A Review on Mining of High Utility Item Sets

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

Abstract—we studied some algorithms based on utility mining. In this paper we will discuss the pros and cons of this algorithm. So we are putting forth an algorithm which will solve the problem of previous algorithm. Here we are finding the high utility pattern and frequent pattern by using UP Growth++ algorithm and FP growth algorithm for e commerce domain; it will be used on real time data base system.

Key words: Data mining, utility mining, high utility item set, frequent pattern, high utility pattern.

I. INTRODUCTION

In utility mining depending on the user preferences weight, price, number of product, profit by certain product and other information is measured. The frequent item set mining discovers a large amount of frequent item. More high utility item sets in the algorithm will generate more processing this can be seen in data mining from data base system therefore the performance of the mining task will reduces while dealing with dense, complex database. This algorithm filters most likely databases from main search item, and directly searches on very likely important databases. In real life we come across number of application on frequent item sets. In this way it will save lots of memory and time and speed up search item precisely. Our approach is finding the high utility item sets and frequent pattern item sets for e commerce domain by using UP Growth++ algorithm and FP growth algorithm we can use this approach on real time database system.

II. LITERATURE SURVEY

A. Apriori Algorithm:

Apriori algorithm works on global node or global item sets. It functions on DGN and DGU, it uses whole data set. It scans whole data and generates new entry and it scans the data step by step, therefore it consumes more memory as well as more time.

B. Up Growth+ Algorithm:

High utility item sets can be found out by using UP growth+ algorithm. It generates the tree of the given data set. After generating the tree it generates high utility item sets from the tree. As the high utility item sets is generated, then it can easily find out most profitable item.

C. Minrp Set and Flexrp Set Algorithm:

MinRP set and FlexRP set algorithm plays huge role in reducing the time and memory. In a way it saves user's time. The memory saved can be used for other profitable item sets. These algorithms find the most profitable and most selling product from the data base. MinRP and flexRP algorithms give us frequent pattern of item set.

D. Potential High Utility Item Sets Algorithm:

Potential high utility item set algorithm will select the products which have highest profit. In this way we can find out the list of products which have maximum profit, and by using this data we can increase the quantity of certain products.

III. BACKGROUND STUDY

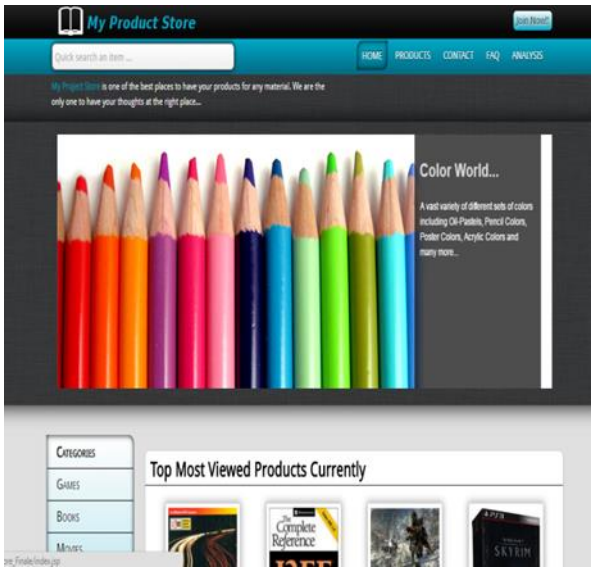

Fig. 1:

My product store is an online website is intended for customer specifically looking for best solution in online purchase of following of item like games, books, movies, computer etc. It gives the best and updated solution in cost and time for the selective item. The web page shown is the first step of this project. This page will give brief idea about the items available. To have access, the user should first register him, and then log in. User wants to access the web side and search the required product.Collection of most of popular brands in the selected item category intelligence search saving the time and best updated best price. The user will be able to access all the available items. The screenshot shows five tabs. When the user will click on the Product tab, user would see category and sub category of the product. The contact tab will allow the user to contact the product store, in case of any query or problem. The FAQs tab shows the questions asked by the different users. We work on real time data base, here we generate 1213 transaction, we create 205 product, we create sub category of product. The whole information of user saves on user details column.

IV. RESULT AND CONCLUSION

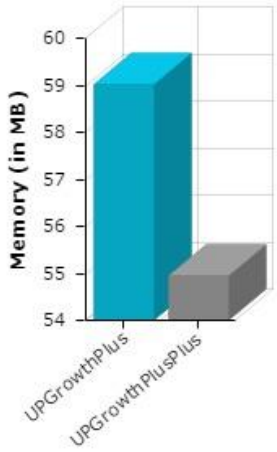

Fig. 2:

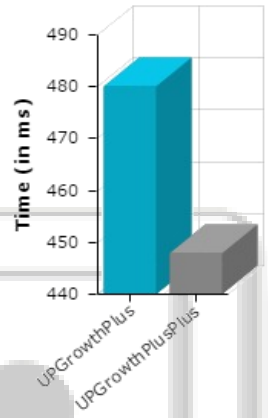

Fig. 3:

Both the graphs show the result of the project. These graphs show the comparison between two algorithms i.e. UP Growth+ algorithm and UP Growth++ algorithm. In Upgrowth+ algorithm, the scanning of database repeats again and again therefore it occupies more memory and more time. In UP Growth++ algorithm data will be scanned in one time, thus it saves more memory and time.

V. CONCLUSION

The algorithm is used to resolve the problem in scanning the data base again and again. The main work was to design a UP Growth++ algorithm for reducing the scanning. We apply the UP Growth++ algorithm to reduce the scanning, so that it will save the more memory as well as time.

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.

