ASSOCIATION RULE MINING USING IMPROVED FP-GROWTH ALGORITHM

[1]Palak Patel (Student), [2]Prof.Purnima Gandhi (Guide)
Department of Computer Engineering
Alpha College Of Engineering and Technology
Gandhinagar, India.

Abstract: Association rule mining is one of the important technique of which can be defined as discovering meaningful patterns from huge amount of data. Mining frequent item set form databases is very fundamental part of association rule mining. FP-Growth algorithm construct conditional frequent pattern tree and conditional pattern based from database which satisfies the minimum support. FP growth algorithm is required tree storage structure so its execution time is more. For incremental mining Simple FP-Growth algorithm is not Better approach so in this paper, changed the data structure of FP-Growth algorithm and it is applied on incremental database. So, this algorithm can work better for incremental database and produces frequent items and reduce time rather than other incremental mining algorithm.

Keywords: Association Rule Mining, Support, Confidence, FP-Growth Algorithm, Incremental mining

I. INTRODUCTION

Association rule mining is one of the major techniques of data mining. Association rule mining finding frequent patterns, associations, correlations among sets of items or objects in transaction databases. The volume of data is increasing dramatically as the data generated by day-to-day activities. Therefore, mining association rules from large amount of data in the database is interested for many industries which can help in many business decision making processes, such as cross-marketing, and Basket data analysis [1]. Association rule mining can be defined formally as follows: I= {i1, i2, i3… in} is a set of items, such as products like printer, papers …. Let DB be a set of database transactions where each transaction T is a set of items such that T⊆I. Each transaction is associated with unique identifier, transaction identifier (TID). Let X, Y be a set of items, an association rule has the form X→Y, X∩Y=Φ.X is called the antecedent and Y is called the consequent of the rule where, X, Y is a set of items is called as an item set or a pattern [10]. Let freq(X) be the number of rows (transactions) containing the item set in the given database. The support of an association rule is the support of the union of X and Y,
Support (X→Y) =(X U Y)/D. The confidence of an association rule is defined as the percentage of rows in D containing item set X that also contain item set Y,
Confidence (X→Y)=P(X/Y)=support(X U Y) / support(X)
In Association rule mining different approaches are used like horizontal layout based, vertical layout based and projected layout based.

Many researchers have proposed several algorithms to generate frequent pattern but most of them are static. In real world application continuously data is growing. The static calculations are not fit to work productively at whatever point any change happens to the original database.. So for that condition re-apply the algorithm on new database, but in this case it takes more time [8]. This approach is expensive whenever little amount of data is inserted into databases.. The advantage of incremental Mining is to process only incremented part and to discover frequent patterns. Efficiency of these algorithms is depends on the number of passes as well as scans required for processing and number of frequent patterns are generated. The Huge amount of data can be handled efficiently by Incremental Data Mining.

II. IMPROVED FP- GROWTH ALGORITHM

In real world application day by day new transactions are coming and according to that databases are updated. According to that mining frequent pattern is necessary to build. For that many algorithms are already available in association rule mining. Here proposed FP-Growth algorithm which is applied on incremental mining.

Example: 1

TID	**Items**
T100	f,a,c,d,g,i,m,p
T200	a,b,c,f,l,m,o
T300	b,f,h,j,o
T400	b,c,k,s,p
T500	a,f,c,e,l,p,m,n

Table 1: First day Transactions

Minimum support=3
Step1: Scan all transactions and count item's support and put it into ascending order according to its support and put it into table name All_item and item header table is built. Item header table consist:

Itemid	Support count	Node-link
..	..	..

Table 2: Header table

d=1, e=1, g=1, h=1, i=1, j=1, k=1, n=1, s=1, l=2, o=2, a=3, b=3, m=3, p=3, c=4, f=4
Step 2: Then check min support and check condition if item's support=>minimum support and satisfied item's

support put that item's with its support into another table in L-order in linear list table. Linear list table consist:

Item	Count	Pointer

Table 3: Linear list table

Items which are satisfied with minimum support:
a=3,b=3,m=3,p=3,c=4,f=4

Step 3: According to flowchart all transactions are running and generate frequent item sets which are as follows according to Linear list table transactions are scanning one by one.

TID	Items
T100	a,m,p,c,f
T200	a,b,m,c,f
T300	b,f
T400	b,p,c
T500	a,m,p,c,f

Table 4: Transactions (T100 to T500) Linear list table

1) a,m,p,c,f

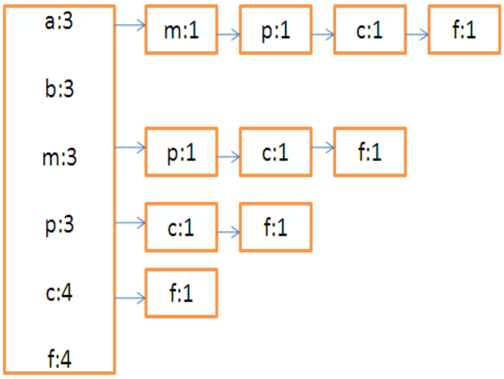

Fig. 1: After scanning T100's Transaction

2) a,b,m,c,f

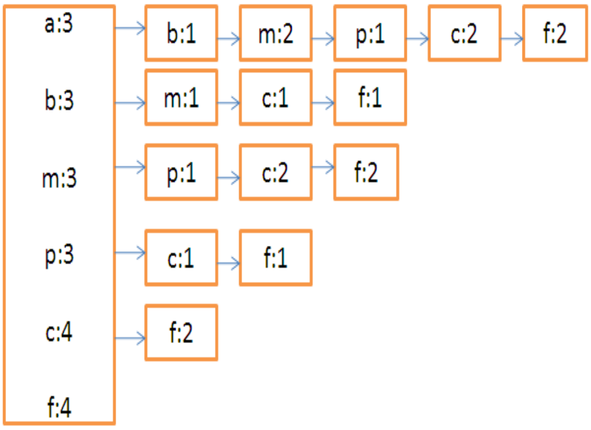

Fig. 2: After scanning T200's Transaction

After that same as apply transactions 3 and 4.At the end of 5th transaction are scanned and frequent item will be generated based on the last transaction.

3) a,m,p,c,f

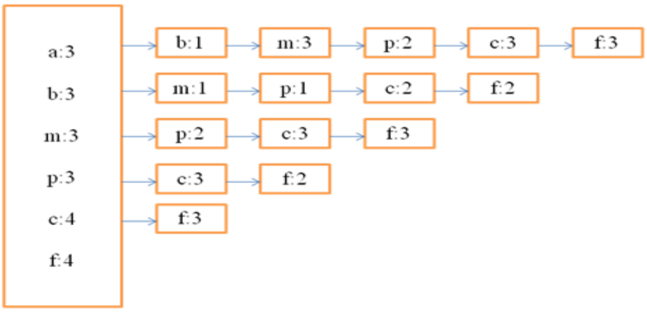

Fig. 3: After scanning T500's Transaction

Rules For Generating Frequent item:

- Take a minimum support in all the item which satisfies item's support=>Minimum Support. Example: Suppose c:4 and f:3 generate frequent pattern cf,so this will contain cf:3,bcause in this minimum support=3.
- In that not included those frequent pattern's which are not greater than equal to Minimum Support. Example: Suppose a: 3 and b:1 are not included as frequent pattern.

Generated Frequent item sets:

Items	Frequent Pattern
f	(f:4)
c	(c:4)(cf:3)
p	(p:3)(pc:3)
m	(m:3)(mc:3)(mf:3)(mcf:3)
b	(b:3)
a	(a:3)(am:3)(ac:3)(af:3)(amc:3)(acf:3)(amf:3)(amcf:3)

Table 5: Generated frequent pattern for T100 To T500

III. IMPROVED FP- GROWTH BASED ON INCREMENTAL DATA SET

For Incremental dataset, after generating frequent item sets move All-item's table database into All-item2 table database. So whenever new day new transactions arrived at that time All-item table is empty.

Step 4: Whenever new transactions are arrived at that time again scan all new transactions and count Item's support and put it into table All-item. Suppose 2 new transactions are arrived it's database is as follows:

TID	Items
T600	a,c,f,m,g,o,l
T700	f,b,a,c,l,m,o,n

Table 6: Second day transactions Example-1

Scanning item's support.
b=1, g=1,l=1,n=1,a=2, c=2, f=2, o=2, m=2

Step 5: Compare item's into All_item table with item's into All-item2 Table. When both item name are same then add that support and put into all-item table. And according to minimum support condition checked and item put into linear list table and then scans transactions one by one. And check condition if item's support do not change then do not consider that item into again scan transcations.so according all conditions linear list table generated and one by one transactions are scanned.

TID	Items
T100	a,m,c,f
T200	b,l,o,a,m,c,f
T300	b,o,f
T400	b,c
T500	l,a,m,c,f
T600	l,o,a,m,c,f
T700	b,l,o,a,m,c,f

Table 7: Transactions (T100 to T700) Linear list table

1) a,m,c,f

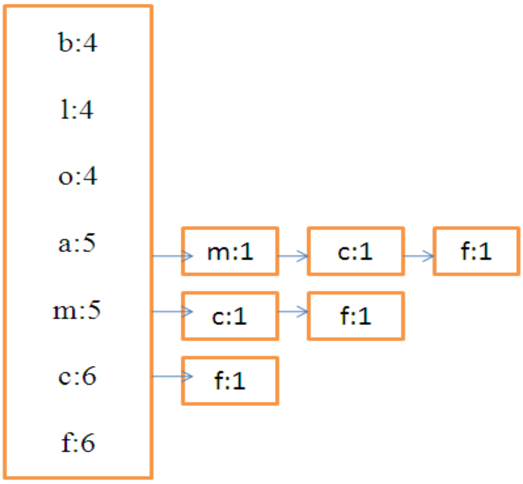

Fig. 4: Scanning T100's Transaction After adding second day transactions

After that same as apply remaining transactions. At the end of last transaction are scanned and frequent item will be generated based on the last transaction.

7) b,l,o,a,m,c,f

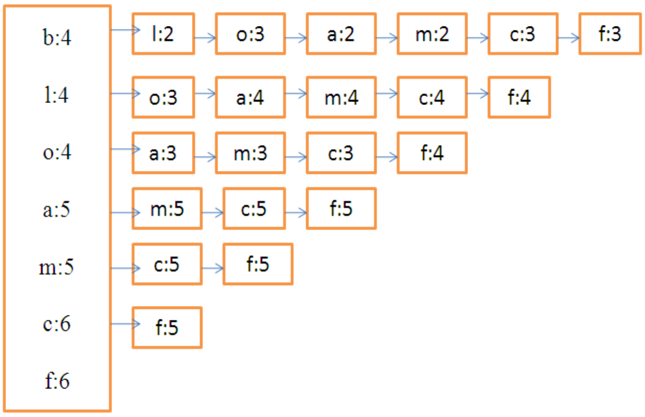

Fig. 5: Scanning T700's Transaction After adding second day transactions

Generated Frequent item sets:

Items	Frequent Pattern
f	(f:6)
c	(c:6)(cf:5)
m	(m:5)(mc:5)(mf:5)(mcf:5)
a	(a:5)(am:5)(ac:5)(af:5)(acf:5)(amc:5)(amf:5)(amcf:5)
o	(o:4)(oa:3)(om:3)(oc:3)(of:4)(ocf:3)(omc:3)(omf:3)(omcf:3)(oam:3)(oac:3)(oaf:3)(oacf:3)(oamc:3)(oamf:3)(oamcf:3)
l	(l:4)(lo:3)(la:4)(lm:4)(lc:4)(lf:4)(lcf:4)(lmc:4)(lmf:4)(lmcf:4)(lam:4)(lac:4)(laf:4)(lacf:4)(lamc:4)(lamf:4)(lamcf:4)(loa:3)(lom:3)(loc:3)(lof:3)(locf:3)(lomc:3)(lomf:3)(lomcf:3)(loam:3)(loac:3)(loaf:3)(loacf:3)(loamc:3)(loamf:3)(loamcf:3)
b	(b:4)(bo:3)(bc:3)(bf:3)(bof:3)(boc:3)(bcf:3)(bocf:3)
p	(p:3)(pc:3)

Table 8: Generated frequent pattern for T100 To T700

IV. RESULTS AND DISCUSSIONS

Dataset characteristics

Datasets	No. of transactions	Items
Dataset-1	7	8
Dataset-2	10	4
Food mart	10000	150

Table 9: Dataset Characteristics

Comparisons between Proposed and simple Fp-growth before applying incremental dataset

Transactions	Proposed Fp Growth Algorithm (in ms)	Simple Fp Growth Algorithm (in ms)
Example-1	1280	1600
Example-2	180	900
Real Dataset	129600	189600

Table 10: Comparisons before applying incremental dataset

Comparisons between Proposed and simple Fp-growth after applying incremental dataset

Transactions	Proposed Fp Growth Algorithm (in ms)	Simple Fp Growth Algorithm (in ms)
Example-1	657	3136
Example-2	540	608
Real Dataset	7312	8650

Table 11: Comparisons before applying incremental dataset

According to Table 10 and Table 11 Proposed Fp- Growth Algorithm takes less time than Simple Fp Growth Algorithm.

V. CONCLUSION

In this paper, new method is proposed for finding frequent pattern in case of incremental database .Experimental results shows that an algorithm is implemented for transactional database and it is generated better output than Simple Fp-Growth Algorithm in terms of time. In Future, by using other

Data structure reduces space complexity.

REFERENCES

[1] Jeetesh Kumar Jain, Nirupama Tiwari, Manoj Ramaiya "A Survey: On Association Rule Mining", International Journal of Engineering Research and Applications, February-2013

[2] Yunlong Song, Ran Wei , "Research on Application of Data Mining Based on FP-Growth Algorithm for Digital Library", IEEE,July, 2011

[3] Kuldeep Singh Malik, Neeraj Raheja "Improving performance of frequent item set algorithm" International Journal of Research in Engineering & Applied Sciences, March, 2013

[4] Chen zhuo, Lu nannan, Li shiqi , Han tao "Research of Association Rule Mining Algorithm Based on Improved FP-Tree", I.J. Engineering and Manufacturing, October, 2012

[5] Banu Priya.M*, Umarani. V "Enriching the Efficiency of Association Rule Mining using Enhanced IFP-Growth Algorithm", International Journal of Advanced Research in Computer Science and Software Engineering, January, 2013

[6] Ding Zhenguo, Wei Qinqin, Ding Xianhua"An Improved FP-growth Algorithm Based on Compound Single Linked List"International Conference on Information and Computing Science(IEEE)March,2009

[7] LIU Jian-ping, WANG Ying, YANG Fan-ding "Incremental mining algorithm Pre-FP in association rules based on FP-tree" First International Conference on Networking and Distributed Computing (IEEE), October**,**2010

[8] Jyoti Jadhav, Lata Ragha, Vijay Katkar "Incremental Frequent Pattern Mining", International Journal of Engineering and Advanced Technology (IJEAT),August, 2012

[9] Chia-Han Yang and Don-Lin Yang "IMBT - A Binary Tree for Efficient Support Counting of Incremental Data", International Conference on Computational Science and Engineering(IEEE) ,August,2009

BOOKS

[10] Ha, Jiawei, Micheline Kamber, and Jian Pei. Data mining: concepts and techniques. Morgan kaufmann, 2006.