A Review on High Utility Itemset Algorithms

M. Suleman basha, K. Nageswara reddy, P. Arun babu

Abstract*: At present situation, Utility Mining is most advanced research domain in data mining and utility mining determines the High-Utility Itemset (HUI) in data sets. The utility can be any user describes option and to allot this utility value is based on the component of the dataset. The utility can be is used in many applications like marketing. The HUI Mining is targeted for recognizing various itemsets consists of utility values which are in a desired threshold limits. As well as, the utility values are greater than the minimum user pre-requisite supported threshold value; then the itemset is treated as a low-utility itemset. In this paper, various algorithms and its performance factors are proposed and the effectiveness is evaluated based on space & time for HUI mining.

Index Terms*: Faster high utility itemset, High-Utility Itemset, High utility itemset miner, HUP-Miner, Patter mining.

I. INTRODUCTION

The HUI mining is advancement of frequent-pattern mining used in various mining methodologies [3]. The frequent pattern technique is entailed for finding the various frequent patterns in transaction database. For these mining systems, several algorithms are proposed like, apriori, FPGrowth, LCM, etc. The regular association rules only examine the existence of desired items in a transaction; the utility functions of items are un-recognized with the aid of HUIM determines above-mentioned issues. The HUI mining is used to evaluate the all itemset values which have greater user-utility defined value of minimum utility range. The values of acquired profit correlation with each item in a database sequence are called as utility of defined itemset.

For example, the computer is highly profitable over the telephone. The utility determine as Interestingness, profitability or importance of item. The Utility is concerning to cost-profit and/or other user option. The Utility of certain items in database is implicated with respective aspects:

(1) The significance of defined items called as utility in external range (e).

(2) The significance of defined items in a transaction, named as utility in internal region (i).

Utility of Itemset (U) = external utility (e) * internal utility (i).

In many operations High utility mining are widely used [2], like e-commerce, web-stream analysis, cross-marketing in retail shops and acquiring the significant pattern in bio-medical applications.

Revised Manuscript Received on December 22, 2018.

M. Suleman Basha, Department of Information Technology, RGM College of Engineering & Technology, Nandayal (Andhra Pradesh), INDIA.

K. Nageswara Reddy, Department of Information Technology, RGM College of Engineering & Technology, Nandayal (Andhra Pradesh), INDIA.

P. Arun Babu, Department of Information Technology, RGM College of Engineering & Technology, Nandayal (Andhra Pradesh), INDIA.

II. HIGH UTILITY ITESMSET ALGORITHMS

R. Chan et. al., [6], explores the algorithms of high-utility itemset mining over the non-binary databases with items of distinct values. The main intension of HUI's mining is to implement the itemsets which brings the prominent profit to the users, thus they are treated as non-frequent itemsets.

A. High Utility Itemset Miner (HUI-Mining)

Liu & Qu [5] explores the HUI-mining algorithm, it is very high-utility itemset with a listed-data structures named as utility list. Generally, it first develops a basic utility-list for defined itemsets with a range of length-1 for favourable items. Then, HUI-Mining is designed recursively based on utility-list for every itemset with a length (k) by using a pair of utility-lists for itemsets with a length of (k-1). For high-range utility itemsets, every utility list for an itemset maintains the TIDs information of all transactions which comprises of utility values of every itemsets, sum of remaining utility item sets that can be incorporate to super itemsets of very itemset in a transaction. The favourable merit of HUI-Mining system is highly avoiding the cost-effective utility computation and generation.

B. HUP-Miner

The HUP-Miner [4] extends the HUI-Miner algorithm [5], [9] for defining the high-range utility itemsets in a transaction database encloses the utility information. It supports the basic idea of database separation and pruning-scheme named LA-prune. The major defect of HUP-Miner is that the user pre-requisite the set of extra specifications, which is defined based on number of partitions. Furthermore, HUP-Miner is more flexible and faster than HUI-Miner but very slower than FHM. The high-utility itemset mining is used in various applications like discovering the item groups in transactions of stores that produce good & more profitable. Any database comprising of information related to utility in a database [10], where the items can have unit prices and quantities. Hence, the respective algorithms are often introduced in the context of market-basket analysis, there exist several applications.

C. Faster High utility itemset Miner (FHM)

The Philippe-Fournier-Viger proposes the FHM algorithm [1], it improves the HUI-algorithm. It is operated as Depth-first search algorithm and evaluates the exact utility of itemsets. This algorithm is consolidated as novel scheme named as Estimated Utility Co-occurrence Pruning (EUCP) to minimize the number of join operations when mining the high-utility itemsets by using utility-list data sequence. Assessed EUCS system [11], [12] furnishes the transaction in weighted utility [TWU) of all 2-itemsets.

It is designed during the initial database scans, EUCS signified as a triangular matrix or hash-map. The fastness of FHM is upto 6 times of HUI-Miner.

III. EXPERIMENTAL RESULTS

We evaluated the performance of FHM, HUI-Miner & HUP-Miner on the bench mark datasets i.e., chess, Foodmart & retail utility. Experiments were performed on a computer with a 7th generation 64-bit Core i3 processor running Windows 10 and 4 GB of free RAM. We compared the performance of FHM with the state-of-the-art algorithm HUI-Mine and HUP-Miner for high-utility itemset mining. In this paper Fig 1, Fig.2 & Fig.3 shows the Time Comparisons of FHM, HUI & HUP-Miner and Fig.4, Fig.5 & Fig.6 shows the Space/Memory Comparison of FHM, HUI-Miner & HUP-Miner. The minimum utility from 650000 to 70000 in the chess dataset. For Food mart dataset minimum utility from 10000 to 20000 and for retail_uility data set minimum utility from 55000 to 60000.

A. Time Comparisons

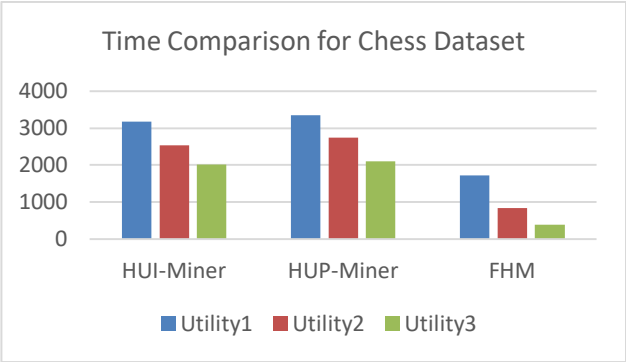

Fig. 1: Comparison of Execution time in ms for Chess Dataset

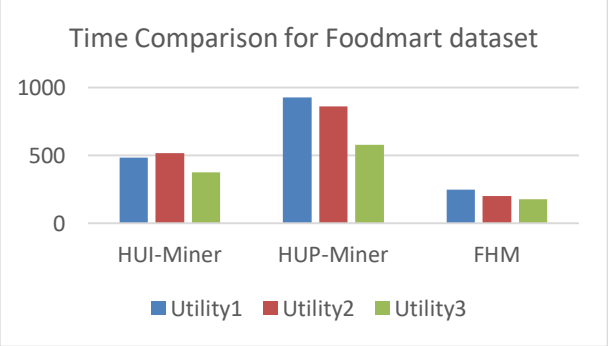

Fig. 2: Comparison of Execution time in ms for Foodmart Dataset

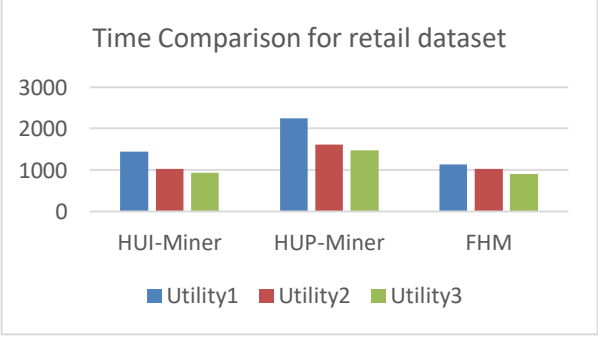

Fig. 3: Comparison of Execution time in ms for retail Dataset

B. Memory or space Comparisons

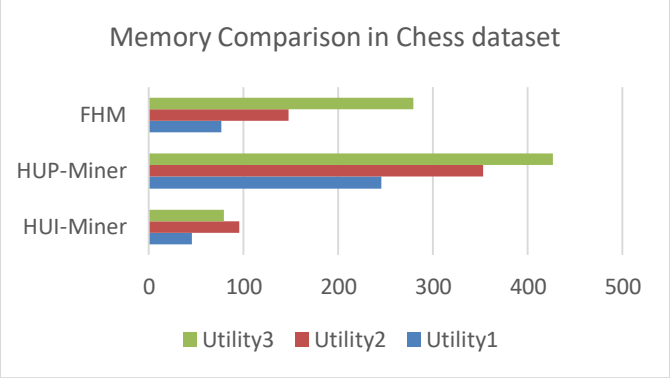

Fig. 4: Comparison of memory space occupied in mb in Chess Dataset

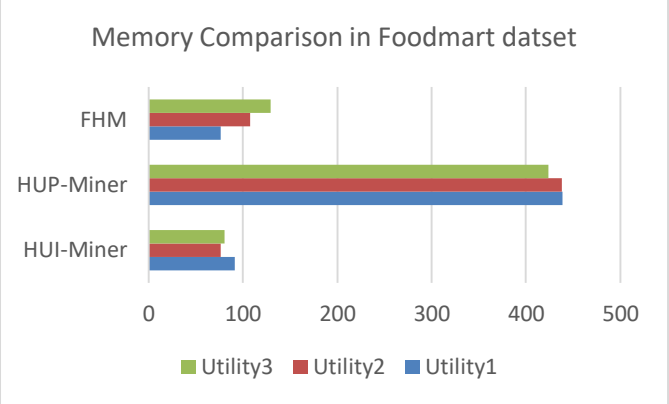

Fig. 5: Comparison of memory space occupied in mb in Foodmart Dataset

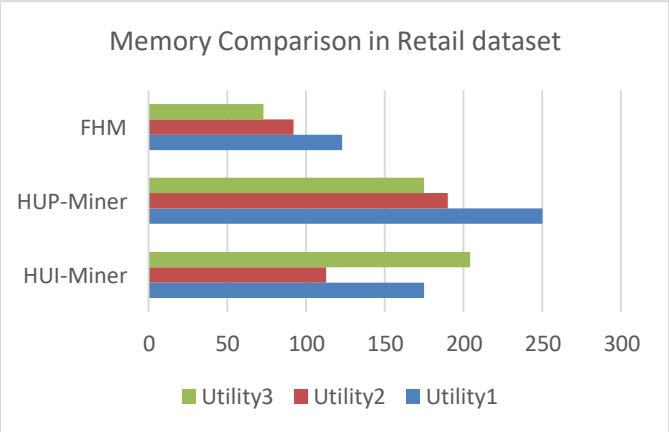

Fig. 6: Comparison of memory space occupied in mb in retail Dataset

IV. CONCLUSION

Mining high utility item sets turn into more significant. In this paper, the High utility item set miner (HUI) algorithms are evaluated. These algorithms are implemented for three different sizes of data for different support threshold value and analyse the performance factors like time and space of FHM with the state-of-the-art algorithm HUI-Miner& HUP-Miner.

REFERENCES

1. Fournier-Viger P, Wu CW, Zida S, Tseng VS (2014), “FHM: faster high-utility itemset mining using estimated utility co-occurrence pruning”. In: International symposium on methodologies for intelligent systems, pp 83–92.
2. Hong TP, Lee CH, Wang SL (2011), “Effective utility mining with the measure of average utility”. Expert Syst Appl 38(7):8259– 8265.
3. J. Han, J. Pei, Y. Yin, and R. Mao (2004), “Mining frequent patterns without candidate generation: A frequent-pattern tree approach,” Data mining and knowledge discovery, vol. 8, no. 1, pp. 53–87.
4. Krishnamoorthy, S. (2014), “Pruning Strategies for Mining High-Utility Itemsets”. Expert Systems with Applications.
5. Liu, M., Qu, J. (2012), “Mining High Utility Itemsets without Candidate Generation”. In: Proceedings of CIKM 2012, pp. 55–64.
6. R Chan, Q. Yang, and Y. D. Shen (2003), “Mining high utility itemsets,” in Data Mining, ICDM 2003. pp. 19–26.
7. U Kanimozhi, J K Kavitha, D Manjula (2014), “Mining High Utility Itemsets – A Recent Survey”, International Journal of Scientific Engineering and Technology, Volume No.3 Issue No.11, pp: 1339-1344.
8. Snehal D. Ambulkar, Prashant N. Chatur (2017), “Efficient Algorithms for Mining High Utility Itemset”. International Conference on Recent Trends in Electrical, Electronics and Computing Technologies (ICRTEECT).
9. S. Shankar, T.P. Purusothoman, S. Jayanthi, N. Babu (2009), "A fast algorithm for mining high utility itemsets", *Proceedings of IEEE International Advance Computing Conference (IACC 2009)*, pp. 1459-1464.
10. 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,” IEEE Transactions on Knowledge and Data Engineering, Vol. 21, No. 12, pp. 1708– 1721.
11. Arati Borker W. (2014), “Utility mining algorithm for high utility itemsets from transactional databases,” International Organization of Scientific R Journals (IOSR Journal of Computer Engineering), Vol. 16, No. 2, pp. 34– 40.
12. Meenakshi S., Sharmila P. (2017), “A Review of Mining high utility itemsets,” International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE), Vol. 5, No. 8, pp. 14055-14061.