

A Study on Mining High Utility Item sets for Promoting Business Activities

^[1]R. Sheeba Mary Ananthi, ^[2] Dr. V. Joseph Peter

^[1] Research Scholar, ^[2] Associate Professor

^{[1][2]} Department of Computer Science, Kamaraj College, Thoothukudi.

Abstract: In recent era, High Utility Itemset Mining (HUIM) is an emerging critical research topic. In traditional approach, the items which occur frequently together are extracted from a database. But the frequency of Itemset is not sufficient to reflect the actual utility. Utility mining is an extension of frequent Itemset mining by considering the utility of an item. Utility Mining is the process of discovering all item sets whose utility values are equal to or greater than the user specified threshold in a transaction database. Utility Mining covers all aspects of economic utility in data mining and helps in direction of itemset having high utility. The main objective of high utility itemset mining is to find the itemset having maximum utility values. We can extract the high utility from rare itemsets, irregular occurrence, from different discount strategies. In this paper, we present a various algorithms for High Utility Mining to promote business activities.

Index Terms — high utility mining, on-shelf time, rare itemset, utility mining.

I. INTRODUCTION

Data mining is a process to find out interesting pattern, correlations and information from database which is used to make decision making. It is used in wide range of area to predict future trends and behavior analysis. Market Basket Analysis is an important component of analytical system in retail organizations to determine the placements of goods, designing sales promotions for different segments of customers to improve customer satisfaction and hence the profit of the super market. Discovering useful patterns hidden in a database plays an essential role in several data mining tasks, such as frequent pattern mining, weighted frequent pattern mining and high utility pattern mining. The need for considering the economic utility of the itemsets in the data mining process has gained wide acceptance. Association Rule Mining is one of the most widely used techniques in data mining and knowledge discovery. The ARM mostly focused on frequent itemset mining. It is widely used in Market Basket Analysis. It considers only the occurrence of items while the other factors such as price, profit are neglected. The frequency of itemset is not sufficient to identify highly profitable items. In recent decades mining high utility itemsets has become interesting research topic in data mining.

II. TECHNIQUES

A. Utility Mining

The identification of itemsets with high utilities is called as utility mining. Utility is a measure of how useful or profitable an itemset is and it can be measured in terms of

cost, profit or other expressions of user preferences. An itemset is called a high utility itemset if its utility is no less than a user specified minimum utility threshold. We can extract high utility itemset from the transaction database using various techniques they are High Utility from Rare Itemsets, High Utility from Irregular occurrence Itemsets, High Utility Itemsets with Discount strategy, and High Utility Itemsets from On-Shelf time of product. To promote business activities, we should classify the High Utility Itemsets from the total utilities.

B. HUI from Rare and Irregular Occurrence

Now-a-days researches are going on high utility itemsets with occurrence behaviors investigations. Consider only regular occurrence behavior which may not sufficient in some applications, so that the task of mining high utility itemsets with irregular itemsets that is not periodic. It will help to know which products give high profit even it has not been purchased them together and it helps to manage warehouse in order to avoid depreciations of the products.

Frequent itemsets may not generate a very high profit but rare itemsets provide very useful information in the business databases. Rare itemsets are the itemsets that occur infrequently in the transaction data sets. Rare itemsets provide useful information in different decision making domains in business transactions. For example, in a super market microwave oven or LED TV is sold rarely while compared to frequently moving itemsets like bread and jam. But the former transaction provides high utility than the later one. All the itemsets which has value lesser than the support, will considered as infrequent itemset or rare itemset. Items which are below from the min support threshold will be considered as HURI.

UPRI-algorithm has been used to effectively mine High Utility Rare Itemsets from transactional database. After Identifications of high utility rare itemsets, marketers can do the promotion or advertising of such itemsets to increase the overall profit of the business.

A Rare Itemset Miner Algorithm (ARIMA) generates the set of all rare itemsets splits into two sets, the set of rare itemsets having a zero support and the set of itemsets with non-zero support. If an itemset is rare then any extension of that itemset will result a rare itemsets. The advantage of ARIMA can find the rare items without zero itemsets, it depends on two thresholds.

HUIIM – Algorithm is used to mine HUIIS. HUIIM is an efficient single pass algorithm based on the use of utility list structure used to capture essential information of each item occurring in each transaction. It gives a complete set of high utility irregular itemsets from collection of information in the scanned data.

C. HUI with Discount strategies

Discovering high utility items under different discount strategies are highly desirable because discounting is done in most business. There are different types of discount strategies. They are Discount levels varies from 0% to 100% Giving free units if the customer buys huge level of units Giving one unit free for buying a product

HUID and HUI-DTP are the algorithms commonly used for extracting high utility itemsets with different discount strategies. HUID is an efficient algorithm used to mine high utility items using discounts strategies. It is a two phase HUID algorithm which is used to mine HUI with UBTWDC property. HUI-DTP is a two phase algorithm to mine the complete set of high utility itemsets based on a novel downward closure and a vertical TID-List structure.

D. HUI from On-Shelf time of product

In real world applications not all the products in stores are always shelf for sale. Some utility patterns may high utility itemsets within their selling time period, but they are not high in all the time periods. High Utility On-Shelf Itemsets (HUOI) which consider not only individual profits and quantities of products in transactions but also actual on-shelf time periods of products. The whole time interval to be analyzed is split into several time period and any itemsets with a high utility value within the union of

all its ON-Shelf time periods is thought as a high On-Shelf utility itemsets. The On-Shelf utility itemset is the one with its sum of utilities in all On-shelf periods larger than or equal to a threshold.

Two-Phased mining algorithm is efficient to mine High On-shelf utility itemsets. In the first phase, it finds all the possible candidate On-shelf utility itemsets with each time from a database. In the second phase, it scans the database again to find the actual utility values within the union of all its time period and give High Utility On-shelf Itemsets (HUOI). Another algorithm KOSHU (fast –K on-shelf high utility itemset miner) is used to mine the top-k HOU's, while considering on shelf time periods of items, and items having positive and /or negative unit profits.

III.CONCLUSION

In this paper we have listed the various high utility mining algorithms to efficiently discover the high utility from rare itemsets, irregular occurrence, from different discount strategies for the desired itemsets from a database to promote business activities. In future work, we are interested in exploring other interesting problems involving utility in itemset mining.

REFERENCES

- J. Pillai, O.P. Vyas, B. Mulgrew, "User centric approach to itemset utility mining in Market Basket Analysis," International Journal of Computer Sciences and Engineering, pp.393-400, July 2011.
- J. Pillai, O.P. Vyas, "Overview of Itemset Utility Mining and its Applications," International Journal of Computer Applications, pp. 9–13, 2010.
- K. Amphawan, P. Lenca, A. Jitpattanakul, and A. Surarerks, "Mining high utility itemsets with regular occurrence," Journal of ICT Research and Applications, vol. 10, no. 2, pp. 153–176, 2016.
- V. S. Tseng, B.E. Shie, C.W. Wu, and P. S. Yu, "Efficient algorithms for mining high utility itemsets from transactional databases," IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 8, pp. 1772–1786, 2013.
- 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 Transactions on Knowledge and Data Engineering, vol. 21, no. 12, pp. 1708–1721, 2009.

P. Fournier-Viger, “FHN: Efficient mining of high-utility itemsets within negative unit profits,” Advanced Data Mining and Applications, pp. 16-29, 2014.

H. Yao and H. J. Hamilton, “Mining itemset utilities from transaction databases,” Data and Knowledge Engineering, vol. 59, no. 3, pp. 603–626, 2006.

C.W. Lin, T.P. Hong, and W.H. Lu, “An effective tree structure for mining high utility itemsets,” Expert Systems with Applications, vol. 38, no. 6, pp. 7419–7424, 2011.

C.Y. Chang, M.S. Chen, and C.H. Lee, “Mining general temporal association rules for items with different exhibition periods,” The Third IEEE International Conference on Data Mining, 2002.

T. P. Hong, C. Y. Wang, and Y. H. Tao, “A new incremental data mining algorithm using pre-large itemsets,” Intelligent Data Analysis, vol. 5, no. 2, pp. 111–129, 2001.

Thu-lanDam, Kenlili, philippe Fournier-viger, Quang-huyduong, “An efficient algorithm for mining top –K on-shelf high utility itemsets”.