Volume 4, Issue 5, May-2017, pp. 169-170 ISSN (O): 2349-7084

International Journal of Computer Engineering In Research Trends

Available online at: www.ijcert.org

Survey on Mining Partially Ordered Sequential Rules

Mr. Sandipkumar Sagare[1*], Prof. Suresh Shirgave [2]

[1] M.E.C.S.E Student, C.S.E Department, D.K.T.E. Society's Textile and Engineering Institute,Ichalkaranji, 416115, India

[2] Professor, I.T. Department, D.K.T.E. Society's Textile and Engineering Institute,Ichalkaranji, 416115, India

Sagaresandip999@gmail.com1, skshirgave@yahoo.com2

Abstract:- Nowadays in various applications such as stock market analysis, e-commerce sequential rule mining is used to extract important data. It majorly includes identification of common multiple sequential rules from given sequence database. One of the general forms of sequential rule mining is Partially Ordered Sequential rules in which listed items in left and right side of rule does not need to be ordered. These partially ordered sequential rules are identified using RuleGrowth Algorithm, TRuleGrowth Algorithm. These algorithms identify partially ordered sequential rules for more generalized decision making. In this paper we are focusing on such algorithms.

Keywords: Sequential rules, sequential patterns, temporal patterns, pattern mining, sequence, data mining.

1. Introduction

There are important data mining and machine learning application areas where the data to be analyzed consists of a sequence of events [7]. Identifying relationships between occurrence of events stored in database is important in various domains as it provides understanding of relations of events to predict the next event [5,10].

In the field of data mining, one of the most popular set of techniques for discovering temporal relations between events in discrete time series is sequential pattern mining, which consists of finding sequences of events that appear frequently in a sequence database. However, prediction of events cannot be done by knowing a sequence of events that appear frequently in a database as number of rules generated by sequential rule mining algorithms is very specific and this may provide wrong result. For this partially ordered sequential rule (POSR) mining is used. In this paper we are concentrating on POSR and different techniques used for it.

2. Literature review

2.1 Sequential pattern mining:

It consist of discovering subsequences that appear in sequence database having support not less than threshold value of support set by the user [1][2][3][4]. J. Pei, J. Han, B. Mortazavi-Asl, H. Pinto Q. Chen, U. Dayal, and M. Hsu proposed the PrefixSpan Approach for mining sequential patterns by pattern-growth. This method uses sequence database recursively to project it on smaller projected database and sequential pattern is grown by only identifying locally ordered sequential rule.

2.2 Mining sequential rules common to several sequences

In this sequential pattern mining algorithms are applied first and then the output of that is further processed to generate rules between the pairs of sequential patterns. [4][5][6][7]. For example consider a sequence database as follows:

SID	SEQUENCES
S1	{a},{b,c},{g}
S2	{a,c},{b},{a ,e}
S3	{a},{b},{c}
S4	{d},{c},{g}

From the above sequence database sequential patterns which respect minimum support and confidence threshold are generated and then the sequential rules are generated of the form A=>B such that A and B are sequential patterns. For example, rule {a}=>{b} can be found if the minimum support and minimum confidence values are 50 percent.

3. Partially Ordered Sequential Rules

Partially ordered sequential rules are more generalized form of standard sequential rule in which order of left side of the rule and side of the rule is not important. For example, in standard sequential rule each of the following sequential rules are considered as different while in POSR these rules are same.

1. {a}{b}{c}=>{d}
2. {a}{c}{b}=>{d}
3. {b}{a}{c}=>{d}
4. {b}{c}{a}=>{d}
5. {c}{a}{b}=>{d}
6. {c}{b}{a}=>{d}

POSR helps to conclude that above rules describe same situation that is customer who bought items a, b, c in any order then bought item d. So these rules allow taking more helpful and accurate decisions. For mining POSR, pattern rule growth approach is used. This approach is extended to accept window size constraints.

P. Fournier-Viger, R. Nkambou, and V. S. Tseng [8] presented RuleGrowth approach for mining POSR common to several sequences. This approach does not use the previous techniques of generating candidate and then testing them. Instead it uses pattern-growth approach for discovering more efficient and valid rules. It finds rules in incremental fashion. It start with two items and then grows by scanning the database to expand their left or right part of rule.

P. Fournier-Viger, Cheng-Wei Wu, V. S. Tseng, Longbing Cao and Roger Nkambou [9] proposed TRuleGrowth approach. This approach takes extra input to RuleGrowth in the form of window size. As in extension to RuleGrowth algorithm, TRuleGrowth algorithm uses window size to discover the rules occurring within the sliding window. This constraint enforced for rule size of 1*1. Accordingly left and right side of the rule is modified.

4. Future Work

In this paper, we have presented two different techniques for partially ordered sequential rule mining. TRuleGrowth algorithm shows better results in term of generating more accurate results compared to RuleGrowth algorithm. This approach can be further extended to apply multithreading technique on the algorithm used in the TRuleGrowth. Multithreaded TRuleGrowth accepts the same input as in the TRuleGrowth and makes use of the multithreading technique by splitting the data and tasks into parallel subtasks. It lets the underlying architecture manage how threads run, either concurrently on single core, or in parallel on multiple cores.

5. References

1. J. Pei, J. Han, B. Mortazavi-Asl, H. Pinto, Q. Chen, U. Dayal, and M. Hsu, "Mining sequential patterns by pattern-growth: The pre-fixspan approach," IEEE Trans. Knowl. Data Eng., vol. 16, no. 10, pp. 1–17, Oct. 2004.

2. R. Agrawal and R. Srikant, "Mining sequential patterns," in Proc. 11th Int. Conf. Data Eng., 1995, pp. 3–14.

3. J. Pei, J. Han, B. Mortazavi-Asl, H. Pinto, Q. Chen, U. Dayal, and M. Hsu, "Mining sequential patterns by pattern-growth: The pre-fixspan approach," IEEE Trans. Knowl. Data Eng., vol. 16, no. 10, pp. 1–17, Oct. 2004.

4. M. J. Zaki, "SPADE: An efficient algorithm for mining frequent sequences," Mach. Learning, vol. 42, no. 1–2, pp. 31–60, 2001.pp. 1–17, Oct. 2004.

5. D. Lo, S.-.C. Khoo, and L. Wong, "Non-redundant sequential rules—Theory and algorithm," Inf. Syst., vol. 34, no. 4/ 5, pp. 438–453, 2009.

6. Y. Zhao, H. Zhang, L. Cao, C. Zhang, and H. Bohlscheid, "Mining both positive and negative impact-oriented sequential rules from transactional data," in Proc. 13th Pacific-Asia Conf. Knowl. Discovery Data Mining, 2009, pp. 656–663.

