ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 11, November 2013

Incremental Association Rule Mining by Modified Approach of Promising Frequent Itemset Algorithm Based on Bucket Sort Approach

Ms. Anju k.kakkad[1], Ms. Anita Zala[2]

Computer, LDRP-ITR, Gandhinagar, India [1]

Computer, LDRP-ITR, Gandhinagar, India [2]

Abstract: The association rule mining is very useful in market basket analysis, web data analysis, decision making, knowing customer trends etc. there are many application in which stream data mining require association rule mining such as web click stream mining, sensor networks, and network traffic analysis. Data streams are continuous, unbounded, usually come with high speed. In incremental association rule mining as the time goes new transaction are added and old transaction are being obsolete. So old rule may be dropped out and new rule may be arrived in. so in this paper I have introduce new approach of incremental association rule mining for finding frequent itemset and promising frequent itemset based on bucket sort algorithm without scanning old database.

Keywords: dynamic database, promising frequent itemset, incremental association rule mining.

I. INTRODUCTION

one of the most existing application area is association rule mining which discovers hidden knowledge from database.[3]a data stream is an ordered sequence of items that arrives in timely order. Different from data in traditional static databases, data streams have the following characteristics. First, they are continuous, unbounded, and usually come with high speed. Second, the volume of data streams is large and usually with an open end. Third, the data distribution in streams usually changes with time.[2].

The association rule mining problem is to find out all the rules in the form of $X \Rightarrow Y$, where X and $Y \subset I$ are sets of items, called itemsets.[1] The association rule discovery algorithm is usually decomposed into 2 major steps. The first step is find out all large itemsets that have support value exceed a minimum support threshold and the second steps is find out all the association rules that have value exceed a minimum confidence threshold.[1] Traditional association rule mining technique is not working with incremented database so incremental association rule mining technique is necessary. When new transactions are added new rule is required and old rule is obsolete. After adding new transaction old frequent itemset may became infrequent itemset in new database. So most important task is to run incremental association rule in incremented database without scanning old database.

There are many algorithms which are based on incremental association rule mining. 1)FUP(fast update) 2)UWEP(update with early pruning) 3)negative boarder 4)DB-Tree 5)potFP-tree 6)CAN-Tree. This paper is dived into following section. Section II is stream data mining. Section III is proposed algorithm and section IV is conclusion and future work.

II. STREAM DATA MINING

Stream Mining is the process of extracting knowledge structures from continuous, rapid data records. A data stream is an ordered sequence of instances which have limited computing and storage capabilities. In many application of data stream mining, it can be read only once or small number of times. Examples of application where data stream mining is used are computer network traffic, phone conversations, ATM transactions, web searches, and sensor data.[9]

data stream mining can be considered a subfield of data mining, machine learning, and knowledge discovery. In many data stream mining applications, the goal is to predict the class or value of new instances in the data stream given some knowledge about the class membership or values of previous instances in the data stream. Machine learning

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 11, November 2013

techniques can be used to learn this prediction task from labelled examples in an automated fashion.

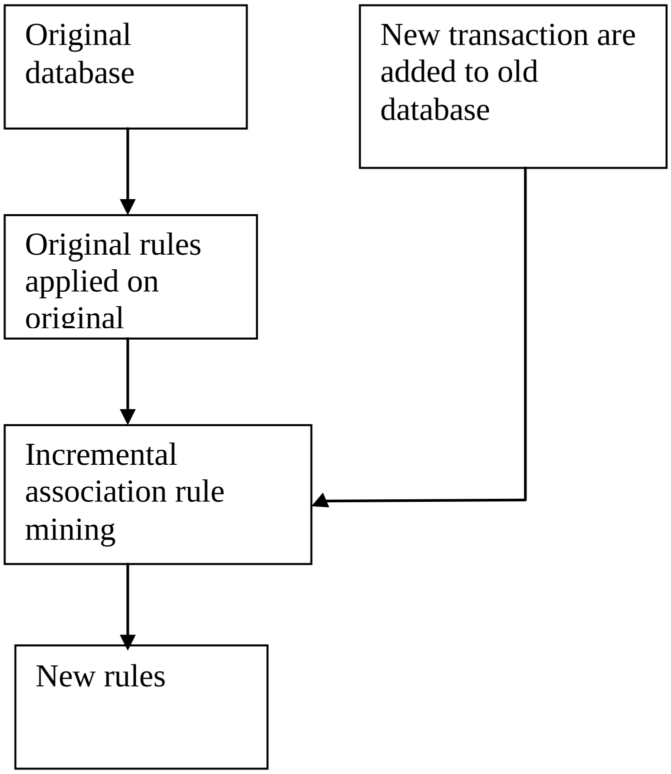


FIG 1:PROCESS OF ASSOCIATION RULE MINING

III EXISTING ALGORITHM

A. FUP (fast update) [Cheung et al., 1996]

FUP is the first algorithm of incremental association rule mining. It works with insertion transaction only it cannot work with deletion transaction. It performs multiple scanning of database i.e. it scans incremented database as well as old database. it performs similar operation for k itemset. Original database D and its corresponding frequent itemsets L = {L1... Lk}. The goal is D'= D∪ Δ+. Here Δ+ is an incremented database.

Cases : Original - Large	**Result**
Case 1 : Large - Large	Always Large
Case 2 : Large – Small	Determined from existing information
Case 3 : Small - Large	Determined by rescanning the original database
Case 4 : Small - Small	Always small

Table I: FUP and its result

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

It is an extension of FUP algorithm. It works with incremented database as well as decremented database. So i.e. it will handle deletion of transaction from old database also. FUP2 is equivalent to FUP for the case of insertion, and is, however, a complementary algorithm of FUP for the case of deletion. For a general case that transactions are added and deleted, algorithm FUP2 can work smoothly with both the deleted portion Δ− and the added portion Δ+ of the whole dataset. It gives poor result if it used with temporal database.

C. *UPDATE WITH EARLY PRUNING(UWEP)*

It is a subset of FUP algorithm. In update with early pruning algorithm it prunes the itemset in original dataset as soon as it became infrequent in updated database D'.[8]. It will not wait until all k^{th} iteration is completed. So it reduces the candidate set generation in incremented database.

D. Negative Border

Negative boarder algorithm is used for improving efficiency of FUP-based algorithm Given a collection of frequent itemsets *L* negative border *Bd*−(*L*) of *L* consists of itemset R which are not in L.[4] In other words, the negative border consists of all itemsets that were candidates of the level-wise method which did not have enough support. This algorithm first scans incremented part of database and then whole database is scanned if and only if itemset outside of negative border gets added to frequent itemset. This may result into increasing size of candidate set generation.

IV PROMISING FREQUENT ITEMSET ALGORITHM BASED ON BUCKET SORT APPROACH

This algorithm is based on incremental association rule mining. In this paper. We introduce new idea of incremental association rule mining which does not scan original database. i.e. without scanning original database it will scan only incremented database. The itemset which are not frequent in original database but it could be frequent when incremented transaction are added to database is called promising frequent itemset.

Maintaining association rules for a dynamic database is an important issue. this paper proposes a new algorithm to deal with such updating situation. This algorithm uses maximum support count of 1-itemset which is calculated before and this will estimate infrequent itemset of original database which is going to be frequent when new transaction are added.

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 11, November 2013

$$\min\_\sup_{DB}-\left(\frac{\max supp}{total\ size}\right)\times inc\_size \le \min\_PL < \min\_\sup_{DB}$$

where min_sup(DB) is minimum support count for an original database, maxsupp is maximum support count of itemsets, current size is a number of transaction of an original database and inc_size is a maximum number of new transactions.

When new transaction is added to an database it may be possible that old frequent temset could be infrequent and promising frequent itemset could be frequent itemset. How to calculate new updated promising frequent itemset is as follows:

$$\min\_PL_{DB\cup db} = \min\_\sup_{DB\cup db} - \left(\frac{\max supp}{total\ size}\times inc\_size\right)$$

After calculating promising frequent itemset of incremented database now apply bucket sort approach on this. Bucket sort approach store promising frequent itemset of incremented database and select bucket based on below formula for frequent itemset.

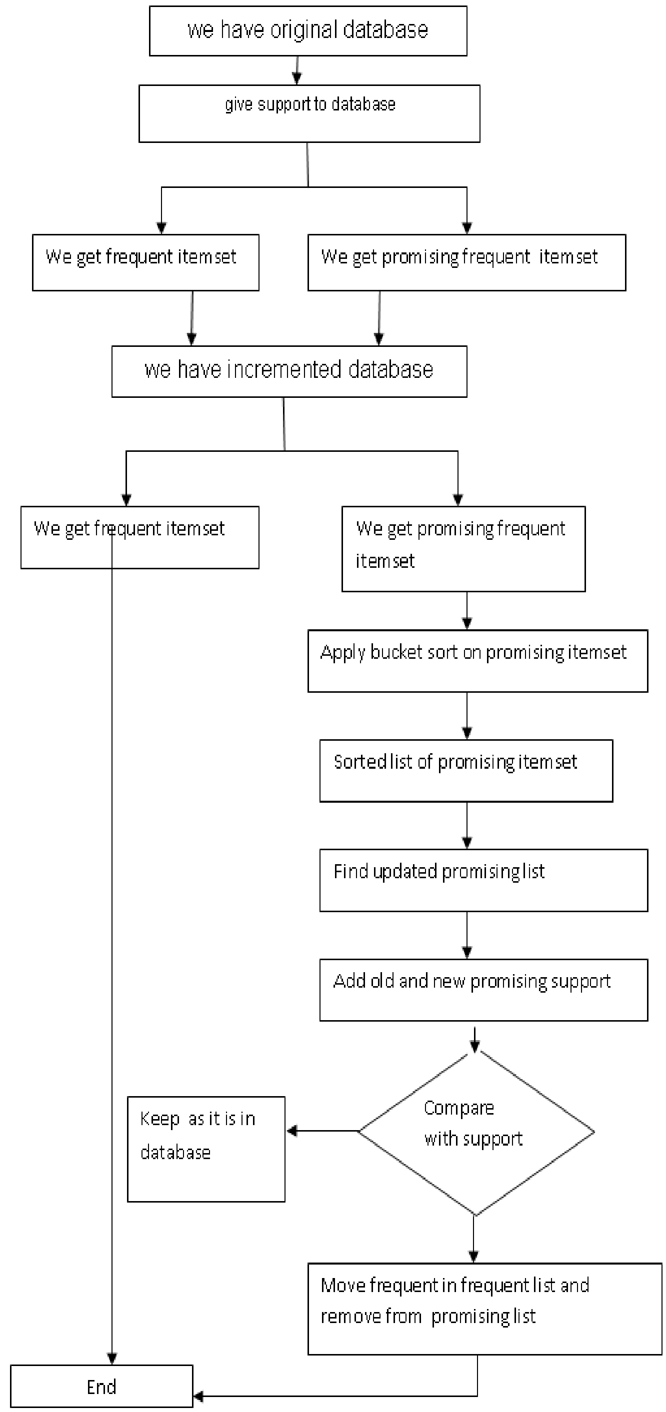
Needed bucket= transaction size / Bucket size

A. Algorithm

Step 1: support is given
Step 2: calculate maximum support of 1-itemset
Step 3: based on above two step calculate promising itemset support

$$\min\_PL_{DB\cup db} = \min\_\sup_{DB\cup db} - \left(\frac{\max supp}{total\ size}\times inc\_size\right)$$

Step 4:check which item are frequent and which are promising frequent item set and make list of both.
Step 5: add incremented database
Step 6: calculate frequent itemset and promising frequent itemset of incremented database
Step 7: now apply bucket sort on incremented promising frequent itemset list.
Step 8: calculate how many buckets are going to be frequent from promising list based on this formula.
Needed bucket = transaction size / Bucket size
Step 9: now add support of needed bucket 's itemset to the support of old promising itemset
Step 10: move new frequent itemset into old frequent itemset list and remove from promising itemset list.



V CONCLUSION FUTURE WORK

This paper concludes that the incremental association approach for mining streamed data works on higher range of the data efficiently. we have covered the old existed algorithm we have reduced the time execution and memory is also reduced, and develop a new approach that directly compare the new as well as old record of the database and

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 11, November 2013

there is no need to add the old database with new coming data and the buckets will give the accurate result. Future work of the paper can be extended to the next higher level with scanning fewer data from the record dataset. Also live out display of new results for streamed data and adding streamed in to main database simultaneously.

VI. ACKNOWLEDGMENT

We would like to thanks my research guide Mrs. Anita Zala for many discussions and for providing a guidelines throughout my research work.

REFERENCES

[1] Jiawei Han and Micheline Kamber Data Mining: Concepts and Techniques, 2nd ed.
[2] chin-chen chang, yu-chiang Li An Efficient Algorithm for Incremental Mining of association Rules 15th international workshop IEEE (2005)
[2] Jun Tan, Yingyong Bu and Haiming Zhao Incremental Maintenance of Association Rules Over Data Streams IEEE 2010
[3] Ratchadaporn Amornchewin Probability-based Incremental Association Rules Discovery Algorithm with Hashing Technique. [International Journal of Machine Learning and Computing, Vol.1, No. 1, April 2011]
[4] Barzan Mozafari, Hetal Thakkar, Carlo Zaniolo Verifying and Mining Frequent Patterns from Large Windows over Data Streams. [IEEE-2008]
[5] Nan Jiang and Le Gruenwald Research Issues in Data Stream Association Rule Mining. [SIGMOD 2006]
[6] Ratchadaporn Amornchewin, Worapoj Kreesuradej Incremental Association Rule Mining Using Promising Frequent Itemset Algorithm [IEEE 2007]
[7] Siddharth Shah, N. C. Chauhan S. D. Bhanderi Incremental Mining of Association Rules: A Survey [
[8] Pradnya A. Shirsath, Vijay Kumar Verma A Recent Survey on Incremental Temporal Association Rule Mining
[9] Lukasz Golab and M. Tamer□Ozsu Issues in Data Stream Management [Lukasz Golab and M. Tamer□Ozsu SIGMOD Record, Vol. 32, No. 2, June 2003