International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2494-9150 Vol-02, Issue 06, Sep 2016.

A Recent Study of Various Utility Pattern Mining Techniques

[1]Shilpa Shrivastabva, [2]Mr. Abhishek Tiwari

[1]Research Scholar, [2]Reader, [1,2]Information Technology Department, MIT, Ujjain, MP, India.

[1]shrivastava.shilpa21@gmail.com, [2]abhi.tiwari23@gmail.com

Abstract : Mining frequent item sets is to identify the sets of items that appear frequently in transactions in a database. The frequency of an item set is measured with the support of the item set, means the number of transactions containing the. Mining of frequent item sets only takes the presence and absence of items into account. Other information about items is not considered, such as the independent utility of an item and the context utility of an item in a transaction. Typically, in a supermarket database, each item has a distinct price/profit, and each item in a transaction is associated with a distinct count which means the quantity of the item one bought. In this paper we present a study over various high utility item mining method and techniques.

Keywords: Pattern mining, Database, Utility pattern, Transaction.

I. INTRODUCTION

The use of computer and online applications data are rapidly grown. New techniques of database facilitate the storage and usage of this huge data for business corporations, governments, and scientific organizations. From the large dataset it is difficult to obtain valuable information. Frequent pattern mining is the one the most commonly used data mining techniques in real life applications. In the Frequent item set mining the main problem is that the occurrence of each item in a transaction is represented by a binary value without considering its quantity or weight such as price or profit. Quantity and weight are significant for addressing real world decision problems that require maximizing the utility in an organization.

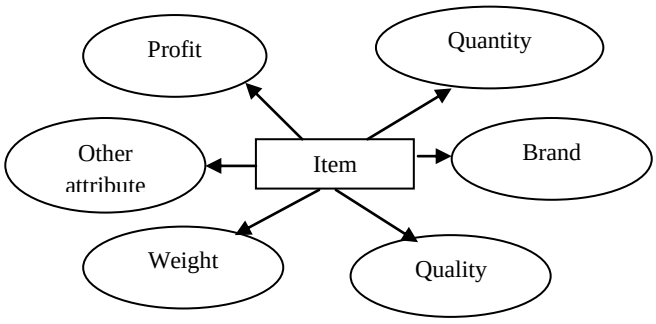

Figure 1. Attribute of an item

II. BASIC TERMINOLOGY

Utility item set mining has various terminologies.

A. Utility of an item - It does not reveal the utility of an item set which can be measured in terms of cost, profit, or other expressions of user preference.

B. High utility - The utility mining problem is to discover all item sets in a transaction database D with utility values higher than the minutil threshold, given a utility table UT.

C. Transaction utility - Reflects the utility in a transaction database. The quantity sold values are the transaction utility values of the items in each transaction.

D. The external utility - The external utility value of an item is a numerical value reflects the utility per item that is independent of transaction.

III. UTILITY PATTERN MINING TECHNIQUES

There are three basis approaches that are commonly used for mining high utility pattern, these include mathematical model based approach, and apriori based approach and UP tree based approach.

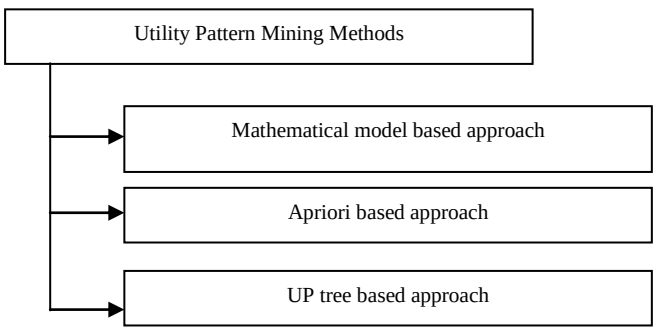

Figure2. Utility pattern mining Techniques

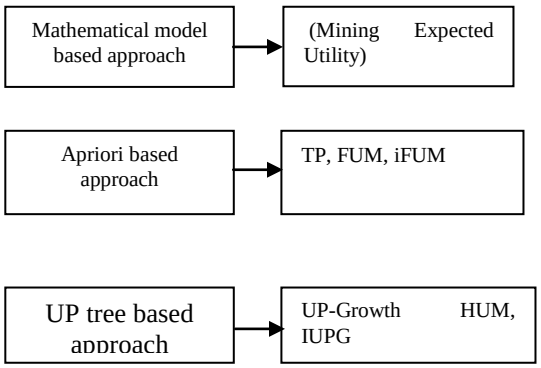

Figure 3. Methods based on techniques

IV. LITERATURE REVIEW

In 2005 Ying Liu Wei-keng Liao Alok Choudhary proposed "A Fast High Utility Item sets Mining Algorithm". They presented a Two-Phase algorithm to efficiently prune down the number of candidates and can precisely obtain the complete set of high utility item sets. In the first phase, they proposed a model that applies the "transaction-weighted downward closure property" on the search space to expedite the identification of candidates. In the second phase, one extra database scan is performed to identify the high utility item sets. They also parallelize our algorithm on shared memory multi-process architecture using Common Count Partitioned Database (CCPD) strategy. The proposed algorithm requires fewer database scans, less memory space and less computational cost. The accuracy, effectiveness and scalability of the proposed algorithm are demonstrated using both real and synthetic data on shared memory parallel machines. Another important feature is that Two-Phase algorithm can easily handle very large databases for which other existing algorithms are infeasible [1].

In 2006 Jia-Dong Ren and Xiao-Lei Zhou Proposed "A New Incremental Updating Algorithm for Mining Sequential Patterns" They discussed how to maintain discovered sequential patterns when some information is deleted from a sequence database. They proposed new algorithm, called MA_D (Maintenance Algorithm when Deleting some information), is presented in order to deal with the maintenance of sequential patterns mining resulted from the updating of database and the algorithm makes full use of the information obtained from previous mining results to cut down the cost of finding new sequential patterns in an updated database [2].

In 2007 Alva Erwin, Raj P. Gopalan & N.R. Achuthan proposed "A Bottom-Up Projection Based Algorithm for Mining High Utility Item sets". They proposed a new algorithm called CTU-PRO that mines high utility item sets by bottom up traversal of a compressed utility pattern (CUP) tree. They tested proposed algorithm on several sparse and dense data sets, comparing it with the recent algorithms for High Utility Item set Mining and the results show that our algorithm works more efficiently [3].

In 2008 Alva Erwin1, Raj P. Gopalan & N.R. Achuthan "Efficient Mining of High Utility Item sets from Large Datasets" We propose an algorithm that uses TWU with pattern growth based on a compact utility pattern tree data structure. Our algorithm implements a parallel projection scheme to use disk storage when the main memory is inadequate for dealing with large datasets. Experimental evaluation shows that our algorithm is more efficient compared to previous algorithms and can mine larger datasets of both dense and sparse data containing long patterns[4].

In 2009 Chowdhury Farhan Ahmed, Syed Khairuzzaman Tanbeer, proposed "An Efficient Candidate Pruning Technique for High Utility Pattern Mining". They proposed a novel tree-based candidate pruning technique HUC-Prune (high utility candidates prune) to efficiently mine high utility patterns without level-wise candidate generation-and-test. It exploits a pattern growth mining approach and needs maximum three database scans in contrast to several database scans of the existing algorithms. From the experimental results they show that proposed technique is very efficient for high utility pattern mining and it outperforms the existing algorithms [5].

In 2010 Vincent S. Tseng, Cheng-Wei Wu, Bai-En Shie and Philip S. Yu proposed " UP-Growth: An Efficient Algorithm for High Utility Item set Mining". They proposed an efficient algorithm, namely UP-Growth (Utility Pattern Growth), for mining high utility item sets with a set of techniques for pruning candidate item sets. The information of high utility item sets is maintained in a special data structure named UP-Tree (Utility Pattern Tree) such that the candidate itemsets can be generated efficiently with only two scans of the database. They have proposed an efficient algorithm named UP-Growth for mining high utility item sets from transaction databases. A data structure named UP-Tree is proposed for maintaining the information of high utility item sets. Hence, the potential high utility item sets can be efficiently generated from the UP-Tree with only two scans of the database [6].

In 2011 S. Kannimuthu Dr. K. Premalatha proposed "iFUM - Improved Fast Utility Mining".. They proposed system we made a significant improvement in FUM algorithm to make the system faster than FUM. They proposed the improved version of FUM algorithm, iFUM for mining all High

Utility Item sets. The proposed algorithm is compared with existing popular algorithms like UMining and FUM by using IBM synthetic data set. From the experimental result they show that iFUM algorithm is faster than other existing algorithms [7].

In 2012 Adinarayana reddy B & Orinivasa Rao proposed "An Improved UP-Growth High Utility Itemset Mining". They proposed modified algorithm UP-Tree (Utility Pattern Tree) which scans database only twice to obtain candidate items and manage them in an efficient data structured way. Applying UP-Tree to the UP-Growth takes more execution time for Phase II. The proposed algorithm aiming to reduce the execution time by effectively identifying high utility item sets[8].

In 2013 Arumugam P and Jose P proposed "Advance Mining Of High Utility Itemsets In Transactional Data". They proposed the novel algorithm for transactional high utility item set mining approach. This make to find association and correlation can generate less number of candidates. So the sales person can use this utility item set transaction for their stocks planning distributor/dealer month wise, product wise, model wise target setting[9].

In 2014 More Rani N & Anbhule Reshma V. proposed "Mining High Utility Item sets From Transaction Database ". They proposed Efficient Algorithm for Mining High Utility Item sets From Transactional Database.. In first scan, Transaction Utility (TU) of each transaction is calculated. At the same time Transaction Weighted Utility (TWU) of each single item is also calculated. In second scan, transaction is inserted into UPTree. Proposed algorithm, not only reduce number of candidate item sets but also work efficiently when database contains lot's of long transactions[10].

In 2015 Prashant V. Barhate & S. R. Chaudhari proposed" Efficient High Utility Itemset Mining using Utility Information Record". They proposed a novel data structure, utility information record, and developed an efficient algorithm, EHUI, for high utility itemset mining. Utility information record provides not only utility information about itemsets but also important pruning information for EHUI [11].

In 2016 Divvela. Srinivasa Rao, & V. Sucharita "Techniques for Mining High Utility Item sets from Transactional Databases". They proposed two algorithms named UP- Growth and UP-Growth+ for mining high utility item sets from transaction databases. A data structure named Comparison results show that the strategies considerably improved performance by reducing both the search space and the number of candidates [12].

V. CONCLUSION

We proposed study of various high utility patterns mining method some of them are based on mathematical model some them are apriori based and some then are UP tree based techniques.

Table 1. Fundamental approach

Algorithms	Approach
EMU	Utility Bound Property
Two-Phase	Top Down
FUM	Top Down
iFUM	Bottom UP
UP-Growth	Linked list

Each and every method tries to generate high utility pattern efficiently. The main objective of each method is to minimize candidate's generation at each level, reduce number of data base scan and reduce complexity in term of space and memory.

REFERENCES

[1] Ying Liu Wei-keng Liao Alok Choudhary *"A Fast High Utility Item sets Mining Algorithm"* UBDM '05, August 21 2005, Chicago, Illinois, USA Copyright 2005 ACM.

[2] Jia-Dong Ren and Xiao-Lei Zhou *"A New Incremental Updating Algorithm for Mining Sequential Patterns"* Journal of Computer Science 2 (4): 318-321, 2006.

[3] Alva Erwin, Raj P. Gopalan & N.R. Achuthan, *"A Bottom-Up Projection Based Algorithm for Mining High Utility Itemsets"* 2007, Australian Computer Society, Inc. Workshop on Integrating AI and Data Mining (AIDM 2007).

[4] Alva Erwin1, Raj P. Gopalan, and N.R. Achuthan, *"Efficient Mining of High Utility Itemsets from Large Data sets"* T. Washio et al. (Eds.): PAKDD 2008, LNAI 5012. Springer.

[5] Chowdhury Farhan Ahmed, Syed Khairuzzaman Tanbeer *"An Efficient Candidate Pruning Technique for High Utility Pattern Mining"* PAKDD 2009, LNAI 5476, 2009 Springer..

[6] Vincent S. Tseng, Cheng-Wei Wu *"UP-Growth: An Efficient Algorithm for High Utility Item set Mining" KDD'10*, July 25–28, 2010, ACM.

[7] S. Kannimuthu &Dr. K. Premalathai *"FUM - Improved Fast Utility Mining"* International Journal of Computer Applications (0975 8887) Volume 27– No.11, August 2011.

[8] Adinarayanareddy B O Srinivasa Rao, *"An Improved UP-Growth High Utility Itemset Mining"* International Journal of Computer Applications (0975 – 8887) Volume 58– No.2, November 2012.

[9] Arumugam P and Jose P *" Advance Mining Of High Utility Item sets In Transactional Data"*. Opinion - International Journal of Business Management (e-ISSN: 2277-4637) Special Issue on Role of Statistics in Management and Allied Sciences Vol. 3 No. 2 Dec. 2013.

[10] More Rani N. &Anbhule Reshma V *"Mining High Utility Item sets From Transaction Database"* International Journal of Latest Trends in Engineering and Technology (IJLTET) Vol. 3 Issue 3 January 2014.

[11] Prashant V. Barhate & S. R. Chaudhari *"Efficient High Utility Item set Mining using Utility Information Record"* International Journal of Computer Applications (0975 – 8887) Volume 120 – No.4, June 2015.

[12] Divvela. Srinivasa Rao & V. Sucharita *"Techniques for Mining High Utility Itemsets from Transactional Databases"* Volume 6, Issue 3, March 2016 International Journal of Advanced Research in Computer Science and Software Engineering.

