International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 22 Issue 1 – MAY 2016.

MINING PARTIALLY-AN EFFECTIVE MEASUREMENT OF ALGORTHIM

R .M Keshav,K .Venkatesan,G Dharani And M .Mohamed Irshadh

B.Tech Computer Science And Engineering Final Year, S.R.M University .Ramapuram Chennai-600089

***Abstract*--Sequential rule mining is very important role in data mining with wide application. Algorithm used to discover sequential rules common which make unable to recognize similar rules .This results in (1) similar rules which can be rated differently.(2) rules are considered as uninteresting when they are taken individually.(3) rules are too specific which help us to making predictions .we address these problems by proposing in sequential terms and in consequence of each rule which are placed in an un-ordered .We propose a algorithm in data mining CM rules for mining the rule .The algorithm play by first finding sequential rules prunes the search the rules that occurs jointly in many sequences in data mining .It eliminates association rules which do not meet the minimum confidence and supports thresholds to the time ordering and scalability . .We categorized the CM rules in three possible ways .First we declare time complexity according to the system analysis provided .Second we compare by performance of an algorithm through literature from CM DEO.IN comparison three types of data are discussed in variety of characteristics .Some of the data sets shows that CM RULES is fast and provide the results in better support . With the help of the algorithm WE propose the application0f the mining in an effective way.**

I. INTRODUCTION

Sequential pattern mining is an important data mining task with wide applications. It consists of discovering sub sequences that are common to multiple sequences. However, sequential patterns found by these algorithms are often misleading for the user. Thereason is that patterns are found solely on the basis of their support (the percentage of sequences in which they occur). For instance, consider the sequential patternVivaldi, Handel, Berlioz meaning that customer bought the music of Vivaldi, Handel and Berlioz in that order. This sequential pattern is said to have a support of 50 %.

A solution to this problem would be to add a measure of the confidence or probability that a pattern will be followed. But adding this information to sequential patterns is not straightforward because they can contain multiple items and sequential pattern mining algorithms have just not been designed for that. An alternative that considers the confidence of a sequential pattern is *sequential mining.*

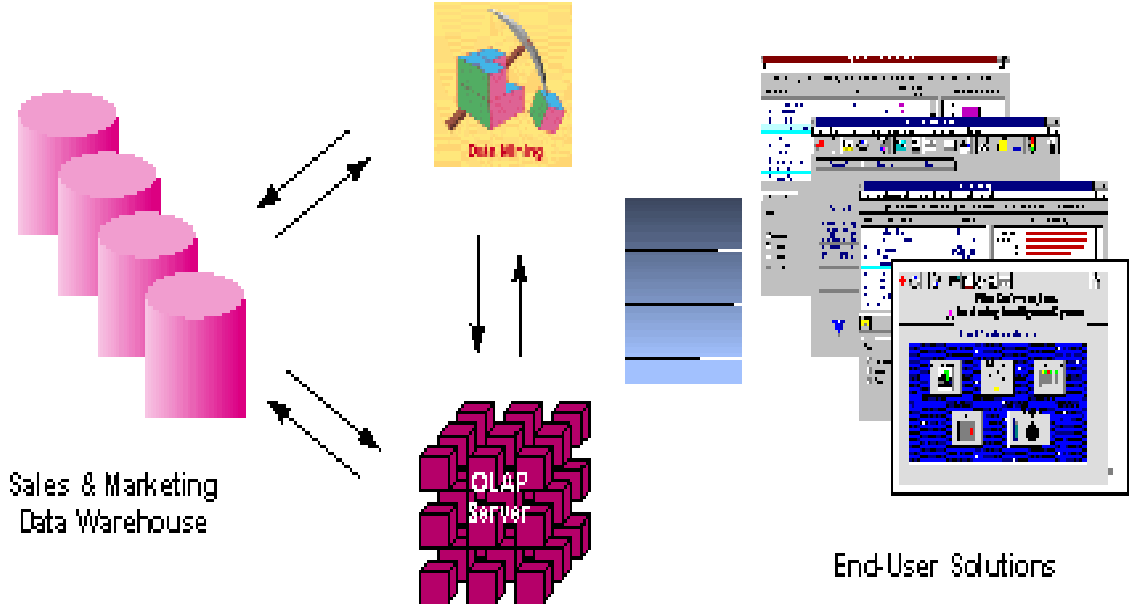
A *sequential rule* indicates that if some event occur, some other event are likely to follow with a given confidence or probability. Sequential rule mining has been applied in several domains such as drought management stock market analysis weather observation reverse engineering elearning and e-commerce. Algorithms for sequential rule mining are designed to either discover rules appearing in a single sequence across sequences or common to multiple sequences.

II. PROBLEM DEFINITION

We note, however, three important problems with thedefinition of a sequential rule as a relationship betweentwo sequential patterns:
1) Rules may have many variations with differentitem ordering. Because sequential patterns specify a strictordering between items, there might be several rules withthe same items but a different ordering. For example,there are 23 variations of {Vivaldi}, {Mozart}, {Handel} ⇒{Berlioz} with the same items ordered differently such aR1: {Vivaldi}, {Mozart}, {Handel} ⇒ {Berlioz},
R2: {Mozart}, {Vivaldi}, {Handel} ⇒ {Berlioz},
R3: {Handel}, {Vivaldi}, {Mozart} ⇒ {Berlioz},
R4: {Handel, Vivaldi}, {Mozart} ⇒ {Berlioz},
R5: {Handel}, {Vivaldi, Mozart} ⇒ {Berlioz},
R6: {Handel, Vivaldi, Mozart} ⇒ {Berlioz}.
But all these variations describe the same situation customers who bought music from Vivaldi, Mozart andHandel in any order, then bought music from Berlioz.
2) Rules and their variations may have important differences how they are rated by the algorithms. Thesedifferences in how variations of the same rules are ratedcan give a wrong impression of the sequential relationshipscontained in the database to the user. In fact, if allthe variations of the same rule were taken as a whole,their support and confidence could be much higher.

III. PROPOSED SYSTEM ARCHITECTURE

To best apply these advanced techniques, they must be fully integrated with a data warehouse as well as flexible interactive business analysis tools. Many data mining tools currently operate outside of the warehouse, requiring extra steps for extracting, importing, and analyzing the data. Furthermore, when new insights require operational implementation, integration with the warehouse simplifies the application of results from data mining. The resulting analytic data warehouse can be applied to improve business processes throughout the organization, in areas such as promotional campaign management, fraud detection, new product rollout, and so on.



The ideal starting point is a data warehouse containing a combination of internal data tracking all customer contact coupled with external market data about competitor activity. Background information on potential customers also provides an excellent basis for prospecting. This warehouse can be implemented in a variety of relational database systems: Sybase, Oracle, Redbrick, and so on, and should be optimized for flexible and fast data access.

- BASELINE ALGORTHIM

In previous work, we proposed two algorithms named CMRules and CMDeo for mining partially-orderedsequential rules, which will be used as baseline algorithmsin this article.CMRules is based on the idea that partially-orderedequential rules can be seen as a subset of associationrules . CMRules performs two steps to discover sequentialrules. First, it ignores the temporal information from the sequence database taken as input to mine associationrules . Then, to obtain sequential rules from association rules, CMRules scans the original database toeliminate rules that do not meet minsup and minconf accordingto the sequential ordering. The main benefits of CMRules are that association rule mining algorithms canbe reused to implement the algorithm and that it performsbetter than CMDeo for some datasets . Its maindrawback is that its performance depends on the numberof association rules. If this set is large, CMRules becomes inefficient .CMDeo proceed It was found that CMDeo performs considerably better than CMRules for some datasets. But for others,the search space is such that CMDeo generates a verylarge number of candidate rules that are invalid, whichmakes CMRules more efficient.

- THE RULE GROWTH ALGORTHIM

A common characteristic of CMRules and CMDeo is thatboth use a generate-candidate-and-test approach, whichconsists of generating candidate rules and then to scanthe database to determine their support and confidence.The problem with this approach is that it often produces alarge amount of candidate rules and that a large proportionare invalid or do not appear in the database. Therefore,these algorithms lot of time to tell apartvalid rules from invalid ones.The RuleGrowth algorithm that we propose in this articleavoids this problem of candidate generation by insteadrelying on a pattern-growth approach partly inspiredby the one used in the PrefixSpan algorithm for sequential pattern mining.

- **ADVANTAGES OF RULE GROWTH ALGORTHIM**

- The performance of RULEGROWTH was compared varying parameters to assess the influence and the performance of each algorithm.
- Second RULEGROWTH was compared to TRULEGROWTH for different windows size values to evaluate a benefits of using the window size constraint.
- Moreover experiment show execution time and the number of valid rules found can be reduced using window constraint.

IV. CONCLUSION

This paper presented two algorithms. RuleGrowth is anovel algorithm for mining sequential rules common tomultiple sequences. Unlike previous algorithms, it uses apattern-growth approach for discovering valid rules suchthat it avoids considering rules not appearing in the database.The second algorithm TRuleGrowthallows the userto specify a slidingwindowconstraintmined. To evaluate RuleGrowth and TRule Growth, weper formed several experiments on four real-life data sets having different characteristics.

REFERENCE

[1] J. Pei, J. Han et al., "Mining Sequential Patterns by Pattern-Growth: ThePrefixSpan Approach," IEEE Trans. Knowledge and Data Eng., vol. 16, no.10, pp. 1-17, 2004

[2] P. Fournier-Viger, Knowledge discovery in problem-solving activities, Ph.D. Thesis, Univ.Quebec in Montreal, Montreal, 2010.

[3] Y.L. Hsieh, D.-L. Yang and J. Wu, "Using Data Mining to Study Upstreamand Downstream Causal Relationship in Stock Market," Proc.2006 Joint Conf.Inf. Sc., 2006

[4] M. J. Zaki, "SPADE: An Efficient Algorithm for Mining Frequent Sequences,"Machine

[5] Learning, vol. 42, no.1-2, pp. 31-60, 2001.

[6] P. Fournier-Viger, A. Gomariz, M. Campos and R. Thomas, "FastVertical Sequential Pattern Mining Using Co-occurrence Information,"Proc. 18th Pacific-Asia Conference on Knowledge Discovery and Data Mining,Springer, pp. 40-52, 2014.

[7] Y. Zhao, H. Zhang, L. Cao, C. Zhang and H. Bohlscheid, "Mining BothPositive and Negative Impact-Oriented Sequential Rules From TransactionalData,"Proc. 13th Pacific-Asia Conference on Knowledge Discovery andData Mining, Springer, pp. 656-663, 2009.

[8] P. Fournier-Viger and V.S. Tseng, "TNS: Mining Top-K NonRedundantSequential Rules," Proc. 28th Symposium on Applied Computing,ACM Press, pp. 164-166, 2013.

[9] http://www.sourcefordgde.com

[10] http://www.networkcomputing.com/

[11] http://www.ieee.org