International Journal on Applications of Information and Communication Engineering
Volume 1: issue 2: Feburary 2015, PP 33-34 www.aetsjournal.com

Mining High Efficacy Frequent itemsets from Databases using FP-Tree

S.P. Siddique Ibrahim

***Abstract*—** Mining useful itemset from any databases refer to the invention of itemsets with high usefulness like profit. Efficiently discovering of these frequent itemset from large databases is the key components of many data mining technologies. In core algorithm generate large number of candidate itemsets with weight. The number of candidate generation normally reduces the performance of the mining process. In particularly, the database contains large transaction with high utility weight then the available algorithm not suitable for efficient rule generation. FP-growth is generally most efficient among available algorithm for rule generation in data mining. The proposed algorithm is tree based structure namely UP-Tree maintained a information of high utility itemsets. Such that tree can be efficiently constructed with only two scans on the database.

***Keywords*—**Data Mining, FP=growth, UP-Tree .

I. INTRODUCTION

Data mining is the process of exploring important, valuable correlations, earlier unidentified, highly valuable information from any kind of large databases like transactional database, web stream database, relational database, hierarchical database, network database, object oriented database. Normally, the user required data in these databases are hidden and retrieve useful pattern from data warehouse act as major role in diverse data mining tasks, like frequent pattern mining, weighted mining, utility mining [1][2]. Data mining tools predict future result and behavior, which helps companies more focus on their most important information to make positive decision.

Association rule mining is a popular rule generation method to enquire into broad transaction databases for association rules [3][4][5][6]. It has turned into crucial research topic in data mining and has numerous feasible application like cross marketing, drug fragment mining, web site click analysis, medical image processing, transactional databases and time series databases [7][8][9][10][11]. Which is first proposed by Agrawal [12] An Association rule is the form bread->milk meaning that if bread is found in the transaction database, and also good chance of getting milk in the same transaction. The probability of finding Y is based on interesting data mining measures support and confidence. Normally rules that have passed such support and confidence based on certain threshold then the generated rules called frequent rules, association rule mining involves finding such frequent patterns. The discovery of these associations helps supermarket retailers develop interactive market tactic by focus internal product like which items are frequently purchased together by consumers, which are the product sell together, which are product give more profit, current stock position and more.

Mining frequent itemsets is the process of identify the set of items that appear most of the transactions in database. The frequency of an itemset is measured with the help of support and confidence measures. i.e the number of transaction containing the itemset. Mining of frequent itemsets only takes presence and absence of items into account.

Mining high utility itemsets from databases is not an easy task because the itemsets must satisfy the downward closure property in frequent itemset. In FP-Tree item pruning for high utility will satisfy the minimum threshold set by the user. In some cases this problem more difficult to pruning like large search space with databases contains large transactions or low utility itemsets. Existing algorithm often generates a huge set of Potential High Utility Itemsets (PHUIs) [13] then identifies the utility of item set. This may degrade the mining performance in particularly when the database contains much long transaction. Consider the database in table 1. There are five items in the profit table and five transaction in the transaction table in the database.

TABLE 1 : An Example Database

TID	Transaction	TU
T1	(A,2)(B,1)(D,1)	18
T2	(A,1)(B,3)(C,1)(E,2)	18
T3	(A,1)(B,2)(D,3)	22
T4	(A,2)(B,3)(E,3)	21
T5	(C,3)(D,4)(E,2)	30

TABLE 2 : Profit Table

Item	A	B	C	D	E
Profit	6	2	4	4	1

II. PROBLEM STATEMENT

Consider a finite set of items $I = \{i_1,i_2,\ldots i_m\}$, each item i has a unit profit p(i). A transaction database $D = \{T_1,T_2,\ldots T_n\}$ contains a set of transactions, and each transaction T_d $(1\leq d\leq n)$ contains a set of transactions, and each transaction T_d has a unique identifier Tid. Each item I in transaction T_d is associated with a profit $p(i,T_d)$

Definition 1.

S.P. Siddique Ibrahim is Assistant Professor and with Department of Computer Science and Engineering, Kumaraguru college of Technology, An Autonomous Institution under Anna University, Coimbatore-641049, TamilNadu, INDIA (email: siddiqueibrahim.sp.cse@kct.ac.in)

An itemset *I* is said to be frequent whose support count is greater than or equal to support threshold ε. Otherwise, it is called infrequent item.

Definition 2.

An itemset is called a high utility itemset if its utility is no less than user specified threshold ε.

For example, in Table 1 and 2,

U({A},T1)= 2 X 6 = 12

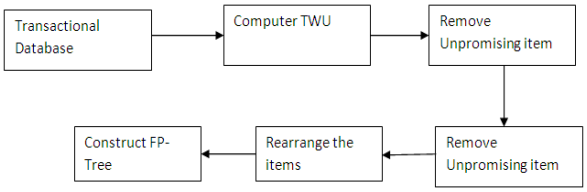
U({A,D},T1)=u({A},T1)+u({D},T1)=12 + 4 = 16

U({A,D})= u({AD},T1) + u({AD},T3) = 16 + 18 =34

Definition 3.

An itemset M is called a high transaction weighted utility itemset if TWU(M) is no less than minimum threshold.

A. Proposed method Work flow



B. Algorithm

Algorithm: UP-Growth(T_n, H_t, I)
// Problem Description: A UP-Tree Tm, a header table Ht for T_n, an itemset I, and minimum threshold min_util.
//Input: Transaction Database T_n
//Output: Identify all PHUIs in T_n
F= ø
Count item Tree <- a new empty FP-Tree
For all weighted transaction in i_k in H_x do
T_u <- Transaction utility(t_q)
If $i_k \geq$ min_sup then
Generate a PHUI F=X U i_k ;
Construct F-LPI
Put local promising items in F-LPI into M_x
End if
End for

REFERENCES

[1] R. Agarwal and R. Srikant," Fast algorithm for mining association rules in large data bases", Proc. of the 20th international conference on very Large Data Base(VLDB'94), Santiago,chile, pp. 487-499, 1994,

[2] J. Han, J. Pei, and Y. Yin, "Mining Frequent Patterns without Candidate Generation", Proc. of the ACM-SIGMOD International conference on Management of Data, Washington,DC, pp. 1-12,2000.

[3] R.Agarwal, T.Imielinski and A.Swami," Mining Association rules between sets of items in large databases", proc. of the ACM-SIGMOD International conference on Management of Data, Washington,DC, pp.207, 1993.

[4] C. Agarwal, C. Procopius and P.S Yu,"Finding Localized Associations in Market Basket Data", IEEE Transaction on Knowledge and Data Eng"., vol.14,No.1, pp.51-62, 2002.

[5] R. Bayardo and R. Agrawal,"Mining the Most Interesting Rules," Pro. ACM-SIGMOD International conference on Knowledge Discovery and Data Mining, pp. 145-154, 1999.

[6] B. Liu, Y.M Wand,"Integrating Classification and Association Rule Mining", Proc. ACM-SIGMOD International conference on Knowledge Discovery and Data Mining, pp. 80-86, 1998.

[7] C.H. Cai, A.W.C. Fu,"Mining Association Rules with Weighted Items", Proc. International Database Engineering and applications, pp. 68-77, 1998.

[8] C.F. Ahmed, S.K. Tanbeer,"Efficient Tree Structures for High Utility Pattern Mining in Incremental Databases", IEEE Transaction on Knowledge and Data Eng"., vol.21, No.12, pp. 1708-1721, 2002.

[9] H.F. Li, H.Y. Huang,"Fast and Memory Efficient Mining of High Utility Itemsets in Data Streams" Proc. IEEE Eighth International conference on Data Mining, pp. 881-886, 2008.

[10] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," Proc. 20th Int'l Conf. Very Large Data Bases (VLDB '94), pp. 487-499, 1994

[11] M.Y. Eltabakh, M. Ouzzani,"Incremental Mining for Frequent Patterns in Evolving Time Series Databases", Technical Report CSD, Purdue University, 2008.

[12] R. Agrawal, A. Swami,"Mining Association Rules between Sets of Items in Large Datasets", Proc. ACM-SIGMOD, pp. 207-216, 1993.

[13] W.Wang, J.Yang and P.S. Yu, "Efficient Mining of Weighted Association Rules (WAR)", IEEE Transaction on Knowledge and Data Eng"., vol.25,No.8, pp.1772-1786, 2013.