Closed Multidimensional Sequential Pattern Mining

Panida Songram Veera Boonjing Sarun Intakosum
Software Systems Engineering Laboratory, Department of Mathematics and Computer Science
King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand

Abstract

We propose a new method, called closed multi-dimensional sequential pattern mining, for mining multidimensional sequential patterns. The new method is an integration of closed sequential pattern mining and closed itemset pattern mining. Based on this method, we show that (1) the number of complete closed multidimensional sequential patterns is not larger than the number of complete multidimensional sequential patterns (2) the set of complete closed multidimensional sequential patterns covers the complete resulting set of multidimensional sequential patterns. In addition, mining using closed itemset pattern mining on multidimensional information would mine only multidimensional information associated with mined closed sequential patterns, and mining using closed sequential pattern mining on sequences would mine only sequences associated with mined closed itemset patterns.

Key-Words: multidimensional pattern mining, closed pattern mining, data mining.

1. Introduction

The purpose of multidimensional sequential pattern mining or MDS pattern mining is to mine sequential patterns in multidimensional circumstances. MDS patterns can cover more useful information than sequential patterns. For example, a computer device shop may find from its database a sequential pattern P1 = (computer, speaker) -> printer -> paper holds for 32% of customers. Such a pattern may be popular for middle age customer group but may not be popular for teenagers. Therefore, the pattern P1 should be associated with a specific customer group to form an MDS pattern such as (middle, P1). This MDS pattern reveals information that the pattern P1 appropriated with middle age customer.

Helen Pinto proposed PSFP and HYBRID algorithms [3] to mine MDS patterns. The PSFP algorithm mines sequences using PrefixSpan algorithm [4] followed by mining multidimensional information associated with mined sequential patterns using FP-growth algorithm [5]. The HYBRID algorithm mines multidimensional information by using BUC algorithm [10] followed by mining sequences associated with mined multidimensional information by using PrefixSpan algorithm. Both PSFP and HYBRID are bottom-up algorithms giving a large set of redundant patterns. The problem of redundancy is similar to those of itemset pattern mining and sequential pattern mining. The closed itemset pattern mining (such as A-CLOSE [13], CLOSET [7], CLOSET+ [9], TFP [6], CHARM [12], GRG [11] and DCI-CLOSE [1]) and the closed sequential pattern mining (such as CloSpan [19], TSP [16] and BIDE [8]) were proposed to eliminate redundant patterns in itemset patterns and sequential patterns, respectively. Therefore, combination of closed itemset pattern mining and closed sequential pattern mining should give closed multidimensional sequential patterns or CMDS patterns. This study proposes to investigate this combination.

The remaining of the paper is organized as follows. In section 2, we define the basic definitions and properties of CMDS patterns. The method of CMDS pattern mining and its correctness are given in section 3 and 4, respectively. Section 5 gives conclusion and future work.

2. Basic definitions and properties

2.1 Basic definitions

Given schema $D = (TID, X_1, X_2, ..., X_m, S)$ is a multidimensional sequence database; where *TID* is a primary key, $X_1, X_2, ..., X_m$ is multidimensional information and *S* is sequences. Let * be any value belong to any domain of $X_1, X_2, ..., X_m$. A multidimensional sequence takes the form of $(a_1, a_2, ..., a_m, s)$, where $a_i \in (X_i \cup \{*\})$ for $(1 \leq i \leq m)$ and *s* is a

Proceedings of the Third International Conference on Information Technology: New Generations (ITNG'06)
0-7695-2497-4/06 $20.00 © 2006 **IEEE**

sequence. A multidimensional sequence $\alpha = (a_1, a_2, ..., a_m, s)$ is said to match a tuple $t = (tid, x_1, x_2, ..., x_m, s_t)$ in multidimensional sequence database if and only if either $a_i = x_i$ or $a_i = *$ and a sequence s is a sub-sequence of a sequence s_t ($s \sqsubseteq s_t$) for $(1 \le i \le m)$. The number of tuples in the database matching a multidimensional sequence α is called the support of α, denoted as $supp(\alpha)$. Given a minimum support threshold *min_support*, a multidimensional sequence α is called an MDS pattern if and only if $supp(\alpha) \ge min\_support$. A set of complete MDS patterns is denoted by $MD = \{\alpha \mid supp(\alpha) \ge min\_support\}$.

Let $I = (a_1, a_2, ..., a_m)$ be multidimensional information and s be a sequence.

Definition 1 A multidimensional sequence $\alpha = (I_m, s_l)$ is a sub-multidimensional sequence of a multidimensional sequence $\beta = (I_n, s_k)$ or $\alpha \subseteq \beta$ if one of the following conditions holds.

1. $I_m \subseteq I_n$ and $s_l = s_k$
2. $I_m = I_n$ and $s_l \sqsubseteq s_k$
3. $I_m \subseteq I_n$ and $s_l \sqsubseteq s_k$

Observation If $supp(\alpha) = supp(\beta)$, then a multidimensional sequence α will be covered by a multidimensional sequence β.

Example 1 Consider table 1, the letters a – h represent items in the sequences, while the number 1 – 3, 4 – 6 and 7 – 9 represent the associated dimension values in unordered dimensions D1, D2 and D3, respectively. Suppose minimum support is 2. Multidimensional sequence (<aa>, *, 6, *) is sub-multidimensional sequence of multidimensional sequence (<aba>, *, 6, 8) ,<aa> $\sqsubseteq$ <aba> and (6) $\subseteq$ (6, 8). Multidimensional sequence (<aba>, *, 6, 8) is called super-multidimensional sequence of multidimensional sequence (<aa>, *, 6, *).

Table 1. Multidimensional sequence database

Id	D1	D2	D3	Sequence
1.	1	5	8	< (b d) c b (a c) >
2.	2	6	7	< (b f) (c e) b (f g)>
3.	1	6	8	< (a h) (b f) a b f>
4.	3	4	9	< (b e)(c e) d >
5.	2	6	8	< a (b d) b c b (a d e)>

Definition 2 Given a sequence s, $P(s)$ is a set of all non-empty sub-sequences of s.

Example 2 From the table1, there are seven sub-sequences of a sequence <cba>:<c>, <b>, <a>, <cb>, <ca>, <ba> and <cba>.

Definition 3 Given multidimensional information I, $P(I)$ is a set of *a*ll non-empty sub-multidimensional information of I.

Example 3 From table 1, there are seven sub-multidimensional information of multidimensional information (1, 2, 8):(1), (2), (8), (1, 2), (1, 8), (2, 8) and (1, 2, 8).

Definition 4 Given a multidimensional sequence $\alpha = (I, s)$, $P(\alpha)$ is a set of cross product of $P(I)$ and $P(s)$.

Example 4 From table 1, there are nine sub-multidimensional sequences of a multidimensional sequence (<ba>, 1, *, 8):(<b>, 1, *, *), (<b>, *, *, 8), (<b>, 1, *, 8), (<a>, 1, *, *), (<a>, *, *, 8), (<a>, 1, *, 8), (<ba>, 1, *, *), (<ba>, *, *, 8) and (<ba>, 1, *, 8).

Definition 5 A set of complete CMDS patterns or *CMD* is defined as followed.

$$CMD = \{\alpha \mid \alpha \in MD \wedge (\nexists\ \beta \in MD \mid \alpha \subseteq \beta \wedge supp(\alpha) = supp(\beta))\}$$

Example 5 From table 1, multidimensional sequence (<ba>, *, 6, 8) is sub-multidimensional sequence of multidimensional sequence (<aba>, *, 6, 8) with the support value 2. Therefore, multidimensional sequence (<ba>, *, 6, 8) is not CMDS pattern. Set of CMDS patterns is {(<aba>, *, 6, 8):2, (<abb>, *, 6, 8):2, (<(bd)bc>, *, *, 8):2, (<(bd)ba>, *, *, 8):2, (<(bf)bf>, *, 6, *):2, (<cba>, *, *, 8):2, (<ba>, 1, *, 8):2, (<bcb>, *, *, 8):2, (<bcb>, 2, 6, *):2 , (<be>, 2, 6, *):2, (<bb>, 1, *, 8):2, (<ba>, *, *, 8):3, (<bb>, *, 6, *):3, (<bb>, *, *, 8):3}.

Definition 6 A set of complete maximal CMDS patterns or *MCMD* is defined as followed.

$$MCMD = \{\alpha \mid \alpha \in CMD \wedge (\nexists \beta \in CMD, \alpha \subset \beta)\}$$

Example 6 From table 1, a multidimensional sequence (<abb>, *, 6, 8) is a maximal CMDS pattern, because there is no a pattern in *CMD* which has a proper super-multidimensional sequence. Therefore, set of complete maximal CMDS patterns is {(<aba>, *, 6, 8), (<abb>, *, 6, 8), (<(bd)bc>, *, *, 8), (<(bd)ba>, *, *, 8), (<(bf)bf>, *, 6, *), (<cba>, *, *, 8), (<ba>, 1, *, 8), (<bcb>, *, *, 8), (<bcb>, 2, 6, *), (<be>, 2, 6, *), (<bb>, 1, *, 8)}.

Definition 7 A set of complete maximal MDS patterns or *MMD* is defined as follow.

$$MMD = \{\alpha \mid \alpha \in MD \wedge (\nexists\ \beta \in MD, \alpha \subset \beta)\}$$

Example 7 From table 1, set of complete maximal MDS patterns is {(<aba>, *, 6, 8), (<abb>, *, 6, 8), (<(bd)bc>, *, *, 8), (<(bd)ba>, *, *, 8), (<(bf)bf>, *, 6,

*), (<cba>, *, *, 8), (<ba>, 1, *, 8), (<bcb>, *, *, 8), (<bcb>, 2, 6, *),(<be>,2,6, *), (<bb>, 1, *, 8)}.

2.2 Properties

Properties of itemset pattern [2, 13, 14 15] have been used to find closed itemset patterns. Therefore, we redefine these properties for finding CMDS patterns as follow.

Property 1 All sub-multidimensional sequences of an MDS pattern are MDS patterns.

Example 8 From table 1, all sub-multidimensional sequences of a MDS pattern (<(bd)>, *, *, 8):2 consist of (<b>, *, *, 8):3, (<d>, *, *, 8):2 and (<(bd)>, *, *, 8):2. Their supports are not lower than minimum support, so they are MDS patterns.

Property 2 All super-multidimensional sequences of a non MDS pattern are not MDS patterns.

Example 9 From table 1, a multidimensional sequence (<(bd)bcb(ade)>, 2, 6, *):1 is not MDS pattern. All super-multidimensional sequences of this pattern consist of (<(bd)bcb(ade)>, 2, 6, 8):1, (<a(bd)bcb(ade)>, 2, 6, *):1 and (<a(bd)bcb(ade)>, 2, 8, 6):1. They are not MDS patterns because their supports are lower than minimum support.

Property 3 The set of maximal MDS patterns is identical to the set of CMDS patterns.

Example 10 From example 6 and 7, we found that the set of maximal MDS patterns and the set of maximal CMDS patterns are the same.

Property 4 The support of a MDS pattern α is equal to the support of the smallest CMDS pattern containing α.

Example 11 The support of a MDS pattern (<ba>, *, 6, *) is equal to the support of a CMDS pattern (<aba>, *, 6, 8), which is the smallest CMDS pattern containing (<ba>, *, 6, *).

3. The method of closed multidimensional sequential pattern mining

Let $CI = \{x_1, x_2, \ldots, x_h\}$ be a set of closed itemset patterns, $CS = \{y_1, y_2, \ldots, y_k\}$ be a set of closed sequential patterns.

The method of CMDS pattern mining consists of two major steps:

1. Combine closed itemset pattern mining with closed sequential pattern mining.
2. Eliminate redundant patterns from the result of the first step.

3.1 Combination of closed itemset pattern mining with closed sequential pattern mining

We integrate closed sequential pattern mining and closed itemset pattern mining using the following two approaches.

- The first approach, *CS* is firstly mined from sequences using closed sequential pattern mining. Then *CCMD* (set of candidate CMDS patterns) for each closed sequential pattern in *CS* is obtained, as shown below, by mining *CI* from multidimensional information using closed itemset pattern mining. The union of all *CCMDs* is called $CCMD_{CS}$.

$$CCMD_1 = y_1 \times CI_1$$
$$CCMD_2 = y_2 \times CI_2$$
$$\ldots$$
$$\ldots$$
$$CCMD_k = y_k \times CI_k$$

- The second approach, *CI* is firstly mined from multidimensional information using closed itemset pattern mining. Then *CCMD* for each closed itemset pattern in *CI* is obtained, as shown below, by mining *CS* from sequences using closed sequential pattern mining. The union of all *CCMDs* is called $CCMD_{CI}$.

$$CCMD_1 = x_1 \times CS_1$$
$$CCMD_2 = x_2 \times CS_2$$
$$\ldots$$
$$\ldots$$
$$CCMD_h = x_h \times CS_h$$

Note that both $CCMD_{CS}$ and $CCMD_{CI}$ are not necessary be sets of CMDS patterns as shown in the following example.

Example 12 From table 1, (6):3 is a closed itemset pattern and <be>:2 is a closed sequential pattern associated with (6). They can be combined to form a MDS pattern (*, 6, *, <be>):2 which is a candidate CMDS pattern in *CCMD*. (2, 6):2 is a closed itemset pattern and <be>:2 is a closed sequential pattern associated with (2, 6). They can be combined to form a MDS pattern (2, 6, *, <be>):2 which is a candidate CMDS pattern in *CCMD*. Definition 1, (2, 6, *, <be>) is a super-multidimensional sequence of (*, 6, *, <be>), and its support is equal to the support of (*, 6, *, <be>). From definition 5, (*, 6, *, <be>) is not a CMDS pattern which is need to eliminate.

3.2 Elimination of redundancy

Given $c, c' \in CCMD$. If $c' \subseteq c$ and *supp(c′) = supp(c)*, then a candidate CMDS pattern c' is redundant and can be eliminated. A set of CMDS patterns containing no redundant element is denoted by *CMD*.

Example 13 From table 1, we know that the supports of both (<bc>,* ,* , 8) and (<(bd)bc>, *, *, 8) are 2 and (<bc>,* ,* , 8) $\subseteq$ (<(bd)bc>, *, *, 8). Then, the pattern (<bc>, *, *, 8) is redundant and can be eliminated.

4. Correctness

The following theorems show correctness of the proposed method of CMDS pattern mining.

Theorem 1 $CMD \subseteq MD$, where *CMD* and *MD* are sets of CMDS patterns and MDS patterns, respectively.

Proof Assume $(y,x) \in CCMD$, where $y \in CS$ and $x \in CI$. We know that $CI \subseteq FI$, where *CI* is a set of closed itemset patterns, *FI* is a set of itemset patterns; and $CS \subseteq FS$, where *CS* is a set of closed sequential patterns and *FS* is a set of sequential patterns. Therefore, $y \in FS$ and $x \in FI$. This implies $(y,x) \in MD$. Therefore, $CCMD \subseteq MD$.

From the proposed method, we know that $CMD \subseteq CCMD$. Therefore, $CMD \subseteq MD$ ∈

Theorem 1 implies that the number of set of CMDS patterns mined from a database is not larger than the number of set MDS patterns mined from the same database. This is denoted by $|CMD| \leq |MD|$.

Theorem 2 The set of MDS patterns can be generated from the set of CMDS patterns.

Proof Based on property 3, all MDS patterns can be derived from the maximal CMDS patterns. Based on property 4, the support of each MDS pattern can be derived from the support of CMDS pattern. We can conclude that the set of MDS patterns can be generated from the set of CMDS patterns. ∈

Theorem 3 $(s \sqsubseteq s') \wedge (supp(s)=supp(s')) \rightarrow I_s = I_{s'}$, where $I_{s'}$ and I_s are sets of transaction id of multidimensional information associated with a sequence s' and s, respectively.

Proof Assume $(s \sqsubseteq s')$, $(supp(s)=supp(s')$ and $I_s \neq I_{s'}$, so there are two cases to be considered:

Case 1: $|I_s| \neq |I_{s'}|$. We have $supp(s) \neq supp(s')$. This contradicts our assumption. (1)

Case 2: $|I_s| = |I_{s'}|$. We have $s \not\sqsubseteq s'$. This contradicts our assumption. (2)

From (1) and (2), we can conclude that

$(s \sqsubseteq s') \wedge (supp(s) = supp(s')) \rightarrow I_s = I_{s'}$ ∈

Theorem 3 implies that multidimensional information associated with sub-sequences of s' will be the same as multidimensional information associated with s' if the support of sub-sequences of s' is equal to the support of s'. Therefore, closed itemset pattern mining on multidimensional information would mine only multidimensional information associated with mined closed sequential patterns.

Example 14 From table 1, sequence <cb> is sub-sequence of <bcb> with the same support as 3. The multidimensional information associated with a sequence <bcb> is found in transaction id 1, 2 and 5 as shown in table 2. Moreover, the multidimensional information associated with sequence <cb> is also found in transaction id 1, 2 and 5. Therefore, the multidimensional information associated with sequence <bcb> and the multidimensional information associated with sequence <cb> are the same.

Table 2. Multidimensional information associated with sequence <bcb>

Id	D1	D 2	D3
1.	1	5	8
2.	2	6	7
5.	2	6	8

Theorem 4 $(I \subseteq I') \wedge (supp(I) = supp(I')) \rightarrow S_I = S_{I'}$, where $S_{I'}$ and S_I are sets of transaction id of sequences associated with multidimensional information I' and *I*, respectively.

Proof Assume $(I \subseteq I')$, $(supp(I) = supp(I')$, and $S_I \neq S_{I'}$ so there are two cases to be considered:

Case 1: $|S_I| \neq |S_{I'}|$. We have $supp(I) \neq supp(I')$. This contradicts our assumption. (1)

Case 2: $|S_I| = |S_{I'}|$. We have $I \subseteq I'$. This contradicts our assumption. (2)

From (1) and (2), we can conclude that

$(I \subseteq I') \wedge (supp(I) = supp(I')) \rightarrow S_I = S_{I'}$ ∈

Theorem 4 implies that that sequences associated with sub-multidimensional information of I' will be the same as sequences associated with I' if the support of sub-multidimensional information of I' is equal to the support of I'. Therefore, closed sequential pattern mining on sequences would mine only sequences associated with mined closed itemset patterns.

Example 15 From table 1, multidimensional information (2) is sub-multidimensional information of (2, 6) and the support of multidimensional information (2) is equal to the support of multidimensional information (2, 6) as 2. The sequences associated with multidimensional information (2, 6) are found in transaction id 2 and 5 as shown in table 3.

Table 3. Sequences associated with multidimensional information (2, 6)

Id	Sequence
2.	< (b f) (c e) b (f g)>
5.	< a (b d) b c b (a d e)>

Furthermore, the sequences associated with multidimensional information (2) are found in transaction id 2 and 5 too. Thus, the sequences associated with multidimensional information (2) and (2, 6) are the same.

In summary, theorem 1 and 2 guarantee that the number of CMDS patterns is not larger than the number of MDS patterns and CMDS patterns cover all MDS patterns, respectively. Theorem 3 guarantees that mining using closed itemset pattern mining on multidimensional information would mine only multidimensional information associated with mined closed sequential patterns. Theorem 4 guarantees that mining using closed sequential pattern mining on sequences would mine only sequences associated with mined closed itemset patterns.

5. Conclusion and future work

In this study, we propose a new method to mine CMDS patterns based on closed itemset pattern mining and closed sequential pattern mining. The new method consists of two major steps; (1) combination closed itemset pattern mining with closed sequential pattern mining (2) elimination of redundant patterns. The second step is based on the result that patterns obtained from the first step are not necessary CMDS patterns. Based on this method, we show that (1) the number of CMDS patterns is not larger than the number of MDS patterns (2) the set of CMDS patterns can cover the set of MDS patterns (3) mining using closed itemset pattern mining on multidimensional information would mine only multidimensional information associated with mined closed sequential patterns, and mining using closed sequential pattern mining on sequences would mine only sequences associated with mined closed itemset patterns. However, performance of mining using closed itemset pattern mining on multidimensional information associated with mined closed sequential patterns compared with mining using closed sequential pattern mining on sequences associated with mined closed itemset patterns is under experimental investigation. The result of such mining combined with an efficient algorithm for redundancy elimination can be an efficient algorithm for mining CMDS patterns. In addition, the interesting method is the method that can generate non-redundant CMDS patterns. This is another interesting topic for future research.

6. References

[1] C. Lucchese, S. Orlando and R. Perego, "DCI-CLOSED: A Fast and Memory Efficient Algorithm to Mine Frequent Closed Itemsets," In Proc. of IEEE ICDM workshop on Frequent Itemset Mininig Implemention (FIMI'04), Vol. 126, Brighton, UK, November 2004.

[2] D. Lin and Z. Kedem, "Pincer-Search: A New Algorithm for Discovering the Maximum Frequent Set," In Proc. of the 6th Int'l Conference on Extending Database Technology (EDBT'98), pp. 105-119, Valencia, Spain, March 1998.

[3] H. Pinto, J. Han, J. Pei, K. Wang, Q. Chen and U. Dayal, Multi-dimensional Sequential Pattern Mining, M. Sc. Thesis, Simon Fraser University, Canada, April 2001.

[4] J. Han, J. Pei, B. Mortazavi-Asl, H. Pinto, Q. Chen, U. Dayal and M-C. Hsu, "PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth," In Proc. 2001 Int. Conf. Data Engineering (ICDE'01), pp. 215-224, Heidelberg, Germany, April 2001.

[5] J. Han, J. Pei and Y. Yin, "Mining Frequent Patterns without Candidate Generation," In Proc. 2000 ACM-SIGMOD Int. Conf. Management of Data (SIGMOD'00), pp.1-12, Dallas, TX, May 2000.

[6] J. Han, J. Wang, Y. Lu and P. Tzvetkov, "Mining Top-k Frequent Closed Patterns without Minimum Support," In Proc. 2002 Int. Conf. on Data Mining (ICDM'02), pp. 211-218, Maebahi, Japan, December 2002.

[7] J. Pei, J. Han and R. Mao, "CLOSET: An Efficient Algorithm for Mining Frequent Closed Itemsets," In Proc. 2000 ACM-SIGMOD Int. Workshop Data Mining and Knowledge Discovery (DMKD'00), pp. 21-30, May 2000.

[8] J. Wang and J. Han, "BIDE: Efficient Mining of Frequent Closed Sequences," In Proc. of the 20th Int. Conf. on Data Engineering (ICDE'04), pp. 79-90, Boston, MA, USA, April 2004.

[9] J. Wang, J. Han and J. Pei, "CLOSET+: Searching for the Best Strategies for Mining Frequent Closed Itemsets," In Proc. of the 9th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'03), pp. 236-245, Washington, DC, August 2003.

[10] K. Beyer and R. Ramakrishnan, "Bottom-up computation of sparse and iceberg cubes," In Proc. 1999 ACM-SIGMOD Int. Conf. Management of Data

0-7695-2497-4/06 $20.00 © 2006 IEEE

(SIGMOD'99), pp. 359-370, Philadelphia, Pennsylvnia, USA, June 1999.

[11] L. Li, D. Zhai and F. Jin, "GRG: An Efficient Method for Association Rules Mining on Frequent Closed Itemsets," In Proc. of the 2003 IEEE Int. Sym. on Intelligent Control (ISIC'03), pp. 854-859, Houston, Texas, October 2003.

[12] M.J. Zaki and C.J. Hsiao, "CHARM: An Efficient Algorithm for Closed Itemsets Mining," In Proc. 2002 SAIM Int. Conf. Data Mining (SDM'02), pp. 457-473, Arlington, VA, April 2002.

[13] N. Pasquire, Y. Bastide, R. Taovil and L. Lakhal, "Discovering Frequent Closed Itemsets for Association Rules," In Proc. of the 7th Int. Conf. on Database Theory(ICDT'99), pp. 398-416, Jerusalem, Israel, January 1999.

[14] N. Pasquier, Y. Bastide, R. Taouil and L. Lakhal, "Efficient Mining of Association Rules Using Closed Itemset Lattices," Journal of Information Systems, Vol. 24, No.1, pp. 25-46, 1999.

[15] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal, "Mining Based for Associate Rules using Galois Closed Sets," In Research Report, Computer Science Dept., Nice Sophia Antipolis University, 2000.

[16] P. Tzvetkov, X. Yan and J. Han, "TSP: Mining Top-k Closed Sequential Patterns," In Proc. of the 3rd IEEE Int. Conf. on Data Mining (ICDM'03), pp. 347-354, Melbourne, Florida, USA, December 2003.

[17] R. Afshar, Mining Frequent Max and Closed Sequential Patterns, M. Sc. Thesis, University of Alberta, Alberta, 2000.

[18] V. Pudi and R. Haritsa, "Generalized Closed Itemsets for Association Rule Mining," In Proc. of the 19th Int. Conf. on Data Engineering (ICDE'03), pp. 714-716, Bangalore, India, March 2003.

[19] X. Yan, J. Han and R. Afshar, "Clospan: Mining Closed Sequenctial Patterns in Large Database," In Proc. 2003 SIAM Int. Conf. Data Mining (SDM'03), pp. 166-177, San Francisco, CA, May 2003.