International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 21 Issue 2 – APRIL 2016.

DYNAMIC METHOD FOR MINING THE CONCISE AND LOSSLESS PORTRAYALS OF HIGH UTILITY ITEMSETS

A.Brigetta[1] V.Ajitha[2]

[1]*PG Scholar, Computer Science and Engineering, Saveetha Engineering College, Chennai, Tamil Nadu.*
[2]*Assistant Professor, Department of CSE, Saveetha Engineering College, Chennai, Tamil Nadu.*

[1]abrigetta@gmail.com
[2]ajithanice@gmail.com

***Abstract*— Data mining is the measure of finding data or patterns among diverse fields and here it refers to the discovery of itemset with high utilities which means the interest to user and profit of the product. If large number of items are mined then the utility discovery of collection of dataset is tedious, which also degrades the effectiveness of the mining progression. To attain the high effectiveness for the mining job and provide a concise mining outcome, a closed high utility itemset (CHUIs) is used. The closed high utility itemsets mining identifies the itemsets whose utility satisfies the controller with its profit level based on the products view count and sold count in the transaction record. The rating of product allows users to quantify the usefulness or preferences of items by the people with different values. Apriori is a classic algorithm for learning association rules. Apriori is constructed to operate on databases containing transactions (for example, collections of items bought by customers). DAHU (Derive All High Utility Item sets) a method is used to recover all items from the set of itemset in the transaction without accessing the database. Thus the result will be concise and the items with high utility are found without any loss. The ranking automation is more useful in easy finding of highly utilized product by the customer in the shopping website using the transaction records which hold the view count and the sold count information. Results on absolute and synthetic datasets show that the recommended techniques are very efficient and that our approaches achieve a massive reduction and lossless representation in the number of HUI.**

Keywords—High Utility itemset, Utility mining, Apriori based algorithm.

I. INTRODUCTION

A. Data Mining

Data mining is the measure of revealing nontrivial, previously unknown and potentially useful information from large databases. Discovering useful patterns invisible in a database plays an essential role in numerous data mining tasks, such as frequent pattern mining, weighted frequent pattern mining, and high utility pattern mining. The data mining has been considerably used in the analysis of customer transactions in retail study where it is described as market basket analysis. The process of data mining consists of three stages: the initial examination, construction of model and model detection with validation, verification, and operation. The goal of such techniques is to extract all the frequent item sets, then generate all the valid association rules A → B from frequent itemset A∪B whose confidence has at least the user defined confidence threshold.

B. Frequent Itemset Mining

Frequent itemset mining means finding items that occurs in a database and that such products are frequently purchased by the customer. The FIM do not think about the quantity or benefit of the purchased items. Therefore it is not d for the dynamic user who wants to find the importance of the items in the transaction table (or) database. The high utility mining technique refers to finding itemsets from database which gives high utility. Utility of items is calculated by multiplying internal utility and external utility. The single transaction of itemset is said to be internal utility and the different transaction database included in itemset is said to be external utility. High utility itemset is the itemset which have the sold count is high and if having more view count of the product then it is known as low-utility itemset. In many applications like cross-marketing in retail stores mining such high utility itemsets from databases is an important task.

C. Utility Mining

Utility mining which refers to the quantitative representation of user preference i.e. the utility rate of an itemset is the measurement of the importance of that itemset in the user's perspective. For e.g. if a sales investigator involved in some retail research needs to find out which itemsets in the stores earn the maximum sales revenue for the stores he or she will define the utility of any itemset as the monetary profit that the store earns by selling each unit of that itemset. The utility mining defines two types of utility measures for any itemset, transaction utility and external utility. The external utility of an itemset is based on the information given by the user and is not available in the transactions. For e.g. in case of sales database the external utility may be the profit associated with the sale of itemsets.

II. RELATED WORK

Data mining is the process of revealing nontrivial, previously unknown and potentially useful information from large databases. One of the popular applications is market basket analysis, which assign to the discovery of sets of items

(itemsets) that are frequently purchased together by customers. Hence, FIM cannot satisfy the requirement of users who desire to discover itemsets with high utilities such as high profits. The utility of an itemset represents its importance, which can be measured in terms of weight, profit, cost, quantity or other information depending on the user preference.

A. Transactional Database Mining

In the tranastional database the mining high utility itemsets is to discover the itemsets with high utility like profits (or) income. In terms of time and memory with the large number of itemsets in the transaction then the mining process becomes difficult and that degrades the mining performance. The database containing a long transactions or long high utility itemsets the situation may become difficult. For that algorithms, namely utility pattern growth (UP-Growth) and UP-Growth+, where introduced for mining high utility itemsets with a set of successful strategies for pruning item which have the less utility. The high utility itemsets is managed in a tree-based data structure named utility pattern tree (UP-Tree) such that high itemsets can be generated powerfully with only two scans of database.

B. Fast algorithms for mining association rules

The problem of discovering association rules between items in a large database of sales transactions. It presents two new algorithms for solving this problem that are fundamentally different from the known algorithms. It also show how the best features of the two proposed algorithms can be combined into a hybrid algorithm, called Apriori Hybrid. Apriori Hybrid also has excellent scale-up properties with respect to the transaction size and the number of items in the database. Use the large itemsets to generate the desired rules. For every large itemset l, find all non-empty subsets of l. Algorithms for discovering large itemsets make multiple passes over the data. In the first pass, it counts the support of individual items and determines which of them are large, i.e. have minimum support and minimum confidence and based on that the high utility of the item is found.

C. Mining Top-K Frequent Patterns

Retrieve the top k itemset based on the memory limit and their control. The existing works only concentrate on the efficiency of mining task among the huge list of itemset in the transaction and also in the improving the memory size with reducing the number of unpromising or the isolated item in the transaction which helps to find the frequent itemset mining. Two algorithms used in the top k frequent pattern MTK and MTK _Close which works in mining without specifying the minimum support values. In some cases the human can also specify the range and in some cases the calculated value is also used in the mining the closed utility itemset form a large range of itemset.. The execution efficiency is not compared in the MTK algorithm. The ranking of item in a group of items in a transaction table based on the profit is easy in identification of high utility itemset. The view count and the sold count of the product are manually computed by the controller using the algorithm mining the top k frequent pattern mining. The ranking process is performed automatically when the sold count of the product is incremented and that item will be listed in the top of the transaction log also the product rating and view count of the of particular product is considered.

D. Isolated Items Discarding Strategy

The appearance of an item in a transaction is considered in the association rule mining; also specify the item with binary values. Share mining model is the generalized form of utility mining, which experiment and overcome the problem of having the unpromising item in the transaction list. The idea called Isolated Items Discarding Strategy (IIDS), which can be applied to any level-wise utility mining method to reduce the low utility and to improve the performance. In a transaction database where each transaction contains the set of items or products, the association rules identifies the profitable itemsets from the database. The association rule finds the support and confidence values that should be higher than the minimum support (*minSup*) and minimum confidence (*minConf*) thresholds.

III. PROPOSED SYSTEM

Mining closed high utility itemsets (CHUIs), which serves as compact and lossless portrayals of HUIs. The reduction of low utility itemsets from a group is directly mined in FIM but in the recent techniques everything is considered because a superset of a low utility itemset can be a high utility itemset in some of the subset. In the utility mining, each item has a weight which means the unit profit and that can be shown in the each transaction table. The utility of an item by the people represents its importance, which can be measured in terms of weight, profit, cost, quantity and other information depending on the user preference. The three capable algorithms named, Apriori HC-D, Apriori HC and closed high utility itemset greatly improve its performance.

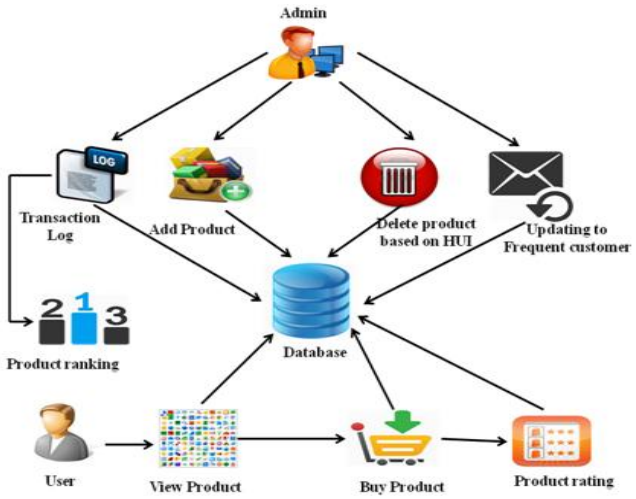

Fig 1: High Utility Itemset

Admin add products into database with product details which include product id, brand name, and product

description with rate to main database with secure operations. It can be retrieved by user in website with product name. Admin views table of transaction in which products transaction details displayed. Updating in transaction table can be accessed by admin. Admin views table of transaction in which products transaction details displayed. Updation in transaction table can be accessed by admin. After viewing the table details, the high utility item sets is backed up and less utility is deleted by admin with view and sold count. The product which has high view count without sold count or only view count is less utility itemset which will be deleted by admin. The product rating is performed by the user based on the previous purchase and by this the usability of the product is known to the unused user. The ranking of product is based on the sold count if the sold count incremented then the the item will be moved to the top of the list among the large dataset. The frequent customer details are collected and send the updated information of their interested product using the email id.

A. Association Rule Mining

Association rules discovery methods find the coincident occurrence of items and build the affinities among them in a transactional database. The methods in the literature are of two types: exhaustive search based algorithms and evolutionary based algorithms. Hence, the frequent itemsets mining phase of the ARM methods plays a vital role in the rule discovery process. The two widely used methods: Apriori and FP-Growth (Frequent Pattern Growth) for discovering the frequent discovering itemsets. Association rule generation by its functionality it is divided into two steps: 1. Apply minimum support to find the frequent itemset in the database 2. These frequent itemsets and the minimum confidence constraint are used to form rules this property states that if an itemset is not a frequent, no superset of that itemset is frequent. This property reduces the search space, but it takes more database scan in order to calculate the frequency of itemsets and it results in increase in execution time and memory overhead.

B. Apriori Algorithm

In data mining Apriori is said to be a classic algorithm for knowing the association rules. Apriori is intended to work on databases containing transactions. , also it is an influential algorithm for mining the frequent itemsets in Boolean association rules.

Key Concepts

•Frequent Itemsets: The number of item which has more support.

•Apriori Property: The frequent itemset must be in subset which is frequent.

•Join Operation: To find specific item from a group of item the difference of the whole itemset is computed. A number of of the methods that are used to improve the efficiency of the apriori algorithm are specified. Apriori generally uses breadth-first search and a tree structure to calculate candidate item sets competently. The downward closure property, the itemset contain all frequent *k*-length item sets. After that, it searches the transaction database to verify frequent item sets among the candidates.

•**Hash-based itemset counting**: For a-itemset hashing bucket range is counted if it is below the range in it is not a frequent purchased item.

•**Transaction reduction**: when frequently scanning the transaction table then number items in the transaction with less hope will be decreased.

•**Partitioning**: While partitioning the database the frequently purchased item will be must in any one of the partition.

•**Sampling**: Here mining the subset of the input data which have less value by that method the completeness is determined.

•**Dynamic itemset counting**: After all the determination of the frequently purchased items the new item for sale is added.

III. IMPLEMENTATION.

A. Product Insertion

The proposed system has many numbers of users, each need an individual login id as username and password for privacy purpose and to get into the purchase point shopping website. The users of the system are only the customers who want to buy the product. For new users the registration process is the first step, registration system requires the basic details of users like username, password, Email id and Address. Details of users are collected and stored in database and during login the username and password are verified, and then get enter into homepage of purchase point.

Collecting data from online sources and enter into database which is known as the data collection. The administrator is commonly called as admin (or) system operator can add products into database with product details which include product id, product name, product price, company, category of the product and product description with rate to main database with secure operations. It can be retrieved by user in website with product name. The image of the product is shown to the customer during the online purchasing in the shopping website. After the successful completion of adding the product the command message called product added successfully is displayed to the admin and now it is stored in the database and the inserted new product can be viewed by the customers (or) user and also the admin in the purchase point website. The uploading are maintained and controlled by the system operator in database, the information about the product the price tag, the category of the product in the shopping website their description are verified completely and also the contact details, services offered by the customer to buy a product and those information are maintained also the contact details.

B. Maintain Transaction log

A transaction indicates a unit of work implemented within a database management system (or identical system) against a database, and treated in a logical and consistent way independent of other transactions. In this project the transaction represents the purchase quantity of the customer. After viewing the product the customer will buy the product and for that they view the product first and that view count of

the product in shopping website also be calculated in the transaction log and to calculate the sold count of the product the confirmation of the order is clicked by the customer then the sold count of the product in the websites are stored in the transaction log. The add to cart button is clicked which means the order is placed, to confirm the order confirm to order button is pressed the customer and now the order is placed. After buying the product by the customer, the transaction details are stored are stored in the transaction table. The transaction table contains attributes such as brand name, brand price, file, product id, the page view, sold no and have the delete button to discard unwanted product from the transaction table which have less utility. Admin views table of transaction in which products transaction details displayed. The updation (or) any modification in transaction table can be accessed by admin.

C. Calculation of View and Sold Count of the Product

The view count and sold count is the major activity performed automatically in the transaction table, at each time the details of the product purchased and product viewed are updated in the transaction log. Also in the table it shows the product with its product id, product name also with the quantity of product purchased by the customer. In Transaction table the number of view count in a product and number of sold count in a product has been displayed. To find the high utility item set without any loss and also to produce the compact portrayals of data the items view count and sold count were calculated. Because without the view count and sold count the high utility of the item set cannot be calculated correctly. When any users who visit the product for purchase and not buy that product then the view count in the transaction table will increase by one. And when the user buys the product based on the quantity the transaction table sold count will be incremented.

D. High Utility Itemset

High utility itemset is itemset which have the utility of high sold count as well as the view count of the product, if having a high view count and the profit of the product is less then it is called a low-utility itemset. In some applications like cross-marketing in retail stores, the mining high utility of the product from databases is an significant task. After viewing the table details, the high utility item sets is backed up and less utility is deleted by admin with view count and sold count. The product which has high view count without sold count or only view count is less utility item set which is considered as less utility item set and product (or) item will be deleted by admin. The product which has high view count and sold count (or) having high sold count that is considered as the high utility item set and that will be remain in the transaction for longer. Normally the HUI describe the frequently purchased itemset that is found only by using the number of view count and sold count in the transaction table.

Utility of each item is cdetermined by multiplying the utilities like internal utility and external utility. Itemset which is under single transaction is said to be internal utility and itemset which is under different transaction database is said to be external utility. A utility mining model was defined to discover additional important knowledge from a database, which can handle the dataset with non binary frequency values of each item in transactions and that is about handling the quantity of the itemset in the transaction table and also with different profit values of each item. Always the utility mining represents the real world market data. By the calculation of utility mining, some important business region decisions like maximizing revenue or minimizing marketing or inventory costs can be considered and knowledge about itemsets/customers contributing to the majority of the profit can be discovered.

IV. RESULT ANALYSIS

The graph shows the number of itemsets with respect to these utility. In the itemsets there are number of items which shows their different utilities in different itemsets. The graph states the total utility of the itemsets. From this graph the itemset which have more utility is found and then discard the low utility itemsets which have the less sold count and more view count .If the utility of the product is determined in terms of the product with its price then the loss of data is high ,for the lossless representation of items with its utility the sold as well as the view count of the product is checked completely and produce the total utility of the product with respect to the itemset.

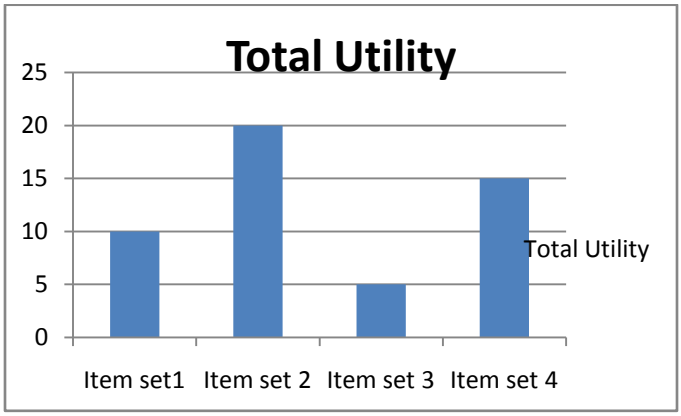

Fig 2: Total Utility Graph

V. CONCLUSION

The high utility itemset mining is developed with the effective algorithm that allows the lossless and compact representation of itemset .The view count and sold count of the product (or) item is added in the transaction log during the purchase by that the item with high utility is found and the low utility items are discarded, which serves as to avoid the large representation of itemset. With the help of this mining technique the amount of performance in mining the high utility itemset from a large amount of itemset is very efficient.

REFERENCES

[1] Agrawal.R and Srikant.R ,(1994)"Fast algorithms for mining association rules," in Proc. 20th Int. Conf. Very Large Data Bases, pp. 487–499.

[2] Ahmed C. F, Tanbeer S. K, Jeong B.S., and Lee Y.K ,(2009) "Efficient tree structures for high utility pattern mining in incremental databases,"IEEETrans.Knowl. Data Eng., vol. 21, no. 12, pp. 1708–1721.

[3] Boulicaut J.F, Bykowski .A , and Rigotti .C ,(2003)"Free-sets: A condensed representation of Boolean data for the approximation offrequency queries, Data Mining Knowl. Discovery, vol.7, no.1,pp.5-22.

[4] Calders.T and Goethals.T, (2002) "Mining all non-derivable frequent itemsets," in Proc. Int. Conf. Eur. Conf. Principles Data Mining Knowl. Discovery pp. 74–85.

[5] Chuang.K, Huang.J, and Chen.M(2008) "Mining top-k frequent patterns in the presence of the memory constraint," VLDB J., vol. 17, pp. 1321–1344.

[6] R.Chan,Yang.G and Shen.Y, (2003) "Mining high utility itemsets," in Proc. IEEE Int. Conf. Data Min., pp. 19–26.

[7] Erwin.A,Gopalan R.P, and Achuthan,N.R (2008) "Efficient mining of high utility itemsets from large datasets," in Proc. Int. Conf. PacificAsia Conf. Knowl. Discovery Data Mining, pp. 554–561.

[8] Gouda.K and Zaki M.J, (2001
"Efficiently mining maximal frequent itemsets," in Proc. IEEE Int. Conf. Data Mining, pp. 163–170.

[9] Hamrouni.T, (2012) "Key roles of closed sets and minimal generators in concise representations of frequent patterns," Intell.Data Anal vol. 16, no. 4, pp. 581–631

[10] Han.J, Pei.J, and Yin.Y, (2000) "Mining frequent patterns without candidate generation," in Proc. ACM SIGMOD Int. Conf. Manage. Data, pp. 1–12.

[11] Hamrouni, Yahia.S, and Nguifo E.M, (2009) "Sweeping the disjunctive search space towards mining new exact concise representations of frequent itemsets," Data Knowl. Eng., vol. 68, no. 10,pp. 1091–1111.

[12] Li.H J, Huang H.Y, Chen.Y, Liu.Y.J , and Lee S.Y, (2008) "Fast and memory efficient mining of high utility itemsets in datastreams," in Proc. IEEE Int. Conf. Data Mining,, pp. 881–886.

[13] Lin C.W, Hong T.P, and Lu W.P, (2011) "An effective tree structure for mining high utility itemsets," Expert Syst. Appl., vol. 38, no. 6, pp. 7419–7424.

[14] Lan G, Hong T, and Tseng S, (2014) "An efficient Projection based indexing approach for mining high utility itemsets," Knowl.Inf. Syst, vol. 38, no. 1, pp. 85–107.

[15] Li H, Li J, Wong J, Feng Y, and Tan Y, (2005) "Relative risk and odds ratio: A data mining perspective," in Proc. ACM SIGACT-SIG -OD-SIGART Symp. Principles Database Syst., pp. 368–377.

[16] Le. B, Nguyen. H, Cao T.A, and B. Vo, (2009) "A novel algorithm for mining high utility itemsets," in Proc. 1st Asian Conf. Intell. Inf.Database Syst., pp. 13–17.

[17] Liu. Y, Liao. W, and Choudhary. A, (2005) "A fast high utility itemsetsmining algorithm," in Proc. Utility-Based Data Mining Workshop , pp. 90–99.

[18] Lucchese .C, Orlando.S, and Perego.R, (2006) "Fast and memory efficient mining of frequent closed itemsets," IEEE Trans. Knowl. DataEng., vol. 18, no. 1, pp. 21–36.

[19] Li. Y. C, .Yeh. J, and Chang. C, (2008) "Isolated items discardingstrategy for discovering high utility itemsets," Data Knowl.Eng.,vol. 64, no. 1, pp. 198–217.

[20] Pasquier. N, Bastide. Y, Taouil. R, and L. Lakhal(199)"Efficient mining of association rules using closed itemset lattice," J. Inf. Syst.,vol 24, no. 1, pp. 25–46.