Journal of Theoretical and Applied Information Technology
30th April 2016. Vol.86. No.3

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

# REGULAR PATTERN MINING ON DYNAMIC DATABASES USING VERTICAL FORMATE ON GIVEN USER REGULARITY THRESHOLD

[1]M.N.S.Gangadhar, [2]Dr M.Sreedevi
[1]M.tech., Department of Computer Science and Engineering, , KL UNIVERSITY,Guntur
[2]Professor., Department of Computer Science and Engineering, , KL UNIVERSITY,Guntur
E-mail: [1]gangamns1001@gmail.com, [2]msreedevi_27@kluniversity.in

**ABSTRACT**

Ascertaining stimulating patterns in effervescent databases is often an exigent task in data mining due to number of transactions arriving dynamically with rapid rates, which is ridiculously expensive in mutually time and space. Mining frequent patterns is a outmoded mining technique in data mining applications, which is based on support count of item set. Support count of item set is not only sufficient, regularity is also considered and required to mine interesting patterns in data mining. As of late, standard itemset mining picked up parcel of consideration in information mining research as a result of its event conduct. In this paper we propose a strategy called MRPDyD (mining standard examples on element information bases) to mine general itemsets in element databases utilizing vertical information position. Our MRPDyD system creates complete arrangement of standard examples in element databases for a client given consistency edge. Our exploratory results demonstrate that this strategy is effective in memory use and execution time.
**Keywords:** *Regular pattern; Dynamic databases; vertical data format; Sliding window technique*

## 1. INTRODUCTION

Recurrent pattern mining is a customary, crucial and vital territory in information mining [1], [2], [3], [4]. Successive arrangement mining has abundant of uses including inquiry access designs, disclosure of D&R NA groupings, client errands arrangements, website page arrangements and securities exchange and so on., Mining interesting patterns on dynamic databases plays a vital role in several applications and information from dynamic databases is crucial due to its regularity, unbounded and its high speed. Conversely the prodigiousness of a pattern may not generally rely on upon its recurrence livelihood. The criticalness of a pattern might likewise rely on their event attributes, for example, happening at customary interims in value-based databases. There is no calculation have been proposed for mining general examples in element databases.

Thus in this article we intend a new routine called MRPDyD routine to mine systematic patterns on vibrant databases using erect data organisation. Fixed enumerates are mined based on consumer given predictability verge. The chief idea of our new routine is to develop a modest, authoritative method to mine regular patterns which occur at regular intervals in dynamic databases using erect records format. The investigational outcomes show that the efficiency of MRPDyD routine for discovering fixed patterns on vibrant databases

The continuing of this article is systematized as follows. Section 2 refer to related work, Section 3 refer to problem definition, section 4 refer to the process of mining regular patterns, section 5 refer to investigational results and finally section 6 will conclude the article.

## 2. RELATED WORK

Insinuation rule mining is a standout amongst the furthermost critical systems in information mining and mining visit thing sets was initially presented by Agarwal and Srikanth in 1993[5], [6].This separates every successive example, connections, relationship among sets of things in value-based databases. The principle downside with this established calculation is that it needs rehashed sweeps to produce hopeful sets. Han et aI.,[7] presented a tree based information structure called

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

FP tree to produce regular examples without creating hopeful sets. This calculation needs just dual catalogue outputs to mine incessant examples. Tanbeer et aI., [8] presented another issue of finding standard examples that take after transient normality in their event conduct. With the assistance of consistency measure at which design happens in a catalogue at a client given most extreme interim is called normal example. They proposed a tree based information structure called RP Tree, finds general examples in a value-based catalogues which needs just two catalogue checks. Vijay Kumar et al., proposed VDSRP system to create complete arrangement of general examples over an information stream at a client given consistency edge utilizing vertical information position [9][10][11].

## 3. PROBLEM DEFINITION

Give I = {i1, i2, … , im} be an arrangement of things. A set X = {i1, i2, ..., xn} Є I, where 1≤ i ≤ n is called a pattern or an recorded and dynamic database DDB has T = (tid, *X*) is a tuple where *tid* is a unique operation identifier and *X* is a pattern. The item set with k number of items is called k-sized item set over the DDB.

TRANSACTION SLIDING WINDOW $T_{sw}$ of DYNAMIC DATABASE:

Transaction sliding window $T_{sw}$ contains affixed number of transactions from dynamic database. Glide of the window familiarize and expires the glide size i.e, 1<=slide-size>=|W|, the transaction addicted to and from the operation window. If X transpires in $t_j$, the transaction-id of X is represented as $t_j^x$, j Є [1,|W|]. Therefore $T_w^x=\{t_j^x,.,.,t_k^x\}$ j, k ∈[1,|W|] and j ≤ k for the set of all transaction-ids where X occurs in transaction window $T_{SW}$.

PERIOD OF X IN $T_{SW}$:

Let $t_j$ X and $t_{j+1}$ X are two consecutive transaction-ids in operation sliding window $T_{sw}$. The number of trades between $t_j^x$ and $t_{j+1}^x$ is defined as a period of X, say $p_x$ where $p_x = t_{j+1}^x - t_j^x$ , j∈(1, |W|). We consider the first exchange is $t_f$ which is invalid exchange i.e $t_f$= 0 and last exchange is $t_l$ which is last exchange in the $T_{sw}$.

For example, consider the Table1 of dynamic database where first transaction sliding window $T_{sw1}$ consists of eight trades from tid-1 to tid-8 and size the of the transaction sliding window is 8. Second transaction sliding window $T_{sw2}$ contains 8 transactions starting from tid-3 to tid-10.In the first transaction sliding window $T_{sw1}$ pattern < b > appears in the transactions {2, 3, 5} and the periods for pattern < b> are (< $t_f$ - 2 > = 2, < 3-2 > = 1, < 5-3 > = 2, < $t_l$ – 5> = 3). Similarly pattern < b c > appears in transactions {1, 5, 7} and the periods for < b c > are (< 1-$t_f$) = 1, < 5-1 > = 4, < 7-5 > = 2, < $t_l$-7 > = 1) where $t_f$ = 0 and $t_l$ = 8. The items < b > and < b c > are provided the information about their occurrence periods of the window $T_{sw1}$.

REGULARITY OF X IN TRANSACTION SLIDING WINDOW $T_{sw}$:

Let $p_w$ be the set of all periods of X in the $T_{sw}$i.e $p_w(x) = \{p_1^x ,\ldots\ldots p_q^x\}$ where q is the highest transaction number for X appears in the window. The regularity of X in the window is defined as reg(X) = max ($\{p_1^x,\ldots\ldots p_q^x\}$). For instance regularity of pattern < b > in $T_{SW1}$ is 3, that is $P_w$ (b) = max < 2, 1, 2, 3 > = 3 and regularity of pattern < b c > is 4 i.e., $P_w$(b c) = max (1, 4, 2, 1) = 4. We say the pattern is regular if it occurs in the specified period otherwise it is not regular pattern. So the regularity of pattern depends on the occurrence behaviour of pattern in the $T_{SW}$.

**REGULAR ITEMSET:**

A arrangement is called ordered pattern if its predictability is not additional than the customer given extreme monotony brink max-reg() is called regular item set.

## 4. MINING REGULAR PATTERNS:

We consider the specimen value-based databank which is in [3] as our seriatim illustration to mine consistent examples. At the initial place we changed over the value-based database DB (Table 1) into erect information design (Table 2) i.e., (X, Tid). X is an itemised and Tid is a value-based Id. Table 2 contains thing sets and their relating exchange ids with regularities. Mining customary itemises strategy is given beneath.

**PROCEDURE**

Input: Dynamic Data base DDB, 0 (support)
Output: Complete set of regular Itemsets
Let $X^k$ <;;;; I a k-item set
2. $P^k$ X = 0 for all $X^k$
3. For each $X^k$

4. Find the period of $X^k$
5. PX=$P_{i+1}$-$P_i$
6. repeat
7. reg($X^k$) = max($P_i$)
8. if reg($X^k$) < = A
9. $X^k$ is regular item set
10. else
11. Delete $X^k$

Journal of Theoretical and Applied Information Technology
30th April 2016. Vol.86. No.3

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

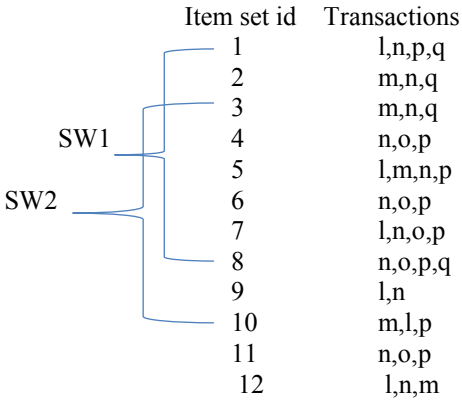



*Table1 Sample Dataset*

Consider a sample transactional sliding window(SW)data in table1 of dynamic database and the size of the $T_{SW}$ is 8 one suppose minimum regularity dawn is. We change of exchanges into succession. Presently we produce standard example which comprises of eight exchanges structure T1 to T8,

*Table1 Sliding Window1 One Itemset*

| One itemset | Period of itemsets X | Regularity value |
|---|---|---|
| L | 1,4,2 | 4 |
| M | 2,1,2,3 | 3 |
| N | 1,1,1,1,1,1,1,1 | 1 |
| O | 4,2,1,1 | 4 |
| P | 1,3,1,1,1,1 | 3 |
| Q | 1,1,1,5 | 5 |

Items < m, n, p > are the only entry which are ordered and entry < l,o,q > are not fixed because they have the monotony significance superior than the given monotony verge assessment.

*Table2 Sliding Window1 Two Itemset*

| Two itemset | Period of itemset X | Regularity value |
|---|---|---|
| (m,n) | 2,1,1,3 | 3 |
| (n,p) | 5 | 5 |
| (n,e) | 1,3,1,1,1,1 | 3 |

For dualistic itemises we can perceive from table 3 merely (m, n) and (n,p) are ordered item set since they have the less normality limit esteem with the prearranged least consistency edge. For three thing set we perform consistent AND manoeuvre on general two thing set.

*Table3 Sliding Window1 Three Itemset*

| Three itemset | period of itemset X | Regularity value |
|---|---|---|
| (m,n,p) | 5,3 | 5 |

In Sliding Window1 the regular pattern are { ( m ), ( n ), ( p ), (m , n), (n, p). In window2 around are eight trades from T2 to T9. We have to deliberate solitary the item of operation T9

*Table4 Sliding Window2 One Itemset.*

| One itemset | Period of itemset X | Regularity value |
|---|---|---|
| L | 4,2,2 | 4 |
| M | 1,1,2,4 | 4 |
| N | 1,1,1,1,1,1,1,1 | 1 |
| O | 3,2,1,1,1 | 3 |
| P | 3,1,1,1,1,1 | 3 |
| Q | 1,1,5,1 | 5 |

The thing n, o, p are one and only normal thing set in window1 in light of the fact that they have consistency limit not exactly the given least normality edge esteem 4

*Table5 Sliding Window2 Using Logical Operation*

| Itemset | Period of itemset X | Regularity value |
|---|---|---|
| N | 1,1,1,1,1,1,1,1 | 1 |
| O | 3,2,1,1,1 | 3 |
| P | 3,1,1,1,1,1 | 3 |

On behalf of three standard itemset we accomplish intelligent AND operation on dual universal itemset.

*Table6 Sliding Window2 Two Itemset*

| Two sItemset | Period of itemset X | Regularity |
|---|---|---|
| n,o | 3,2,1,1,1 | 3 |
| n,p | 3,4,5,6,7 | 7 |
| o,p | 3,2,1,1,1 | 3 |

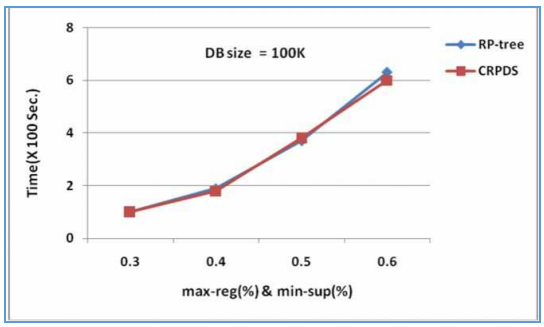
Itemset(n, o) and (o, p) are customary thing set in window 2 since they have normality limit esteem less the are equivalent to the given least consistency edge esteem 4.

*Table7 Sliding Window3 Three Itemset*

| Three Itemset | Period of itemset X | Regularity |
|---|---|---|
| n,o,e | 3,2,1,1,1 | 3 |

Journal of Theoretical and Applied Information Technology
30th April 2016. Vol.86. No.3

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




*Execution Time Of Resulted Work*

## 5. EXPERMENTAL RESULT

In this section we produced our results for closed regular patterns in data streams. We used java to develop our algorithm with the computer configuration of 3.33 GHz CPU with 4GB main memory on windows7 Operating system. We applied our mining process on Kosarak (real data set) and T10I4D100K (synthetic data set).These data sets are often used in successive example mining examination which are created at IBM Almaden quest research group and which are acquired from http://cvs.buu.ac.th/mining/datasets/synthesis_ data and UCI machine repository. The real data set provided by Ferenc Bodan which comprises click brook information of Hungarian on-line news portal.

We used T10I4D100K and Kosarak datasets with various normality and bolster qualities to contrast our outcomes and RP-tree that finds just consistent itemsets. T10I4D100K dataset contains 870 elements with average length of10.10 of 1,00,759 total operations. Kosarak dataset contains 41,270 items with average operation span of 8.10 of 9,90,000 transactions. To produce our results we consider 100K and 500K size of T10I4D100K and kosarak datasets which are shown in figure1 and figure2 respectively. Experimental results shown that higher max-reg() and min- sup() values larger the time vital which are uncovered in both the charts.

## 6. CONCLUSION

Mining regular patterns on dynamic databases using erect data arrangement is completely a novel tactic in data mining applications. We Proposed algorithm to MRPDyD (mining regular patterns on dynamic data bases) to mine systematic itemises in dynamic databases using perpendicular data format with sliding window model. The advantage of our proposed algorithm is it requires guileless actions like addition, subtraction, arrays etc. Our investigational fallouts have shown the effectiveness of MRPDyD in terms of execution time

## REFERENCES

[1] J. Han, Y. Yin, “Mining frequent patterns with out candidata generation”. In Proc. ACM SIGMOD international conf on managt of
Data. pages. 1-12 2000.
[2] Jiawei Han, Micheline Kamber “Data Mining: concepts and techniques.” 2nd ed. An Imprint of Elsevier, Morgan Kaufmann
publishers, pages. 232-248 2006.
[3]R. Agarwal, Imielinski, T. swamy A.N. “Mining association rules between sets of items in large data bases’ ACM, SIGMOD
conference of management of data. pages 207-216 1993.
[4] R. Agarwal., R. srikanth “Fast algorithms for mining association rules” In proc of International conference on very large databases,
Santiago, chile, pages 487-499 sept 1994.
[5]R. Agarwal, Imielinski, T. swamy A.N. "Mining association rules between sets of items in large data bases" ACM, SIGMOD
conference of management of data. pages 207-216 1993.
[6] R. Agarwal., R. srikanth "Fast algorithms for mining association rules" In proc of International conference on very
large databases, Santiago, chile, pages 487-499 sept 1994.
[7] J. Han, Y. Yin, "Mining frequent patterns with out candidate generation". In Proc. ACM SIGMOD international conf on
management of Data. pages. 1-12 2000
[8] S.K. Tanbeer, C.F. Ahmed, B.S.jeong and Y.K.Lee "Mining regular patterns in transactional databases." IEICE trans. On
Infor systems, pages 2568-2577 2008.
[9] Vijay Kumar, G., Sreedevi, M., Pavan kumar, N.V.S.,(2012) “ Mining Regular Patterns in data streams using Vertical format”, IJCSS volume 6 Issue 2, pp142-149.

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

[10] Vijay Kumar, M. Sreedevi, NVS Pavan Kumar "A Vertical Format to Mine Regular Patterns in Incremental Transactional Databases" Journal of Computing, Volume (3) Issue (11) 2011.

[11]G. Vijay Kumar, M. Sreedevi, NVS Pavan Kumar "Mining Regular Patterns in Transactional Databases using Vertical Format" International Journal of Advanced Research in Computer Science, Volume (2), Issue (5) 2011.