ISSN:0975-9646

Mining High BeneficialItemsets from Transactional Database

Supriya P. Bhosale

Department of Information Technology, DGO FOE
Swami-chincoli, Daund-pune.

***Abstract*- Mining high beneficial itemsets from a transactional dB means to identify the itemsets with high utility as profits. Although a number of Algorithms have been invented but there is a problem as it generate large set of candidate Itemsets for High Beneficial Itemsets also require large number of times scan the database.The unit profits and purchased quantities of the items are not taken into considerations in frequent itemset mining and weighted mining only weight is to be considered but not consider the profit. Large number of Item sets decreases the performance of mining with respect to execution time and space requirement. This situation may worse when database contains a large number of Transactions.**

In proposed system for create UP_tree and UP_tree mining algorithms information of high beneficial itemsets is recorded in tree based data structure called Utility Pattern Tree which is a compact representation of items in transaction database. With the help of Utility Pattern Tree, candidate itemsets generated with only two scans of the database. Proposed algorithms not only reduce a number of candidate itemsets but also work efficiently when database has lots of long transactions.

***Keywords*-Candidate itemsets, data mining, high utility itemsets, utility mining**

I. INTRODUCTION

Data Mining is method of retrieving useful information from large databases that are previously unknown. This information has very useful application in data mining tasks. One of the applications is frequent pattern mining but weight of an item is not considered so weighted rule mining came into picture. Still there is a problem because it can not consider quantities so can't satisfy's requirement of user that want itemsets with high profit.

Considering all these issues utility mining came in focus with great advantage in mining high utility mining. High utility itemsets mean items that give high profit to the user, for that minimum utility value is set, and the itemsets that have value greater than that predefined value are favorable items that consider for further processing.Two novel algorithms are proposed as Create UP_tree, that insert transaction DB into tree structure that is compact and UP_tree Mining that give potential high utility itemsets from transaction DB. Discard unfavorable item, reduce a node utility are methods used for these two algorithms.

It plays very important role retail store, e -commerce management. Data mining is the fabulous method of retrieving itemsets from database. Proposed system use transactional dB and mine high beneficial itemsets from it. High utility itemsets are nothing but the itemsets that have highest profits. In existing System, HUP Algorithm is used to mining High Utility Itemsets from database but there are some disadvantages like, it generates huge set of Potential High Utility Itemsets. This system uses UP_Mining Algorithm. Main advantages of this Algorithm are, it scans two time database only, and it generates less set of PHUIs. Aim is to find high beneficial items from the transaction dB when unit profit, quantity is provided.

II. LITERATURE SURVEY

Substantial studies have been done for mining frequent pattern [2] in that Apriori algorithm is used which consider only how frequent item is sold not its weight and quantity also to generate result multiple database scan require. Also, it generates the number of candidates, so next study is on [3] mining frequent pattern without generation of candidates; Frequent pattern tree is generated. It gives frequent items without any candidate key and Search database only twice. It treats all items with the same price. From this algorithm, we get Advantage in our proposed system i.e. scans two times database only.

Comprehensive studies have been done for mining weighted association rule [4] .The association rules related to items that are frequent, and based on weight. In this category work is done on weighted support, and outstanding framework [5]. It considers the importance of the item, in transaction dB items quantities in transactions are not taken into account. From this algorithm, we get Advantage in our proposed system i.e. Considers weight of the item.

Significant studies have been done on high beneficial itemsets mining [6].Now it consider item's frequency, weight and efficiency but still generates a large number of candidates. So to reduce candidates [7] Isolated Items Discarding Strategy is used. In the same category work is done on High Utility Pattern Mining in Incremental Databases using significant tree structure [8]

It provides a remarkable method. Three variations of the tree structure have been proposed. IHUPL-Tree is arranged according to items lexicographic order. IHUP Transaction Frequency Tree (IHUPTF-Tree), which obtains a small size. UP-Transaction-Weighted Utilization Tree (IHUPTWU-Tree) is designed based on the Transaction Weighted Utility value of items in descending order. But it creates a large set of Potential High Utility Items. This situation leads to mining performance is reduced.

III. IMPLEMENTATION DETAILS

3.1 System Architecture and Design:

If you consider previous technology, in that Algorithm multiple numbers of candidate itemsets, are generated. For removing this drawback of previous system, new Algorithm is proposed i.e. create UP_tree and UP_Mining Algorithm. Main intension of this system is reducing itemsets over calculated utilities.

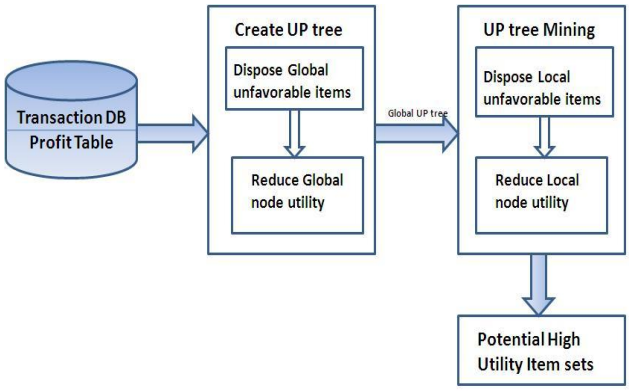

Fig. 1. System Architecture

Figure contains the following blocks:

Transaction DB and Profit table are input to system from that from that system discovers potential high beneficial itemsets.

Create UP_tree: UP_tree is created using dispose of unfavorable the global items and reducing global node utility. UP_tree has fields as Node.name which contain name of the item, Node.count, Node.nu, Node. parent, Node. hlink.
Dispose of global unfavorable items: After calculating transaction utility and transaction weighted utility, the item sets which have utility less than predefined minimum threshold utility is disposed.

Reduce global node utility: After disposing the unfavorable items the node utilities are reduced. And nodes are inserted into UP tree using create UP_tree algorithm.

Dispose local unfavorable items: Construct conditional pattern base of bottom item entry in header table .Retrieve the entire path related to that item CPB. Conditional UP tree created by two scans over CPB. Local unfavorable items removed using path utility of each item in CPB paths are organized in descending order.

Reduce local node utility: Reorganized path is inserted into conditional utility pattern tree using reduce local node utility strategy.

Potential High Utility Itemsets: identify potential high utility itemsets and their utilities form UP tree mining using. Dispose of local unfavorable items and Reduce local node utility.

3.2 Mathematical Model

Variable used in Mathematical Model

Util-un-prom-item=utility of unpromising item
RTU=Recognized Transaction Utility
TranU=Transaction Utility
TranWU=Transaction Weighted Utility
Inputs:

I= {D,P}
Where,
Transactional DB, D = $\{T_1,T_2,..., T_n\}$ is a set of transactions, and for each transaction T_d (1< =d< =n) has a unique id , called T_{id}.

T_d ={(i_1,q_1) ,(i_2,q_2),…..,(i_n,q_n)}
Each item i_p(1<=p<=n) is associated with a quantity $q_p(i_p,T_d)$ that is, the purchased quantity of ip in T_d.

Profit DB, P = {pr(i_1),pr(i_2),……,pr(i_n)}
min-util = user defined minimum threshold.

Process:
1) TranU=$\sum i_p \in T_d$ [pr(i_p)*$q_p(i_p,T_d)$]
2) TranWU(i_p) = $\sum$ TranU $\in i_p$
3) RTU (T_d):= TranU (T_d) - Util-un-prom-item
4) create UP tree and Mine it.

With Node.name, Node. count, Node.nu, Node. parent,

Node.hlink.

Output:

All Potential High Utility Itemsets in T_x

3.3 System Modules

Input to system:
1 Transaction _DB
2 Profit table

List of Modules:
1. Dispose global unfavorable items
2. Reduce global node utility
3. Dispose local unfavorable items
4. Reduce local node utility

Output of system:
Potential high utility itemsets

IV. PRACTICAL ENVIRONMENT

In this section we represent the input dataset and its type, practical results and environment.

4.1 Input Dataset

Transaction database and profit table database are used for the experiment.

4.2 Hardware and software Used

Hardware Requirements:-
1) Operating System: windows XP/ Win7
2) Processor: Pentium IV or advanced
3) RAM: 256 MB (min)
4) HDD: 20 GB (min)

Software Requirements:-
1) Programming Language: Java
2) Framework: Net beans 6.8 or more
3) Development Kit: JDK 1.6 or more

V. RESULT ANALYSIS

This section analyze how system works with respect to different transaction size and different minimum utility. Performance of the proposed algorithms is evaluated in this section. The algorithms are implemented in Java language. Synthetic data sets are used in the experiments.

5.1 TimeVs. No of Transactions

We show the results under different parameters in this part. First, the performance under varied average transaction length (T) is shown in Fig. 2. This experiment is performed on synthetic data sets and minimum utility is set to 50. In Fig. 2, runtimeof UP GROWTH increases with increasing T because when T is larger, transactions and databases become longer and larger.

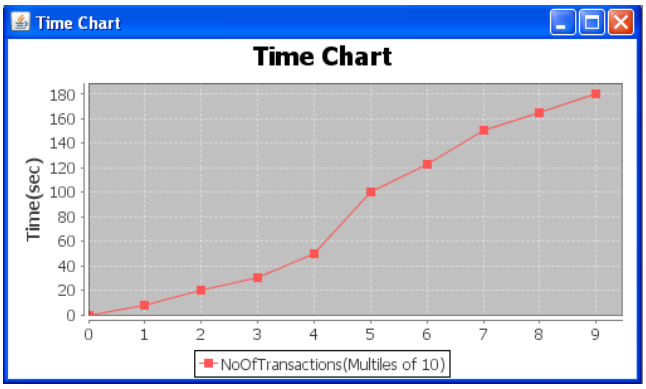

Fig2: Time Vs. No of Transactions

5.2 Utility vs. PHUI

We show the results under different parameters like utility and number of candidates in this part. First, the performance under varied average minimum utility is shown in Fig. 3. This experiment is performed on synthetic data sets and average transaction length is set to 100 In Fig. 3, Number of potential high utility item set decreases with increase in utility threshold.

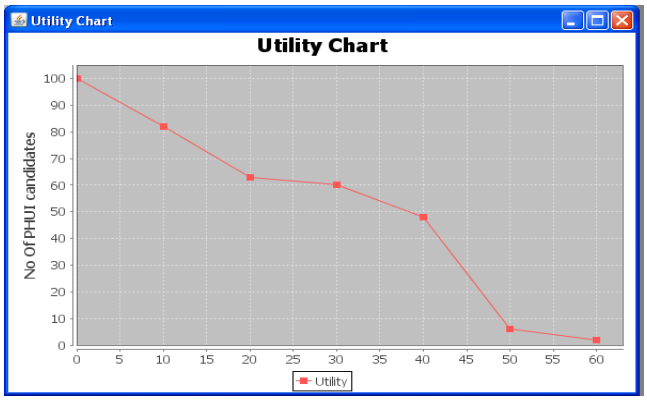

Fig3: Utility vs. PHUI

VI. CONCLUSIONS

Proposed system has tree-based algorithms, called create UP_tree and UP_tree Mining for discovering High beneficial itemsets from databases. Data Structure UP-Tree for recording the information of high beneficial itemsets and four effective strategies, DGU, RGN, DLU and RLN, to minimize search space and the number of candidates for utility mining. Potential high utility itemsets can be generated from Utility Pattern Tree with only two scans of the database. UP-Mining Algorithm is faster than previous algorithms when database have lots of long transactions.

REFERENCES

[1] Vincent S. Tseng, Bai-En Shie, Cheng-Wei Wu, and Philip S. Yu, Fellow,"Efficient Algorithms for Mining High Utility Itemsets from Transactional Databases",IEEE Transaction on knowledge and data engineering, vol. 25, no. 8, Aug 2013.

[2] R. Agrawal and R. Srikant,"Mining Sequential Patterns ", Proc.11th Intl Conf. Data Eng. , pp. 3-14, Mar. 1995.

[3] J. Han, J. Pei, and Y. Yin,"Mining Frequent Patterns without Candidate Generation ", ACM-SIGMOD Intl Conf. Manage-ment of Data ,pp. 1-12, 2000.

[4] C.H. Cai, A.W.C. Fu, C.H. Cheng, and W.W. Kwong,"Mining Association Rules with Weighted Items ", Proc. Intl Database Eng.and Applications Symp. (IDEAS 98) , pp. 68-77, 1998.

[5] F. Tao, F. Murtagh, and M. Farid,"Weighted Association Rule Mining Using Weighted Support and Significance Framework ",ACM SIGKDD Conf. Knowledge Discovery and Data Mining (KDD 03) , pp. 661-666, 2003.

[6] Y. Liu, W. Liao, and A. Choudhary, "A Fast High Utility Itemsets Mining Algorithm ", Proc. Utility-Based Data Mining Workshop 2005.

[7] Y.-C. Li, J.-S. Yeh, and C.-C. Chang,"Isolated Items Discarding Strategy for Discovering High Utility Itemsets ", Data and Knowl-edge Eng vol. 64, no. 1, pp. 198-217, Jan. 2008.

[8] C.F. Ahmed, S.K. Tanbeer, B.-S. Jeong, and Y.-K. Lee, "Efficient Tree Structures for High Utility Pattern Mining in Incremental Databases", IEEE Trans. Knowledge and Data Eng., vol. 21, no. 12,pp. 1708-1721, Dec. 2009