International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-3 Issue-1, June 2013

A Recent Survey on Incremental Temporal Association Rule Mining

Pradnya A. Shirsath, Vijay Kumar Verma

Abstract*— One of the most challenging areas in data mining is Association rule mining. Several algorithms have been developed to solve this problem. These algorithms work efficiently with static datasets. But if new records are added time to time to the datasets means if the datasets are incremental in nature, scenario of association rules may changed. Some of the new itemsets may become frequent, while some previously derived frequent set may become infrequent. Due to updated dataset some rules that are already derived may dropped and some new rules may arrive up. For the up to-date rules over the updated dataset, if the association mining technique redo the rule generation process for the whole dataset, based on the frequent itemsets, simply by discarding the earlier computed results, it will inefficient. It is mostly due to the multiple scanning over the older dataset. Recently, temporal data mining has become a core technical data processing technique to deal with changing data. Actually, temporal databases are continually appended or updated so that the discovered rules need to be updated. In this paper we represent the survey of various methods for incremental as well as temporal association rule mining.

Keywords*— Mining, Incremental, Temporal, Inefficient, Frequent pattern.

I. INTRODUCTION

In business applications previous transaction summary is useful for identifying behavior of customer and thus help in business decisions. Recent important applications have called for the need of incremental mining. This is due to the increasing use of the record-based databases whose data is being continuously updated e.g. super market data, stock market data, sales data etc. In general Missing values comprise an important and unavoidable problem in data mining, data management and analysis. Conventional mining techniques are not capable of working with amusingness directly, and require some form of work around or pre-processing [6].

The update problem can be solved by generating the new frequent itemsets. After that, the new association rules can be compute from the new frequent itemsets. After getting the new frequent itemsets an old frequent itemsets could become infrequent in the updated database and an old infrequent itemset might become frequent in the new database. The most important aim of incremental mining techniques is to re-run the mining algorithm on the only updated dataset. In next section various algorithms are discussed for the incremental mining. The concept of incremental mining is shown in below Fig.

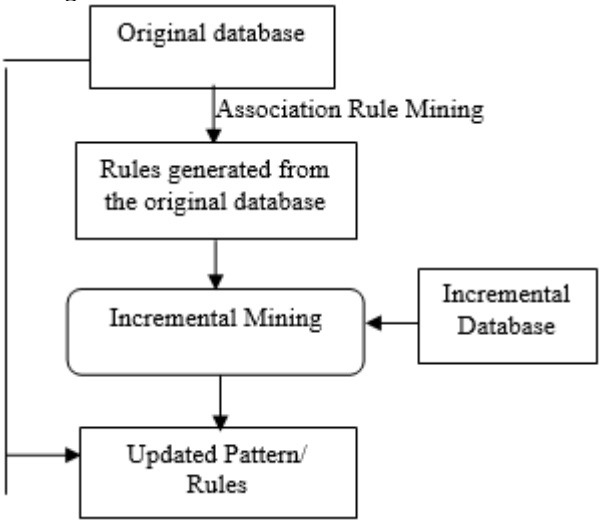

Fig I Process of Incremental Mining

The algorithms proposed for incremental mining were unable to handle data which include time expression. So to solve this problem new concept TAR (Temporal association rule mining) was introduced.

Temporal association rule mining was first introduced by Wang, Yang and Muntz in years 1999-2001. TAR (Temporal Association Rule) algorithm was introduced first to solve the problem on handling time-series by including time expression into association rules. It helps to find the valuable relationship among the different item sets, in temporal database. Temporal association rules are different from traditional association rules. Temporal association rules attempt to model temporal relationships in the data.

Consider a simple example of temporal dataset as shown in below Fig. There are four products A, B, C and D each one has unique exhibition period.

Products A, B and C are exhibited from 2005 to 2013. Product D exhibited from 2007 to 2013 .Product E is exhibited from 2009 to 2013.

However, product F is exhibited from 2011 to 2013. Traditional mining techniques do not take into account the exhibition period of each item and use a unique minimum support threshold [5].

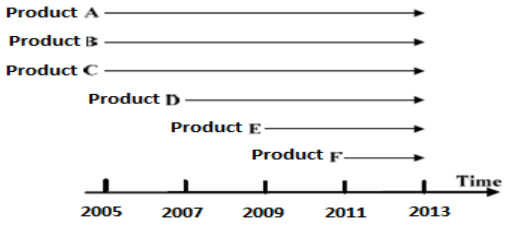

Fig II Example of temporal dataset

Manuscript published on 30 June 2013.
*Correspondence Author(s)
Pradnya A. Shirsath, Department of Computer Science and Engineering, Lord Krishna College of Technology, Indore, MP, India.
Prof. Vijay Kumar Verma, Department of Computer Science and Engineering, Lord Krishna College of Technology, Indore, MP, India.

II. RELATED WORK

Several algorithms have been proposed for mining temporal association rule. These algorithms are based on dividing temporal databases into several partitions according to time and then mining temporal association rules by finding frequent temporal itemsets within these partitions.

A. FUP [Cheung et al., 1996]

Algorithm FUP (Fast Update) handles databases with transaction insertion only, but it is unable to deal with deletion. FUP first scans the incremental part of dataset.

Then the whole dataset (old and new) next it performs similar operation for k itemsets. Finally, after multiple scanning it finds multiple frequent itemsets. The FUP algorithm works efficiently with incremented datasets but it cannot work efficiently with temporal datasets [3].

B. FUP2 [Cheung et al. (1997)]

An extension to FUP algorithm FUP2 is introduced for updating the existing association rules when transactions are added to and deleted from the database. It can deal with insertion as well as deletion [3], [7].

It extracts the rules from the final dataset by considering newly added part and deleted part. FUP2 algorithm also has poor performance if it applies on temporal datasets.

C. PPM [C. H. Lee et al (2002)]

Progressive Partition Miner is proposed to discover general temporal association rule. In PPM the database is first partitioned by the size of time granularity. The PPM algorithm is applying with a filtering threshold mechanism on each partition of the database to prune out those cumulatively infrequent 2-itemsets. PPM is scan reduction technique.

PPM algorithm works efficiently with temporal databases. However, limitation of this technique is its ability to deal with problems of incremental mining [3].

D. SPF [Cheng. Y. Chang et (2002)]

Segment Progressive filter algorithm first divides the database into certain imposed time granularity. Then in illumination of exhibition period of each item, it further segments the database based on their common starting and ending times. For each part of the database it finds the 2-candidate item set with a cumulative filtering threshold.

SPF is also employs scan reduction technique for generating candidate K-item set. After generating all candidates it generates the sub-candidate and counts for the value of support.

Temporal databases are continuously updated or appended. This algorithm first finds out the old frequent itemsets that will remain frequent in the updated dataset also. Then it checks for possible new frequent itemset, and if found then new candidates are generated. Limitation of SPF algorithm is it does not perform any incremental mining technique on the refreshing database [10].

E. NFUP [Chin-Chen Chang Yu-Chiang Li 2005]

The basic idea behind of previous incremental mining techniques is to reduce the number of times that databases need to be scanned. Although those techniques may avoid some unnecessary scanning, they do rescan the original database [6]. The original database is normally much bigger than the incremental database. Therefore, scanning the original database is time-consuming. So new fast update algorithm (NFUP) is introduced.

NFUP does not require the rescanning of the original database to detect new frequent itemsets [5], [18]. NFUP partition the incremental database (db) by the time interval.

The frequent set of itemsets of db is known in advance. NFUP scans each partition backward; the last partition is scanned first.

F. Modified Borders [Das and Bhattacharyya 2005]

This algorithm is a modified version of the borders algorithm that minimizes unnecessary candidate generations. However, this algorithm uses an additional user parameter apart from the parameter support count which is sensitive. With proper tuning of these parameters only- a better performance of the algorithm is possible.

When this additional parameter's value is closer to the support count, the algorithm converges to the borders algorithm. Depending on this parameter, the border sets has been divided into four different sets B', B", B"' and B"". The probability of becoming promoted border set is the highest for the elements of B' and lowest for B"" [7], [3].

G. ISPF [Mohsin Naqvi1, Kashi Hussain1, Sohail Asghar1, Simon Fong (2007)]

Incremental Standing method for Segment Progressive Filter modifies the frequent patterns in pace with changes to the database over the time.

ISPF first divides the database according to the common start and end times of the item sets, and it considers the updates of the temporal association rules. ISPF optimized such that scanning of database is minimized [10].

H. ITARM [Tarek et al 2010]

Proposed Incremental Mining of Temporal Association Rules algorithm to discover the temporal frequent item set after the temporal transaction database has been updated. The basic idea of ITARM algorithm depends on previously generated 2-candidate item set with their supports. ITRAM works as it checks first the extension of the pervious partition and attempts to find 2-candidate item set from the new partition; if it succeeds then it merges the current partition with the pervious partition, and from there it finds the 2-candidate item set. This approach is basically introduced to facilitate incremental mining techniques over an ever updating transaction database [5], [9]. ITARM uses tree method to find frequent itemset, but it is very difficult to maintain a tree if databases is incremental in nature.

I. RUPF [Abhay Mundra, Poonam Tomar, Deepak Kulhare]

The maintenance of dynamic dataset is an important problem. RUPF (Rapid Update in Frequent Pattern) algorithm reduces the number of times to scan the database (old and new) to generate frequent pattern. As a result, the algorithm has faster execution time than previous algorithm [8].

III. COMPARATIVE STUDY

Following table shows the comparative study between some of the methods mentioned above

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Table I Comparison between various methods

Name of Method	Scanning Method/ Concept used	Nature of dataset	Performance
FUP	Multiple Scanning and Apriori based	It works on Incremented dataset only	Poor
FUP2	Multiple Scanning	It works on incremented as well as decremented dataset	Poor
NFUP	It scans only incremented dataset not original dataset	It works on incremented datasets only and it can also works with temporal dataset	Average
PPM	Multiple Scanning	It works on incremented datasets only	Good
SPF	Scanning by maximal candidate itemset	It works on incremented, decremented datasets and also temporal datasets	Good
ITARM	Based on FP growth tree	It works on incremented datasets and temporal datasets	Average

The performance of FUP and FUP2 is poor because they need multiple scanning over the original datasets and they cannot work efficiently if they applied on temporal datasets.

The algorithm NFUP has average performance since it scans incremented dataset only no need to rescan the original dataset and it works efficiently with temporal datasets as compared to FUP and FUP2.

The performance of PPM and SPF is good while the performance of ITARM is average because it uses FP growth tree and maintenance of the tree is very difficult

IV. CONCLUSION AND FUTURE WORK

In the real world, databases are dynamic. In dynamic databases new transactions are appended as time advances. Therefore, mining have to be repeated. Valid patterns and rules must to be efficiently generated. Scanning the original database is very expensive and time consuming. Several methods are proposed for avoiding the rescanning of the original database. FUP work for only incremented database and FUP2 will work for both incremented as well as decremented. Other methods like MAAP, Modified borders also reduce the process of database scanning but they have limitation they will not work for incremental temporal datasets. The concept of temporal association rule has been introduced in order to solve the problem of handling time series by including time expressions into association rules. There are several methods like NFUP, ITARM, PPM and SPF where proposed to solve the problem of finding temporal association rules in the transaction database. They also reduce the time needed for generating new candidates by storing candidate 2-itemsets. This study proposes a new algorithm called RFPM (Recurrent Frequent Pattern Mining) algorithm. RFPM partitions the incremental database logically according to unit time interval (month, quarter or year). RFPM progressively accumulates the occurrence count of each candidate according to the partitioning characteristics and works efficiently as compared to previous algorithms.

REFERENCES

1. J. Han, M. Kamber, Data mining, Concepts and techniques, Academic Press, 2003.
2. Arun K. Pujari, Data mining Techniques, University Press (India) Private Limited, 2006 Florida Artificial Intelligence Research Society Conference, 2001
3. B. Nath1, D K Bhattacharyya2 & A Ghosh3, "Discovering Association Rules from Incremental Datasets, International Journal of Computer Science & Communication Vol. 1, No. 2, July-December 2010, pp. 433-441.
4. Tannu Arora1, Rahul Yadav2, "Improved Association Mining Algorithm for Large Dataset", IJCEM International Journal of Computational Engineering & Management, Vol. 13, July 2011
5. Chin-Chen Chang , Yu-Chiang Li, " An Efficient Algorithm for Incremental Mining of Association Rules" Proceedings of the 15th International Workshop on Research Issues in Data Engineering: Stream Data Mining and Applications 2005 IEEE
6. Tarek F. Gharib, Hamed Nassar, Mohamed Taha, Ajith Abraham, "An efficient algorithm for incremental mining of temporal association rules", An International Journal Expert System with Application.
7. Abhay Mundra, Poonam Tomar, Deepak Kulhare, "Rapid Update in Frequent Pattern form Large Dynamic Database to Increase Scalability", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-6, January 2013
8. Abhay Mundra, Poonam Tomar, Deepak Kulhare, " Rapid Update in Frequent Pattern form Large Dynamic Database to Increase Scalability", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-6, January 2013
9. Anour F.A. Dafa-Alla, Ho Sun Shon, Khalid E.K. Saeed, "IMTAR: Incremental Mining of General Temporal Association Rules", Journal of Information Processing Systems, Vol.6, No.2, June 2010
10. Mohsin Naqvi, Kashif Hussain, Sohail Asghar, Simon Fong, "Mining Temporal Association Rules with Incremental Standing for Segment Progressive Filter"
11. Siddharth Shah, N. C. Chauhan, S. D. Bhanderi, "Incremental Mining of Association Rules: A Survey", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (3) , 2012,4071-4074
12. Pauray S.M. Tsai , Chih-Chong Lee , and Arbee L.P. Chen An "Efficient Approach for Incremental Association Rule Mining", Department of Information Management, Ming Hsin Institute of Technology, Hsin-Feng, Hsinchu 304, Taiwan, R.O.C
13. N.L. Sarda N. V. Srinivas, " An Adaptive Algorithm for Incremental Mining of Association Rules" Computer Science and Engineering Indian Institute of Technology Bombay Mumbai, India Downloaded on April 24, 2009 at 06:04 from IEEE Xplore
14. Siddharth Shah, N. C. Chauhan ,S. D. Bhander, "Incremental Mining of Association Rules: A Survey", College, V.V.Nagar, Gujarat, India A. D. Patel Institute of Technology , Gujarat, India International Journal of Computer Science and Information Technologies, Vol. 3 (3) , 2012,4071-4074
15. Animesh Tripathy, Subhalaxmi Das & Prashanta Kumar Patra, "An Association Rule Based Algorithmic Approach to Mine Frequent Pattern in Spatial Database System", International Journal of Computer Science & CommunicationVol. 1, No. 2, July-December 2010, pp. 357-363
16. Sandhya Rani Jetti, Sujatha D, "Mining Frequent Item Sets from incremental database : A single pass approach", International Journal of Scientific & Engineering Research, Volume 2, Issue 12, December-2011 1 ISSN 2229-5518
17. Rakesh Agrawal, Ramakrishnan Srikant, **"**Fast Algorithms for Mining Association Rules", Almaden Research Center 650 Harry Road, San Jose, CA 95120
18. D. Cheung, J. Han, V. Ng, and C. Y. Wong. Maintenance of Discovered Association Rules in Large Databases: An Incremental Updating Technique. Proceedings of the 12th International Conference on Data Engineering, pages 106—114, February 1996

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)