *Proc. Fourth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '98)*, (Fourth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '98)) 1998. p. .

[Burdick et al. (2005)] 'Mafia: A Maximal Frequent Itemset Algorithm'. D Burdick , M Calimlim , J Flannick , J Gehrke , T Yiu . *IEEE Trans.Knowledge and Data Eng* Nov. 2005. 17 (11) p. .

[Agrawal et al. ()] 'Mining Association Rules between Sets of Items in Large Databases'. R Agrawal , T Imielinski , A Swami . *Proc. 1993 ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '93)*, (1993 ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '93)) 1993. p. .

[Sidduri et al.] 'Mining Dynamic Behavior of Frequent Patterns using TP-Mine Algorithm'. Rekharani Sidduri , T P Shekhar , Ganesh Kumar Nune . *IJCTT* 1.

[Han et al. ()] 'Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach'. J W Han , J Pei , Y W Yin , R Y Mao . *Data Mining and Knowledge Discovery* 2004. 8 (1) p. .

[Srikant ()] 'Mining Generalized Association Rules'. R Srikant , R . *Future Generation Computer Systems* 1997. 13 p. .

[Agrawal and Srikant ()] 'Mining Sequential Patterns'. R Agrawal , R Srikant . *Proc.11th Int'l Conf. Data Eng*, (.11th Int'l Conf. Data Eng) 1995. p. .

[Ptp-Mine (2012)] 'Range Based Mining of Transitional Patterns in Transaction databases Global Journal of Computer Science and Technology Volume XII Issue II Version I 28'. Ptp-Mine . *Global Journals Inc* January 2012. 2012. US.

[Wan and An ()] 'Transitional Patterns and Their Significant Milestones'. Q Wan , A An . *Proc. Seventh IEEE Int'l Conf. Data Mining*, (Seventh IEEE Int'l Conf. Data Mining) 2007.