ISSN No:-2456-2165

An Examination of High Utility Item Set Mining using Different Techniques

Milan N Gohel
Assistant Professor,
Information Technology Department,
Atmiya University, Rajkot, Gujarat, India

Abstract:- Finding high utility itemset from transaction databases refers to finding itemset that are profitable and useful. Frequent Itemset mining, which identifies often occurring itemsets, is expanded upon in Itemset Utility Mining. Recognising itemsets with utility values over a certain ven utility threshold is the aim of high utility itemset mining. The user-specified minimum support threshold value must be met for an itemset to be considered a high utility itemset; otherwise, it is treated as a low utility itemset. In this article, we give a literature review of the current state of research, as well as a look at different algorithms and their potential drawbacks for high utility dataset mining.

Keywords:- *Mining of association rules, frequent itemsets, and high utility itemsets.*

I. INTRODUCTION

The technique of extracting non-trivial, unknown, and potentially helpful information from sizable databases is known as data mining. An essential data mining method known as association rule mining (ARM) is used to identify patterns and rules between objects in a sizable database [1]. Identification of groups of items that frequently occur together is the aim of ARM, for instance in a market basket study. Two processes make up mining association rules: the first is producing frequent itemsets. Creating association rules is the second. The biggest difficulty in association rule mining is figuring out popular item sets. One of the crucial tasks in association rule mining is finding frequent itemset. The majority of studies have concentrated on how to create frequent itemsets because the solution to the second sub-problem is simple.

The itemsets that appear frequently in the transaction database are known as frequent itemsets. Finding all of the frequently occurring itemsets in a transaction database is the goal of frequent itemset mining. Additionally, products with high and low selling frequencies may have different profit margins. For instance, some regularly purchased commodities, like bread, milk, and pens, may have lower profit margins than more expensive, occasionally purchased items, like gold, platinum, and diamonds.. As a result, the requirement to locate the most valuable item sets or customers that contribute to the majority of the overall earnings in the real world retail database cannot be satisfied by detecting simply typical frequent patterns in a database. This provides inspiration to create a mining model to identify the consumers and item sets that account for the majority of the profit. A frequent itemset is one whose frequency support exceeds a minimal threshold determined by the user. [1]

II. MINING HIGH UTILITY ITEMS

Finding all itemsets with utility greater than a user-specified minimum utility value is the goal of the high utility itemset mining challenge. The utility of the itemset refers to the worth or profit each item in a database is associated with. In terms of profit, for instance, computer systems are more profitable than telephones.[2]

Utility is defined as an item's interest, profitability, or significance. In terms of cost, profit, or another user choice, utility is measured. Items in transaction databases can be useful in the following two ways:

(1) The relevance of things in transactions, termed internal utility (i), i.e. quantity, and (2) The importance of distinct items, called external utility (e), i.e. unit profit.

Utility of Itemset(U)= externalutility(e)* internalutility(i).

High utility mining is widely employed in a variety of applications, including cross-marketing in retail stores, online e-commerce management, internet click stream analysis, and identifying significant patterns in biological applications. [2]

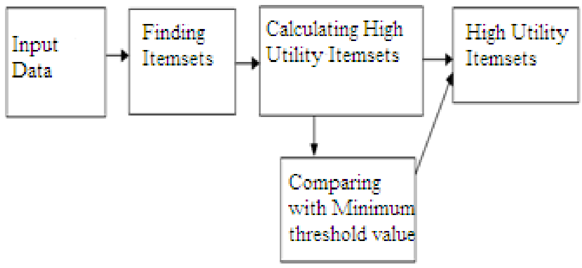

Fig. 1: Flow Diagram of HUIM

III. LITERATURESURVEY

A. Two phase algorithm

A Transaction-weighted Downward Closure Property is kept by this method [3]. Therefore, during the level-wise search, only the combinations of high transaction weighted utilisation itemsets are added to the candidate set at each level. Phase I never underestimates any item sets, even when it overestimates some low utility item sets. Phase II just requires one additional database search to filter the overestimated itemsets. Due to the level-wise strategy used in Two-Phase, numerous databases must be scanned and a large number of candidate itemsets must be generated.

ISSN No:-2456-2165

B. Compressed Transaction Utility (CTU-Mine)

The traditional candidate-generate-and-test strategy for selecting high utility itemsets is not appropriate for dense date sets, according to Erwin et al. in [4]. Their research suggests the unique algorithm CTU-Mine, which use the pattern growth method to mine high value itemsets. Yu et al. make a similar point. As a result of being column enumeration based and using an apriori-style candidate set creation and test technique, existing high utility mining algorithms are insufficient for datasets with high dimensions.

C. Temporal High utility itemset(THUI)

In order to efficiently and successfully mine temporal high utility itemsets from data streams, V.S. Tseng et al. presented THUI (Temporal High Utility Itemsets) -Mine in[3]. The innovative feature of THUI-Mine is its ability to efficiently discover temporal high utility itemsets by producing less temporal high transaction weighted utilisation 2 - itemsets, hence significantly reducing the mining time for all high utility itemsets in data streams. With less candidate itemsets, less memory, and less CPU I/O time, it is possible to effectively discover all temporal high utility itemsets across all time windows of data streams. This satisfies the essential criteria for time and space efficiency for data mining. The experimental findings demonstrate that, in contrast to previous algorithms, THUI-Mine can find temporal high utility itemsets with high performance and fewer candidate itemsets. In addition, it works well under huge databases in terms of execution speed. For mining temporal high utility itemsets in data streams, THUI-Mine is hence promising.

D. Utility patterngrowth(UP-growth)

The PHU (Potential High Utility) model is used in the recently proposed UP-Growth [5] (V.S. Tseng et al., 2010) algorithm to address the problem of producing a high number of candidates. The UP-Growth employs four strategies: DGU (Discarding Global Unpromising items), DGN (Decreasing Global Node utilities), DLU (Discarding Local Unpromising items), and DLN (Decreasing Local Node utilities) for lowering the number of candidate itemsets. Additionally, it builds the UP-Tree tree structure using two database scans and performs mining high utility itemsets. In other words, finding high utility itemsets requires three database scansTWU values for each item are tallied during the initial database scan. Items with TWU values below the user-specified minimal utility threshold are eliminated from each transaction during the second database scan. Additionally, the transactions are entered into the UP-Tree with the objects in the transactions organised in TWU descending order. DGU and DGN are used at this stage to reduce overestimated utilities. The UP-Tree is then used to produce high utility itemsets with DLU and DLN.

E. High utility itemset miner(HUI-Miner)

The HUI-Miner algorithm was proposed by Liu & Qu in 2012 [6]. It is a high utility itemset with a utility list list data structure. It starts by generating an initial utility list for potential itemsets of length 1. Then, using a pair of utility lists for itemsets of length k-1, HUI-Miner generates recursively a utility list for each itemset of length k. Each utility list for an itemset keeps the information about TIDs for all transactions containing the itemset, utility values of the item set, and utility values for mining high utility itemsets.

F. Faster High Utility itemset(FHM)

A super itemset of the itemset in the transactions, and the total utilities of the remaining items that can be added. HUI-Miner has the particular advantage of avoiding the expensive candidate creation and utility computation.

François Fournier-Viger's (2014) FHM algorithm was proposed [6]. The Hui-Miner Algorithm is extended by it. A depth-first search algorithm is used. To determine the precise utility of itemsets, it uses utility-lists. When mining high-utility itemsets using the utility list data structure, this approach incorporates a novel technique known as EUCP (Estimated Utility Co-occurrence Pruning) to minimise the amount of join operations. The transaction weighted utility (TWU) of all 2-itemsets is stored in the estimated utility co-occurrence structure (EUCS). During the initial database scans, it developed. Using a triangular matrix or hash map, EUCS is visualised. The EUCS structure has a modest memory footprint. Compared to HUI-Miner, FHM is up to 6 times faster.

G. Efficient high utility itemset (EFIM)

With the help of EFIM (Efficient high-utility Item set Mining), which presents a number of fresh concepts, it is possible to more quickly identify high-utility item sets in terms of both execution time and memory [7]. To more efficiently trim the search space, EFIM uses two upper-bounds called sub-tree utility and local utility. To determine these upper-bounds in linear time and space, it also presents a novel array-based utility counting technique called Fast Utility Counting. Merging transactions is clearly a desirable goal. To apply it effectively, though, is a major challenge. Sort the original database according to a new total order T on transactions to quickly locate identical transactions in O(n) time. Sorting is completed in a timely manner and just once. Due to transaction merging, the projected databases created by EFIM are frequently relatively small.

Sr. no.	Studies	Year	Dataset	Methodused	algorithm	Limitation
1	Ying Liu, Wei-kengLiao, AlokChoudhary	2005	Transactiondataset	Levelwiseapproach	Twophase	database multiple times and produce a large number of candidate itemsets
2	Alva Erwin, Raj P.Gopa lan, N.R.Achuthan	2007	Transactiondataset	Patterngrowthapproach	CompressedTransactionUtility(CTU-Mine)	Evaluation is difficult because of the tree structure.
3	VincentS.Tseng,Chun-Jung Chu,TyneLiang	2008	Transactiondataset	Patterngrowth	Temporal highutility itemsetmining(THUI)	large memory usage and a large number of false candidate itemset
4	VincentS.Tseng, Cheng-Wei Wu, Bai-En Shie, and Philip S.Yu	2010	Transaction dataset	Pattern growth	Utilitypattern growth(UP-growth)	Evaluation is difficult because of the tree structure.
5	MengchiLiu, JunfengQu	2012	Transaction dataset	Level wiseapproach	High utility itemsetminer(HUI-Miner)	It costs a lot of money to calculate the utility of an itemset joining utility list.
6	PhilippeFournier-Viger, Cheng-WeiWu, SouleymaneZida, Vincent S.Tseng	2014	Transaction dataset	Level wiseapproach	Fasterhigh utility itemsetmining(FHM)	Static database, significant memory use

IV. CONCLUSION

One of the most crucial tasks in data mining is association rule mining. There are many effective algorithms for association rule mining, which takes frequent item sets into account. Utility mining, which adds utility factors during itemset mining, is a new area of study in data mining. All facets of economic usefulness in data mining are covered by utility mining, which also aids in the identification of itemsets with high utility, such as profit. Mining high utility itemsets is particularly advantageous in a number of practical contexts. In this survey work, we provide numerous techniques for mining highly useful datasets, compare them with transaction datasets, and discuss the pros and cons of each methodology.

REFERENCES

[1.] U Kanimozhi, J K Kavitha, D Manjula, Mining High Utility Itemsets – A Recent Survey, International Journal of Scientific Engineering and Technology, Vo lumeNo.3IssueNo.11, pp: 1339-1344,2014.
[2.] Mayajoshi, Mansipatel, A Surveyon High Utility Itemset Mining Using Transaction Databases, International Journal of Computer Science andInformationTechnologies,Vol.5(6),7407-7410,2014.
[3.] Jyothi Pillai, O.P. Vyas, Overview Of Itemset Mining And its Application, International Journal of Computer Applications (0975 –8887),Volume5–No.11,August 2010.
[4.] Sudip Bhattacharya, Deepty Dubey, High Utility Itemset Mining, International Journal of Emerging Technology and Advanced Engineering, Volume2, Issue8, August 2012.
[5.] Hua-Fu Li, Hasin-Yug Huang, Yi-Cheng Chen, Yu-Jiun Liu, Suh-Yin Lee, Fast and Memory Efficient Mining of High Utility. Itemsets in Data Streams, IEEE International Conference on Data Mining, 2008.
[6.] Shekhar Patel B Madhushree, A Survey on Discovering High Utility Itemset Mining from Transactional Database, Information and Knowledge Management, Vol.5,No.12,2015.
[7.] R. Shyamala Devi, D. Shanthi, A Survey Mining High Utility Item Sets And Frequent Item Sets , International Journal of Innovative Research in Science, Engineering and Technology, Vol. 4, Issue 12, December 2015.
[8.] Smita R. Londhe, Rupali A. Mahajan, Bhagyashree J. Bhoyar, “Overview on Methods for Mining High Utility Itemset from Transactional Database”, International Journal of Scientific Engineering and Reaserach (IJSER), Volume 1 Issue 4, Deccember 2013.