

RHUI Based Mining for Efficient Discovery of Itemsets in Knowledge Representation

Ms.M.Vijayalakshmi¹, Mr. S.J.Vivekanandan², Ms.P.Subhashree³

¹PG Scholar, PRIST University, Vallam, Thanjavur.

²Research Scholar, Sathyabama University, Chennai & Asst. Prof., PRIST University, Vallam, Thanjavur.

³Research Scholar, Sathyabama University, Chennai.

Abstract— Mining high utility item sets (HUIs) from databases is an essential data mining task, for finding of itemsets with high utilities. Anyhow, they generate more number of itemsets to user so it decreases the energetic of the mining process. To attain high efficiency the closed high utility itemsets (CHUIs) is produced, which serves as a reduced form of high utility itemsets without loss of information and quantity. CHUD (Closed High Utility Itemsets Discovery) is used to reduce the HUI with combination of DAHU (Derive All High Utility). The number of high utility itemsets on all real dataset may be more, To handle this issue, we use an efficient method called Redundant High utility item (RHUI) to avoid the repeated mining of closed high utility and also it reduces time to find the high utility items. Classify the closed high utility itemsets into three group's high, medium and low then again discover the high utility itemsets from the classification. Final set of high utility itemsets is generated.

Key Words: —Frequent itemsets mining, closed high utility itemsets, utility mining, data mining.

I INTRODUCTION

A. DATA MINING

Data mining is the extraction of unknown predictive information from large databases. The Data mining involves the process of knowledge

discovery in databases (KDD). It estimates and understands the possible patterns and selects the knowledge [1]. It also used for extracting patterns from data.

Data mining is concerned with evaluation of massive volumes of statistics to routinely discover thrilling regularities or relationships which in turn leads to higher understanding of the underlying approaches.

The real data mining task is the automated or semi-automatic analysis of massive portions of information to extract formerly unknown thrilling patterns.

B. FREQUENT ITEMSET MINING

Frequent itemset mining (FIM) [2], [4], [5], [6] is a process to find the itemset which is appears frequently in the transaction database. It is a usual model to discover the frequent itemsets. In this process the items are taken by equal importance and it assumes that every item in a transaction appears in a binary form, i.e., it does not take the quantity of the items. So FIM cannot satisfy the user requirement. However, utility of an itemset like profit, quantity and weight are important for addressing real world decision problems that require maximizing the utility in an organization. In many

areas of business like retail, inventory, marketing research etc. decision making is very important.

C.HIGH UTILITY ITEMSET MINING

The utility of the itemsets is greater than the user given minimum utility, where the utility represents the cost, quantity, and other basic needs depending on the user preferences[7], [8]. The profit of an itemset depends not only on the support (total number of items an itemset appear in a transactional database out of the total number of transactions) of the itemset, but also on the prices of the items in that itemset. Our goal must be in identifying itemsets that have better utilities within the database, regardless of whether or not those itemsets are common itemsets or not [9], [10].

II RELATED WORK

In market basket analysis, Frequent Itemset Mining (FIM) is used to discover the frequent itemset in transactional database. It produce more number of itemsets because FIM only considers whether an item has occurred in transactional database or not, it does not consider their quantity and profits It does not reveal the actual utility of an itemset, which can be measured in terms of cost, quantity, profit, or other expressions of user preference.

However, HUI mining is not a simple job while the downward closure property in FIM does not hold in utility mining [2], [15]. The search space for mining HUIs cannot be directly reduced as it is done in FIM because a superset of a low utility itemset can be a high utility itemsets [2],[3],[6],[7]. Many studies, they produce large number of high utility itemsets. So it difficult to user to understand the results and it degrade the efficiency of the algorithm. The performance of the mining task

decreases greatly for low minimum utility thresholds or when dealing with dense databases.

In existing they discover a reduced form of itemsets but they found in frequent itemsets not in HUI. Many algorithms are used to generate the high utility itemsets; it may slow the mining process and also not suitable for big databases.

III PROPOSED METHOD

In proposed closed high utility itemset is introduced, which integrate the concept of closed itemset into high utility itemset mining. Mine all the high utility itemsets first and then apply the closed constraint. And Mine all the closed itemsets first and then apply the utility constraint. Three efficient algorithms are used to mining the high utility itemsets. A novel framework is used in this paper to find the CHUI.

RHUI (Redundant High Utility Itemset Mining)

RHUI (Redundant High Utility Itemset Mining) is a method used to avoid the repeated mining process of same itemsets during the discovery of closed high utility itemset. The final results of high utility items and closed high utility itemset are arranged in ascending order and then the set of closed high utility itemsets is classified into three groups based upon their utilities like high, medium and low.

A. DATASET COLLECTION

Analysis is one of the most important data-mining models. A database is an organized collection of data (Fig.1). Collect data from the database and clustered the items based upon their quantities, profit, price.

B. FREQUENT ITEMSET MINING

Mining the itemsets which is frequently purchased by the customer using FIM (Frequent Itemset Mining) process, it would not consider the quantity and importance of the items like profit and cost (Fig.1). It produces more number of frequent itemsets so it degrades the mining process.

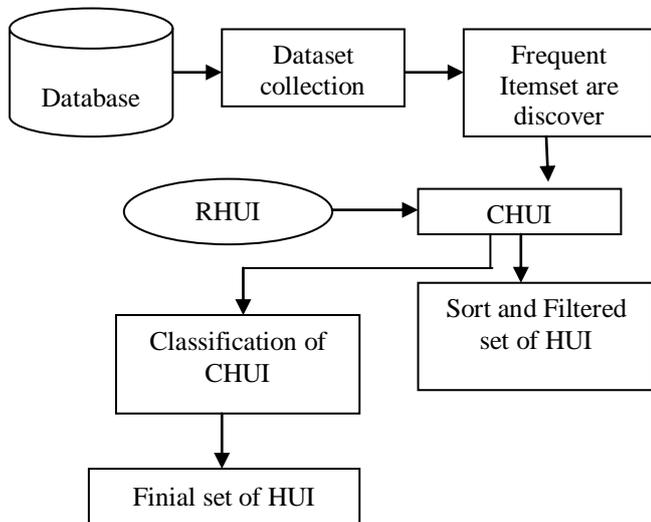


Fig.1.Flow diagram

C. CLOSED HIGH UTILITY ITEMSETS

Closed high utility itemsets are discovering in this module using AprioriHC-D (AprioriHC algorithm with discarding unpromising and isolated items) and CHUD (Closed High Utility Itemset Discovery) algorithms.

In AprioriHC-D they use two effective strategy DGU (Discarding Global Unpromising items) [14] and IIDS (Isolated Items Discarding Strategy) [12] to generate the set of CHUI. In DGU the first examine of database, the Transaction Utility (TU) of each transaction is detected and TWU (Transaction weighted utilization) [11] of each single item is also computed. An item which have greater TWU than the minimum utility

threshold which is specified by the user is called as promising item otherwise the item is unpromising item. Unpromising items are removed from the transaction and their utilities are eliminated from the TU of the transaction since only the supersets of promising items are possible to be the high utility itemsets. IIDS strategy eliminates the isolated items and their exact utilities from the Transaction Utility.

Redundant High Utility Itemset is method used to avoid the repeated combination of itemsets in (Fig.1) closed high utility itemset and their utilities are removed.

D. HIGH UTILITY ITEMSET PROCESS

Mining the high utility itemsets from the closed high utility itemsets then sorting into ascending order. By using the Closed High Utility Itemsets Discovery (CHUD) and DAHU (Derive All High Utility). Now classify the closed high utility itemsets into three classes high, medium and low then again discover the high utility itemsets from the classification (Fig.1). Final set of high utility itemsets is generated.

V RESULTS

The RHUI (Redundant high utility Itemsets) technique is removing the repeated combinations of closed high utility itemsets. Compare with existing set of closed high utility itemsets we discover accurate utility of an itemsets and time consumption of the mining process is reduced.

And also we compare (Fig.2) the first, second phase of HUI and identified the experimental results that showed the second phase provide more efficient results.

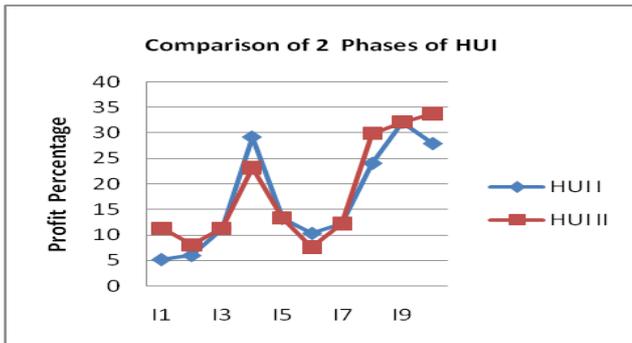


Fig.2.comparison of 2 phases of HUI

V CONCLUSION

In this paper we have proposed redundant high utility itemsets (RHUI) method to avoid the repeated mining process of same itemset in closed high utility itemsets (CHUI). First phase of High utility itemsets is generated by using Closed High Utility itemsets Discovery (CHUD) algorithm along with the combination of DAHU (Derive All High Utility). CHUI is classified into three classes and then again HUI is discovered from this classification. Here we prove the difference between the first phase of HUI and another set of HUI. The proposed system has applications in website click stream analysis, business promotion in chain supermarkets, cross marketing in retail stores, online e-commerce management.

REFERENCES

1. J. Han, M. Kanber, and J. Pei, *Data mining: Concepts and Techniques*, 3rd ed. Burlington, MA, USA: Morgan Kaufmann, 2011.
2. R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in *Proc. 20th Int. Conf. Very Large Data Bases*, 1994, pp. 487–499.
3. C.-W. Lin, T.-P. Hong and W.-H. Lu, "An effective tree structure for mining high utility itemsets," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 7419–7424, 2011.
4. T. Calders and B. Goethals, "Mining all non-derivable frequent itemsets," in *Proc. Int. Conf. Eur. Conf. Principles Data Mining Knowl. Discovery*, 2002, pp. 74–85.
5. K. Chuang, J. Huang, and M. Chen, "Mining top-k frequent patterns in the presence of the memory constraint," *VLDB J.*, vol. 17, pp. 1321–1344, 2008.
6. K. Gouda and M. J. Zaki, "Efficiently mining maximal frequent itemsets," in *Proc. IEEE Int. Conf. Data Mining*, 2001, pp. 163–170.
7. 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 Trans. Knowl. Data Eng.*, vol. 21, no. 12, pp. 1708–1721, Dec. 2009.
8. R. Chan, Q. Yang, and Y. Shen, "Mining high utility itemsets," in *Proc. IEEE Int. Conf. Data Min.*, 2003, pp. 19–26.
9. A. Erwin, R. P. Gopalan, and N. R. Achuthan, "Efficient mining of high utility itemsets from large datasets," in *Proc. Int. Conf. Pacific-Asia Conf. Knowl. Discovery Data Mining*, 2008, pp. 554–561.
10. G.-C. Lan, T.-P. Hong, and V. S. Tseng, "An efficient projection-based indexing approach for mining high utility itemsets," *Knowl. Inf. Syst.*, vol. 38, no. 1, pp. 85–107, 2014. *Inf. Syst.*, vol. 38, no. 1, pp. 85–107, 2014.
11. Y. Liu, W. Liao, and A. Choudhary, "A fast high utility itemsets mining algorithm," in *Proc. Utility-Based Data Mining Workshop*, 2005, pp. 90–99.
12. Y.-C. Li, J.-S. Yeh, and C.-C. Chang, "Isolated items discarding strategy for discovering high utility itemsets," *Data Knowl. Eng.*, vol. 64, no. 1, pp. 198–217, 2008.
13. B. Vo, H. Nguyen, T. B. Ho, and B. Le, "Parallel method for mining high utility itemsets from vertically partitioned distributed databases," in *Proc. Int. Conf. Knowl.-Based Intell. Inf. Eng. Syst.*, 2009, pp. 251–260.
14. 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. ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2010, pp. 253–262.
15. J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2000, pp. 1–12.