A Survey on Approaches for Mining of High Utility Item sets

Nita B. Parmar
G.H. Raisoni College of Engineering
Nagpur (M.S.), India
nbp2910@gmail.com

Asst. Prof. S. K. Jaiswal
G.H. Raisoni College of Engineering
Nagpur (M.S.), India
siddhant.jaiswal@raisoni.net

***Abstract*— we have studied so many algorithms. In this paper we will discuss the advantages and disadvantages of this algorithm. At the end we have proposed an algorithm which will overcome the problems of previous algorithm. Our approach is finding the high utility pattern and frequent pattern for e commerce domain with the help of UP Growth++ algorithm and FP Growth algorithm we apply this approach on real time database system.**

***Keywords*— Utility mining, frequent item set, data mining, high utility item set, frequent pattern, closed+ high utility item set.**

I. INTRODUCTION

In utility mining measured weight, profit, cost, quantity or other information depending on the user preference [1]. The frequent item set mining discovers a large amount of frequent item. In data mining from database system it understood that more high utility item in the algorithm generate the more processing hence it consume the performance of the mining task to decreases greatly while dealing with dense, complex database. The frequent pattern mining and high utility pattern mining to find the frequent pattern and high utility pattern [2]. This algorithm filters most likely databases from main search item, and directly searches on very likely important databases. Frequent item sets find application in a number of real life contexts. The weighted items, is commonly evaluated in terms of the corresponding item weights [4].in this way it will save lots of memory and time and speed up search item precisely. Our approach is finding the high utility pattern and frequent pattern for e commerce domain with the help of up growth++ algorithm and fp growth algorithm we apply this approach on real time database system.

II. LITERATURE SURVEY

Apriori based algorithm for mining high utility closed+ item sets, Apriori HC-D algorithm used discarding unpromising and isolated items, CHUD closed+ high utility item sets discovering to find this representation. The frequent item sets mining discover a large amount of frequent item. Frequent item set mining assumes that every item in a transaction appears in a binary form it means item can be present or absent in a transaction which does not indicate its purchase quantity in the transaction. In utility mining measured weight, profit, cost, quantity or other information depending on the user preference. The performance of the mining task decreases greatly for low minimum utility thresholds. In Frequent item set mining, to reduce the computational cost of the mining task [1].

Utility pattern growth (UP Growth) and UP Growth+ for discovering high utility item set. High utility item sets can be generated from UP tree efficiently with only two scans of original databases. The frequent pattern mining and high utility pattern mining to find the frequent pattern and high utility pattern. High utility item sets from databases refers to finding the item sets with high profits. These item sets utility is interestingness importance or profitability of an item to user. A naïve method is used to find out the high utility item sets from the databases. The potential high utility item sets are found first, and then an additional database scan is performed for identifying their utilities [2].

MinRP set and FlexRP set, to solve the problem. Algorithm MinRP set is similar to RP global to reduce running time and memory usage. MinRP set used a tree structure to store frequent pattern. FlexRP set provides one extra parameter k. This allows user to make a tradeoff between efficiency and the number of representative pattern. Many efficient algorithm have been developed for mining frequent pattern. When the minimum support is low, long pattern can be frequent many frequent pattern have similar items and supporting transactions. The set of frequent closed pattern is lose less representation of the complete set of frequent patterns. Frequent closed patterns supported by exactly the same set of transactions together [3].

Infrequent weighted item sets miner (IWI miner) and Minimal infrequent weighted item sets miner (MIWI miner) IWI miner and MIWI miner are FP Growth like mining algorithm. Frequent item sets find application in a number of real life contexts. The weighted items, is commonly evaluated in terms of the corresponding item weights. The main item sets quantity measures have also been tailored of weighted data. Infrequent item sets that do not contain any infrequent subset. When dealing with optimization problems, minimum and maximum are the most commonly used cost function. In this context, discovering large CPU combinations may be deemed particularly useful by domain experts because they represent large resource sets, which could be reallocated [4].

III. BACKGROUND

Scan the database twice to construct a UP Tree. Recursively generates Potential high utility item sets (PHUIs) from UP-Tree by using UP Growth and by UP Growth++ algorithms [2].Then identify actual high utility item sets from the set of PHUIs. To maintain the information of transactions and high utility item sets [2].Discarding Global Unpromising Items during Constructing a Global UP-Tree. Decreasing Global Node Utilities during Constructing a Global UP-Tree. Constructing a Global UP-Tree by Applying DGU and DGN [2].Discarding Local Unpromising Items during Constructing a Local UP-Tree. Decreasing Local Node Utilities during Constructing a Local UP-Tree. UP-Growth: Mining a UP-Tree by Applying DLU and DLN. To implement UP Growth++, for reducing memory space. Apply UP Growth++ algorithm to find utility pattern. In UP Growth++ algorithm are used to make the estimated pruning values closer to real utility values of the pruned items in database. Apply FP growth algorithm to find the frequent pattern.

IV. CONCLUSIONS

We apply UP Growth++ algorithm to find high utility pattern for maximum profit and use FP growth algorithm to find frequent pattern. Frequent pattern mining from precise data has become popular. Common pattern mining techniques applied to precise data include the mining of contrast patterns, directed acyclic graphs, high utility patterns, and sequential patterns, as well as the stream mining and visual analytics of frequent patterns. To handle uncertain data, we use UP Growth++ algorithm to reduce the number of database scans (down to two scans).Apply UP Growth++ algorithm to improve the mining performance and avoid scanning original database repeatedly. We apply an algorithm UP Growth++ by pushing two more strategies into the framework of FP-Growth.

REFERENCES

[1] Vincent S. Tseng, Cheng-Wei Wu, Philippe Fournier-Viger, and Philip S. Yu, Fellow "Efficient Algorithms for Mining the Concise and Lossless Representation of High Utility Itemsets" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 27, NO. 3, MARCH 2015.

[2] Vincent S. Tseng, Bai-En Shie, Cheng-Wei Wu, and Philip S. Yu, Fellow "Efficient Algorithms for Mining High Utility Itemsets from Transactional Databases" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 8, AUGUST 2013.

[3] Guimei Liu, Haojun Zhang, and Limsoon Wong "A Flexible Approach to Finding Representative Pattern Sets" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 7, JULY 2014.

[4] Luca Cagliero and Paolo Garza "Infrequent Weighted Item set Mining Using Frequent Pattern Growth" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 4, APRIL 2014.

[5] Zhou Zhao, Da Yan, and Wilfred N "Mining Probabilistically Frequent Sequential Patterns in Large Uncertain Databases" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 5, MAY 2014.

[6] Azadeh Soltani and M.-R. Akbarzadeh-T. "Confabulation-Inspired Association Rule Mining for Rare and Frequent Itemsets" IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, VOL. 25, NO. 11, NOVEMBER 2014.

[7] Dennis E.H. Zhuang, Gary C.L. Li, and Andrew K.C. Wong, "Fellow Discovery of Temporal Associations in Multivariate Time Series" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 12, DECEMBER 2014.

[8] V.S. Tseng, C.W. Wu, B.E. Shie, and P.S. Yu. "UP-Growth: An efficient algorithm for high utility itemset mining," ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2010, Washington, DC, United states.

[9] 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, 2009,21(12),pp.1708-1721.

[10] Y. Liu, W.K. Liao and A. Choudhary, eds. A two-phase algorithm for fast discovery of high utility itemsets. Advances in Knowledge Discovery and Data Mining. Vol. 9th Pacific-Asia conference on Advances in Knowledge Discovery and Data Mining. 2005,Berlin:Springer:Hanoi,Vietnam.689-695.

[11] H. Li, H. Huang, Y. Chen, Y. Liu, and S. Lee. "Fast and memory efficient mining of high utility itemsets in data streams," 8th IEEE International Conference on Data Mining, 2008.

[12] J. Liu, K. Wang and B. Fung. "Direct Discovery of High Utility Itemsets without Candidate Generation," 2012 IEEE 12th International Conference on Data Mining, 2012.

[13] M. Liu and J. Qu. "Mining high utility itemsets without candidate generation," 21st ACM International Conference on Information and Knowledge Management, 2012, Maui, HI, United states.

[14] C.W. Wu, B. Shie, V.S. Tseng, and P.S. Yu. "Mining top-K high utility itemsets," 18th ACM SIGKDD international conference on Knowledge discovery and data mining, 2012.

[15] A. Erwin, R.P. Gopalan and N.R. Achuthan. "CTU-mine: An efficient high utility itemset mining algorithm using the pattern growth approach," 7th IEEE International Conference on Computer and Information Technology, 2007.

[16] S. Kannimuthu , Dr. K. Premalatha iFUM - Improved Fast Utility Mining International Journal of Computer Applications (0975 – 8887) Volume 27– No.11, August 2011.

[17] Adinarayanared dy B, O Srinivasa Rao, MHM Krishna Prasad, "An Improved UP-Growth High Utility Itemset Mining" International Journal of Computer publications (0975 – 8887) Volume 58– No.2, November 2012

[18] Song, W., Liu, Y., Li, J.: BAHUI: Fast and memory Efficient mining of high utility itemsets based on bitmap. Intern. Journal of Data Warehousing and Mining. 10(1),1{15 (2014).

[19] R. Chan, Q. Yang, and Y.-D. Shen. Mining high utility itemsets. In Data Mining, 2003. ICDM 2003. Third IEEE International Conference on, pages 19–26, 2003.

[20] K. Sun and F. Bai. Mining weighted association rules without pre assigned weights. Knowledge and Data Engineering, IEEE Transactions on, 20:489–495, 2008.

[21] H. Yao, H. J. Hamilton, and C. J. Butz. A foundational approach to mining itemset utilities from databases. In Proceedings of the Third SIAM International Conference on Data Mining, pages 482–486, 2004.

[22] H. Yao, H. J. Hamilton, and L. Geng. A unified framework for utility-based measures for mining itemsets. In Proceedings of ACM SIGKDD 2nd Workshop on Utility-Based Data Mining, pages 28–37, 2006.