Volume 3, Issue 12, December 2018 ISSN (online): 2456-0006

International Journal of Science Technology Management and Research

Available online at: www.ijstmr.com

Mining Utility Pattern from Large Dynamic Medicine Database

Sheetal Bhagat
M Tech (CSE) IV Sem.
Lord Krishna College of Technology
Indore M P. India
bhagatsheetal4@gmail.com

Vijay Kumar Verma
Asst. Professor CSE department
Lord Krishna College of Technology
Indore M P. India
Vijayvermaonline@gmail.com

Abstract: ***Data mining takes data as input and yields patterns& summaries as output that can be used in classification, clustering, association etc. The most significant tasks in data mining are the process of discovering different types of patterns. Several efficient algorithms have been developed for mining patterns. Utility considerations in data mining tasks are gaining popularity in recent years. Utility-based data mining is used to integrate utility considerations in both predictive and descriptive data mining tasks Utility mining is an extension of pattern mining. Utility can be weight, profit, cost, quantity etc. Utility mining is an important data mining technique for mining patterns based on some utility. Our algorithm for mining utility items from large databases not only reduce data base size during scanning of the but also reduce number of candidates and number of arithmetic calculation which provides a big advantage over the previous algorithms. In this paper we present a novel approach for mining high utility item set based on data compactions techniques***

Keywords: Cryptography, Huffman compression, XOR, private key, public key,

I. INTRODUCTION

Data mining takes data as input and yields patterns& summaries as output that can be used in classification, clustering, association etc. The most significant tasks in data mining are the process of discovering different types of patterns. Several efficient algorithms have been developed for mining patterns. Utility considerations in data mining tasks are gaining popularity in recent years. Utility-based data mining is used to integrate utility considerations in both predictive and descriptive data mining tasks. Utility of an item can be identified by several things some of them are shown in the figure.

.

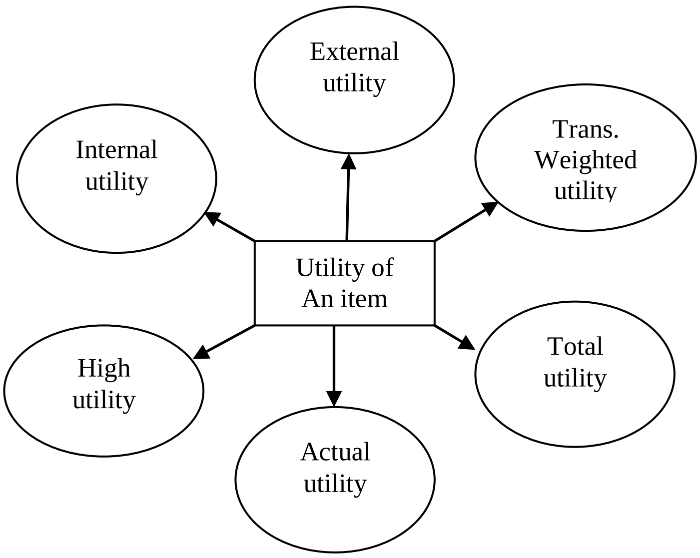

Figure 1-Terms related with utility

There are several terms used to denote utility in utility mining figure 7.1illustrates some of them. The utility of an item in the transaction is known as Internal utility, it is available in a transaction; External utility is not given in transaction but is given by user in the profit table, transaction weighted utility is calculated by transaction number, actual utility is the utility of the item set itself and finally total utility is nothing but the overall utility of the database.

II. UTILITY MINING MODEL

Utility mining is an extension of pattern mining. Utility can be weight, profit, cost, quantity etc. Utility mining is an important data mining technique for mining patterns based on some utility. Utility mining provides sufficient information about the product. In the past there are several algorithms have been developed for mining high utility items set. Each and every algorithm has some advantages and limitations. Efficiency is always a factor for improvement in the existing algorithm. Efficiency can measured in term of execution time, memory requirement or arithmetic complexity. In following section we present a novel approach for mining high utility item set based on data compactions techniques.

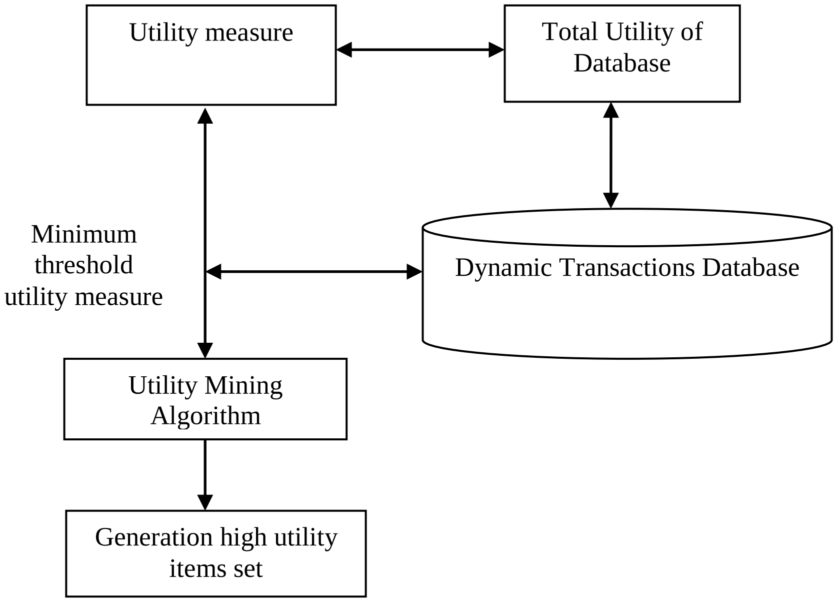

Figure-2 Utility mining models

III. LITERATURE REVIEW

In 2010 Vincent S. Tseng, Cheng-Wei Wu, Bai-En Shie, and Philip S. Yu proposed a data structure, named UP-Tree, and then describe a new algorithm, called UP-Growth, The framework of the UP-Growth: An Efficient Algorithm for High Utility Item set Mining method Proposed approach is not based on the traditional framework of transaction-weighted utilization mining model. UP-Tree facilitate the mining performance and avoid scanning original database repeatedly, proposed algorithm is a compact tree structure, called UP-Tree to maintain the information of transactions and high utility Item set s. The construction of UP-Tree can be performed with two scans of the original database. In the first scan of database, the transaction utility of each transaction is computed. At the same time, TWU of each single item is also accumulated. After scanning database once, items and their TWUs are obtained. By TWDC property, if the TWU of an item is less than minimum utility threshold, its supersets are unpromising to be high utility Item set s. The item is called unpromising items.

In 2011 S. Kannimuthu Dr. K. Premalatha S. Shankar proposed iFUM - Improved Fast Utility Mining. The core step of FUM algorithm is Combination Generator (T) which takes significant time to compute. In the existing system FUM combination generation is performed for Item set s and its subset without checking one important condition. Combination Generator (T) - Generate all possible combinations of Item set TFUM algorithm fails to check this condition so it generates the combinations for the already generated subset of the Item set s too, if it repeats in a later transaction of the input database. Proposed algorithm avoids these extra computations and enhances FUM efficiency.

In 2012 Mengchi Liu and JunfengQu proposed "Mining High Utility Item set s without Candidate Generation" High utility Item set s refer to the sets of items with high utility like profit in a database, and efficient mining of high utility Item set s plays a crucial role in many real life applications and is an important research issue in data mining area. To identify high utility Item set s, most existing algorithms first generate candidate Item set s by overestimating their utilities, and subsequently compute the exact utilities of these candidates. These algorithms incur the problem that a very large number of candidates are generated, but most of the candidates are found out to be not high utility after their exact utilities are computed. They proposed an algorithm, called HUI-Miner (High Utility Item set Miner), for high utility Item set mining. HUI-Miner uses a novel structure, called utility-list, to store both the utility information about an Item set and the heuristic information for pruning the search space of HUI-Miner. By avoiding the costly generation and utility computation of numerous candidates Item set s, HUI-Miner can efficiently mine high utility Item set s from the utility lists Constructed from a mined database.

In 2012 Cheng Wei Wu, Bai-En Shie, Philip S. Yu, Vincent S. Tseng "Mining Top-K High Utility Item set s" Mining high utility Item set s from databases is an emerging topic in data mining, which refers to the discovery of Item set s with utilities higher than a user-specified minimum utility threshold min_util. Although several studies have been carried out on this topic, setting an appropriate minimum utility threshold is a difficult problem for users. If min util is set too low, too many high utility Item set s will be generated, which may cause the mining algorithms to become inefficient or even run out of memory. On the other hand, if min util is set too high, no high utility Item set will be found. Setting appropriate minimum utility thresholds by trial and error is a tedious process for users. They solve this problem by proposing a new framework named top-k high utility Item set mining, where k is the desired number of high utility Item set s to be mined. An efficient algorithm named TKU (Top-K Utility Item set s mining) is proposed for mining such Item set s without setting min util. Several features were designed in TKU to solve the new challenges raised in this problem, like the absence of anti-monotone property and the requirement of lossless results. Moreover, TKU incorporates several novel strategies for pruning the search space to achieve high efficiency.

In 2013 Arumugam P and Jose Proposed "Advance Mining of High Utility Item set s in Transactional Data". The white good industry domain is a dynamic and unpredictable field. Several analysis and algorithms provide investors with some technical tools for managing their stocks and predicting their market field. But these techniques are not enough to produce all the discovery possibilities. The sales executives plan their yearly, month wise target and their historical analysis, data mining approach used extensively in the markets and help in association analysis. It is useful for discovering interesting relationships hidden in large datasets. The uncovered relationships can be represented in the form of association rules. Traditional Apriori algorithm takes more time, space and memory for candidate generation process. 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.

In 2014 D. Usha Nandini, Ezil Sam Leni, M. Maria Nimmy Proposed "Mining of High Utility Item set s from Transactional Databases". Efficient discovery of high utility Item set s from transactional databases crucial task in data mining. UP-Growth and UP-Growth+ algorithms are proposed for mining high utility Item set s. They also proposed a compact tree structure, called Utility pattern tree (UP-Tree) and it maintains the information of high utility Item set s. Previously we proposed FP-Growth algorithm for mining only large number of frequent Item set s, but not generate the high utility Item set s. They have the issue of producing large number of candidate Item set s and probably it degrades mining performance in terms of speed and space requirement. However, our previous study needs more space and execution time. Many algorithms are used to show the performance of UP-Growth and UP-Growth+. UP-Growth and UP-Growth+ becomes more efficient since database contain long transactions and generate fewer number of candidates than FP-Growth. The experimental results and comparison validate its effectiveness.

In 2014 G. Saranya and A .Deepak kumar proposed "Implementation of Efficient Algorithm for Mining High Utility Item set s in Distributed and Dynamic Database" Association Rule Mining (ARM) is finding out the frequent Item set s or patterns among the existing items from the given database. High Utility Pattern Mining has become the recent research with respect to data mining. The proposed work is High Utility Pattern for distributed and dynamic database. The traditional method of mining frequent Item set mining embrace that the data is astride and sedentary, which impose extreme communication overhead when the data is distributed, and they waste calculation resources when the data is dynamic. To overcome this, Utility Pattern Mining Algorithm is proposed, in which Item set s are maintained in a tree based data structure, called as Utility Pattern Tree, and it generates the Item set without store the entire database, and has sparse communication overhead when mining with respect to distributed and dynamic databases. A quick update incremental algorithm is used which scans only the incremental database as well as collects only the support count of newly generated frequent Item set s. Incremental Mining Algorithm not only includes new Item set into a tree but also discard the infrequent Item set from a utility pattern tree structure. Hence it provides faster execution, minimal communication and cost when compared to the existing methods.

In 2015 Junqiang Liu et al proposed "Mining High Utility Patterns in One Phase without Generating Candidates" They proposed a novel algorithm that finds high utility patterns in a single phase without generating candidates. The novelties lie in a high utility pattern growth approach, a look ahead strategy, and a linear data structure. Proposed approach prune search space by utility upper bounding. They ahead to identify high utility patterns without enumeration by a closure property and a singleton property. Linear data structure enables us to compute a tight bound for powerful pruning and to directly identify high utility patterns in an efficient and scalable way, which targets the root cause with prior algorithms. Extensive experiments on sparse and dense, synthetic and real world data suggest that our algorithm is up to 1 to 3 orders of magnitude more efficient and is more scalable than the state-of-the-art algorithms.

In 2016 Dr.R.A.Roseline et al proposed "Efficient Algorithms for Mining High Utility Item sets from Transactional Databases". A number of candidate item sets that are available in the database require more memory consumption and time for processing along with it requires more computational overhead. This problem is resolved by introducing the tree based search paradigm namely UPGrowth and UPGrowth+ that is used to construct a utility tree by discovering only the high potential utility item sets from the transactional database. The tree based methodology improves the accuracy level of retrieving the most utilized potential utility item sets. The computation overhead of tree construction increases in the case of presence of most potential utility item sets. This is resolved in this research work by enhancing the UPGrowth tree construction by integrating the random hashing technique. This approach l allocates memory for every candidate item sets by finding the amount of memory that is required and calculates the node count utility along the route path to allocate the memory space required by them. Experimental tests conducted proves that the proposed approach provides efficient handling of potential utility item sets than the existing approach in terms of improved time complexity and space complexity.

In 2017 P.Sri Varshini et al proposed "Mining High Utility Pattern in One Phase without Candidate Generation using up Growth+ Algorithm" The performance holdup of these algorithms is the generate more no of candidates item sets and increasing of the number of long transaction item sets it cannot work minimum utility threshold, the situation may become worse and also creating more no tree. To overcome these problems, propose an efficient algorithm, namely UP-Growth (Utility Pattern Growth), for mining high utility item sets with pruning techniques for pruning candidate item sets. The information of high utility item sets is stored in a special data structure named UP-Tree (Utility Pattern Tree) such that the candidate item sets can be generated with only two scans of the database. The performance of UP growth+ was evaluated in comparison with the state-of-the-art algorithms on different types of datasets. The experimental results show that UP growth+ outperforms other algorithms in terms of both execution time and memory space under minimum utility threshold is, the more observable its advantage will be it can achieve the level of about two orders of magnitude faster than the state-of-the eart algorithms on dense dataset, and more than one order of magnitude on sparse datasets

IV. PROPOSED METHOD

The proposed method is divided into following steps to mine the utility patterns. For each transaction in database, the following sub steps are performed.

1. Calculate the transaction utility of the transaction.
2. Calculate the transaction-weighted utility of each item as the summation of the transaction utility values of the transactions which include the item.
3 Check whether the transaction-weighted utility of an item is larger than or equal to the minimum utility threshold if it is, put it in the set of high transaction weighted utilization itemsets.
4 Use data compaction techniques to reduce the database size
5 Calculate the transaction utility of each modified transaction.
6 Scan the set of modified transactions to and the transaction-weighted utility
Check whether the transaction-weighted utility of an item is larger than or equal to the minimum utility threshold.

V. EXPERIMENTAL ANLYSIS

The worth of an algorithm cannot be determined if it is not implemented. Thus we also have implemented the above mentioned method for this we have used VB dot net 2010 as front end and SQL server as back end for implementation. All the experiments were performed on an i3 4M Cache, 2.50 GHz Intel PC machine with 2 gigabyte main memory, running Microsoft Windows 7. To evaluate the performance Real life dataset is used. We have implemented three algorithms first one Two Phase (TP), iFUM (Improved Fast Utility Mining) and our method. We have taken 25 different items and records of data about 1000 customers.

Table 1 Show level numbers and number of candidate generated in execution in algorithm Two Phase and proposed approach . Number of items taken 5 and number of transaction are 10.

Level No	Two Phase	Proposed Approach
1	5	5
2	10	6
3	10	3
Total	25	14

Table 1 Level number and number of candidate generated in execution

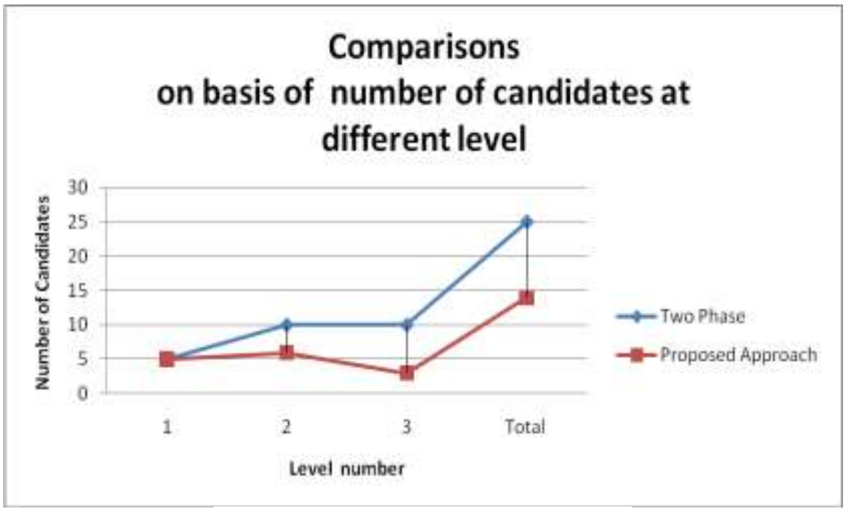

Figure 3 Level number with candidates

CONCLUSION

The complexity factor for frequent pattern mining algorithm includes:
1. Execution time and I/O cost.
2. Arithmetic complexity and Number of Candidates generation.
3. Memory used.
4. Accuracy of the algorithm.
5. Scalability in term of number of records.

We test TP and Proposed approach using different parameter like threshold value , candidate generation ,number of record ,memory used and accuracy of the algorithm .Form the experiment it clear that proposed algorithm perform well and compared to the TP. From Graph it is clear that proposed method outperforms others by reducing arithmetic complexity, number of candidate generation and memory used to execute algorithm. This investigation has presented a new approach, for utility item sets mining

REFERENCE

1. Yao, H., Hamilton, H.J., Buzz, C.J.: A Foundational Approach to Mining Itemset Utilities from Databases. In: 4th SIAM International Conference on Data Mining. Florida USA (2004)
2. Y. Liu, W. Liao, and A. Choudhary. A fast high-utility itemsets mining algorithm. In Proc. of the Utility-Based Data Mining Workshop, 2005.
3. H. Yao, H. J. Hamilton, L. Geng, A unified framework for utility-based measures for mining itemsets. In Proc. of ACM SIGKDD 2nd Workshop on Utility-Based Data Mining, pp. 28-37, 2006.
4. Erwin, A., Gopalan, R.P., Achuthan, N.R.: A Bottom-Up Projection Based Algorithm for Mining High Utility Itemsets. In: International Workshop on Integrating AI and Data Mining. Gold Coast, Australia (2007)
5. H.-F. Li, H.-Y. Huang, Y.-C. Chen, Y.-J. Liu, S.-Y. Lee. Fast and Memory Efficient Mining of High Utility Itemsets in Data Streams. In Proc. of the 8th IEEE Int'l Conf. on Data Mining, pp. 881-886, 2008.
6. A. Erwin, R. P. Gopalan and N. R. Achuthan. Efficient Mining of High-utility Itemsets from Large Datasets. In PAKDD 2008, LNAI 5012, pp. 554-561, 2008. Vid Podpeˇcan A Fast Algorithm for Mining Utility-Frequent"Itemsets All content following this page was uploaded by Nada Lavrac on 27 November 2014
7. V. S. Tseng, C.-W. Wu, B.-E. Shie, and P. S. Yu. UP-Growth: an efficient algorithm for high utility itemset mining. In Proc. of Int'l Conf. on ACM SIGKDD, pp. 253–262, 2010.
8. S. Kannimuthu , Dr. K. Premalatha iFUM - Improved Fast Utility Mining International Journal of Computer Applications (0975 – 8887) Volume 27– No.11, August 2011.
9. Cheng Wei Wu1, Bai-En Shie Philip S. Yu, Mining Top-K High Utility Itemsets. KDD'12, August 12–16, 2012, Beijing, China. Copyright 2012 ACM 978-1-4503- 1462-6/12/08.
10. Arumugam and Jose " Advance Mining Of High Utility Itemsets In Transactional and p-ISSN: 2231–5470) Special Issue on Role of Statistics in Management and Allied Sciences Vol. 3 No. 2 Dec. 2013,.
11. More Rani N. "Mining High Utility Item sets From Transaction Database" International Journal of Latest Trends in Engineering and Technology (IJLTET) Vol. 3 Issue 3 January 2014.
12. Junqiang Jimmy Liu Mining High Utility Patterns in One Phase without Generating Candidates IEEE Transactions On Knowledge And Data Engineering, Vol. 27, For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below. Digital Object Identifier no. 10.1109/TKDE.2015.2510012.
13. Dr.R.A.Roseline, "Efficient Algorithms for Mining High Utility Item sets from Transactional Databases" International Journal of Innovative Research in Computer and Communication Engineering (A High Impact Factor, Monthly, Peer Reviewed Journal) Vol. 4, Issue 1, January 2016.
14. P.Sri Varshini Mining High Utility Pattern in One Phase without Candidate Generation using up Growth+ Algorithm" International Journal of Engineering Trends and Technology (IJETT) – Volume45 Number 4 – March 2017 ISSN: 2231-5381