International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 1 Issue 8, October - 2012

Incremental Frequent Pattern Mining: A Recent Review

Abhay Mundra
M.Tech(CSE) IV Sem
Central India Institute of Technology
Indore By-Pass, Arandiya,
Near Mangliya, Indore, M.P

Chaitanya Singh
Asit. Professor
Central India Institute of
Technology Indore

Poonam Tomar
Asit. Professor
Central India Institute of
Technology Indore

Abstract: - In this paper, we provide the preliminaries of basic concepts about incremental frequent pattern mining and survey the list of existing frequent pattern mining techniques. A single article cannot be a complete review of all the algorithms, but we hope that the references cited will cover the major theoretical issues, guiding the researcher in interesting research directions that have yet to be explored. We cover some algorithm through which we explore incremental frequent pattern mining. Incremental frequent mining is useful for maintaining incremental association rule [1,2]

Keywords: - incremental, frequent pattern Association

Introduction :- As data are inserted or deleted from the database the previous frequent item may no longer be interesting new interesting frequent pattern could appear in the updated database .Generally the process of generating new frequent item set using only the updated part of database and the previously generated frequent item is called incremental frequent pattern mining [1,2] Since the data stored in the database is frequently changed the previous frequent item may become stale while novel frequent item could be appear .If this change is not taken into the account and previous frequent item are used for decision making and business analysis a heavy loss may be incurred .Thus incremental frequent pattern mining is very important to business management[1,3,5] .Usually database dealt with huge data sometimes a database table can have one thousand row .If we generate our desired rules from such a huge database it will cost much time and consume many system resource. When maintaining incremental frequent pattern mining the task is to make full use of the previous result, this will reduce run time cost and system resource therefore achieving lower maintenance cost[1,3]

1 General frequent pattern mining methods:-

Agrawal and srikant propose the problem of mining association rules and decompose this problem into two step sub problems name

(1)Generating All large item sets in the database

(2) using the large item set to generate desired rule

The basic idea of this algorithm is to first get all large item set based on their support and then generate desired association rule by computing confidence among large item sets[2,3] Association rule mining is one of the most challenging areas of data mining which was introduced in Agrawal et al., (1993) to discover the associations or co-occurrence among the different attributes of the dataset. Several algorithms like Apriori(Agrawal et. al., 1993), SETM (Houtsma and Swami, 1993), AprioriTID (Agrawal and Srikant, 1994), DIC (Brin et al., 1997), partition algorithm (Savasere et al.,1995), Pincer search (Lin and Kedem, 1998), FP-tree (Han et al., 2000) etc. have been developed to meet the requirements of this problem[3,5,6]. These algorithms work basically in two phases: frequent itemset generation and rule generation. Since the first phase is the most time consuming, all of the above mentioned algorithms mainly focus on the first phase. A set of attributes is termed as frequent set if the occurrence of the set within the dataset is more than a user specified threshold called minimum support. After discovering the frequent itemsets, in the second phase rules are generated with the help of another user parameter called minimum confidence.[4,6]

2 Drawback :-The computation cost in first step much more than the cost in second step since first step needs many database scans ,most association rule generated algorithms mainly concerned with how to quickly generate all large item set in the database[2,5,6,13]

3 Increasing the Efficiency of frequent Pattern mining [7,8]

The computational cost of association rules mining can be reduced in four ways:

- By reducing the number of passes over the Database
- By sampling the database
- By adding extra constraints on the structure of patterns
- Through parallelization.

In recent years much progress has been made in all these directions

4 Incremental frequent pattern mining Methods:- A straight forward methods to deal with updating Incremental frequent pattern mining is based on the apriori algorithm which generates association rule in new database from the scratch, This approach is not efficient and time consuming since many computational could be eliminated if we utilize previous Incremental frequent pattern in old data base An incremental approach utilize the previous Incremental frequent pattern mining result to generate all association rule in the new database .In face some large item set in the old database could remains large in the new database for these large items set it is unnecessary to recomputed their support from scratch since we already have their support in the old database What we need to do is to scans the changed part between old database and new database and get the support from only the changed part of the database. In this case much computational time can be saved

Association rule mining is one of the most challenging areas of data mining which was introduced in Agrawal et al., (1993) to discover the associations or co-occurrence among the different attributes of the dataset. Several algorithms like Apriori(Agrawal et. al., 1993), SETM (Houtsma and Swami, 1993), AprioriTID (Agrawal and Srikant, 1994), DIC (Brin et al., 1997), partition algorithm (Savasere et al.,1995), Pincer search (Lin and Kedem, 1998), FP-tree (Han et al., 2000) etc. have been developed to meet the requirements of this problem. These algorithms work basically in two phases: frequent itemset generation and rule generation. Since the first phase is the most time consuming, all of the above mentioned algorithms mainly focus on the first phase. A set of attributes is termed as frequent set if the occurrence of the set within the dataset is more than a user specified threshold called minimum support. After discovering the frequent itemsets, in the second phase rules are generated with the help of another user parameter called minimum confidence.[4,6]

In incremental database mostly due to the multiple scanning over the older dataset. If the results of the older dataset are reused for updating the frequent itemsets, then a significant amount of time may be saved. Association mining over dynamic dataset is a challenging area of research for the data mining researchers. Several recent works can be found in the literature to meet this challenge. Some of them are

5 Apriori- Based Algorithms for incremental Mining

5.1 Algorithm FUP (Fast UPdate)

Algorithm FUP (Fast UPdate) is the first algorithm proposed to solve the problem of incremental mining of association rules. It handles databases with transaction insertion only, but is not able to deal with transaction deletion. Specifically, given the original database D and its corresponding frequent itemsets L = {L_1, ..., L_k}. The goal is to reuse the information to efficiently obtain the new frequent itemsets L' = {L_1', ..., L_k'} on the new database D' = D ∪Δ+. By utilizing the definition of support and the constraint of minimum support Smin. The following lemmas are generally used in algorithm FUP.

1. An original frequent itemset X, i.e., X Є L, becomes infrequent in D' if and only if X.support D' < Smin.

2. An original infrequent itemset X, i.e., X does not belongs L, may become frequent in D' only if X.support Δ+ ≥ Smin.

3. If a k-itemset X whose (k-1)-subset(s) becomes infrequent, i.e., the subset

is in Lk−1 but not in Lk'−1, X must be infrequent in D'.[5,7,8]

Consider an example
Taking a simple database with 9 transaction s

	TID	Itemset
D	T1	A,B,C
	T2	A,F
	T3	A,B,C,E
	T4	A,B,D,F
	T5	C,F
	T6	A,B,C
	T7	A,B,C E
	T8	C,D,E
	T9	A,B,D,E

Table 1

One Candidate item set C1

Item	Support Count
A	7/9
B	6/9
C	6/9
D	3/9
E	4/9
F	3/9

Table 2

Frequent one item set L1

Item	Minimum Support Count
A	7/9
B	6/9
C	6/9
E	4/9

Table 3

Two candidate item set C2

Item	Support Count
AB	6/9
AC	4/9
AE	3/9
BC	4/9
BE	3/9
CE	3/9

Table 4

Frequent two items set L2

Item	Support Count
AB	6/9
AC	4/9
BC	4/9

Table 5

Form table 5 it clear that

Frequent one item set A,B,C ,frequent two item set AB,AC,BC and finally after repeating the same process for three item set we get ABC. Now After inserting some new transaction and after deleting some old transaction the updated data set D' is now

		TID	Itemset
D'	D	T1	A,B,C
		T2	A,F
		T3	A,B,C,E
		T4	A,B,D,F
		T5	C,F
		T6	A,B,C
		T7	A,B,C E
		T8	C,D,E
		T9	A,B,D,E
	D+	T10	A,B,D
		T11	D,F
		T12	A,B,C,D

Table 6

Now D' is updated database when we perform the same process for this new database D is included in frequent one item set ,Fast Update

Algorithm work for only incremental data base.[11,13]

5.2 Algorithms FUP 2

For a general case that transactions are added and deleted, algorithm FUP2 can work smoothly with both the deleted portion D− and the added portion D+ of the whole dataset. A very feature is that the old frequent k itemsets L_k from the previous mining result is used for dividing the candidate set C_k into two parts: $P_k = C_k \cap L_k$ and $Q_k = C_k - P_k$. In other words, P_k (Q_k) is the set of candidate itemsets that are previously frequent (infrequent) with respect to D. For the candidate itemsets in Q_k, their supports are unknown since they were infrequent in the original database D, posing some difficulties in generating new frequent itemsets. Fortunately, it is noted that if a candidate itemset in Qk is frequent in D−, it must be infrequent in D−. This itemset is further identified to be infrequent in the updated database D’ if it is also infrequent in D+. This technique helps on effectively reducing the number of candidate itemsets to be further checked against the unchanged portion D− which is usually much larger than either D− or D+. [10,11,15]

Take a simple examples

D’		TID	Itemset
	D-	T1	A,B,C
		T2	A,F
		T3	A,B,C,E
		T4	A,B,D,F
		T5	C,F
		T6	A,B,C
		T7	A,B,C E
		T8	C,D,E
		T9	A,B,D,E
	D+	T10	A,B,D
		T11	D,F
		T12	A,B,C,D

Table 7

When we aplly the same process frequent item set are

Frequent one item set	{A},{B},{C},D}
Frequent two item set	{A,B},{A,D},{B,D}
Frequent three item set	{A,B,D}

Table 8

5.3. 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 are 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"".

Reference:-

(1). Association Rule Mining: A Survey Qiankun Zhao, Nanyang and Sourav S. Technological University, Singapore

(2).Discovering Association Rules from Incremental Datasets B. Nath, D K Bhattacharyya International Journal of Computer Science & Communication Vol. 1, No. 2, July-December 2010,

(3) An Efficient Algorithm for Mining Of frequent items using incremental model Nibedita Panigrahi Konark Institute of Science and Technology International Journal of Computer Science & Informatics, Volume-1, Issue-1,2011

(4)IMTAR: Incremental Mining of General Temporal Association Rules Journal of Information Processing Systems, Vol.6, No.2, June 2010 DOI : 10.3745/JIPS.2010.6.2.163 5

(5)Improved Association Mining Algorithm for Large Dataset Tannu Arora1, Rahul Yadav2 IJCEM International Journal of Computational Engineering & Management, Vol. 13, July 2011 ISSN (Online): 2230-7893

(6)A Comparative Study of Association Rules Mining Algorithms Computer & Software Engineering Department, Politehnica University of Timisoara, Bd. Vasile Parvan 2, Timisoara, Romania

(7) An Algorithm for Frequent Pattern Mining Based On Apriori Goswami D.N. et. International Journal on Computer Science and Engineering Vol. 02, No. 04, 2010,

(8) An Improved Apriori-based Algorithm for Association Rules Mining 2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery

(9) An Implementation of Frequent Pattern Mining Algorithm using Dynamic Function International Journal of Computer Applications (0975 – 8887) Volume 9– No.9, November 2010

(10) Mining Dynamic Databases using Probability-Based Incremental Association Rule Discovery Algorithm Journal of Universal Computer Science, vol. 15, no. 12 (2009), 2409-2428 submitted: 15/12/08, accepted: 25/6/09, appeared: 28/6/09 J.UCS

(11) An Algorithm to Discover Calendar-based Temporal Association Rules with Item's Lifespan Restriction Computer Science Department, Graduate School of Engineering Federal University of Rio de Janeiro PO Box 68511, ZIP code: 21945-970 Rio de Janeiro – Brazil

(12) An incremental algorithm for frequent pattern mining based on bit-sequence Wuzhou Dong, Juan Yi, International Journal of Advancements in Computing Technology(IJACT) Volume3, October 2011

(13) DARM: Decremental Association Rules Mining Journal of Intelligent Learning Systems and Applications, 2011, Published Online August 2011

(14) Incremental Association Rule Mining Using Promising Frequent Itemset Algorithm Ratchadaporn Amornchewin Faculty of Information Technology King Mongkut's Institute of Technology Ladkrabang Bangkok, 10520 Thailand

(15) Incremental Mining on Association Rules Wei-Guang Teng and Ming-Syan Chen Department of Electrical Engineering National Taiwan University Taipei, Taiwan,