

Efficient Finding of Item sets in Knowledge Representation using RHUI Extraction

K. Uma

Department of SITE, VIT University, Vellore, India.

*Corresponding author: E-Mail: dr.k.uma.phd@gmail.com

ABSTRACT

Mining high utility item sets (HUIs) from databases is an essential data mining task, for finding of item sets with high utilities. Anyhow, they generate more number of item sets to user so it decreases the energetic of the mining process. To attain high efficiency the closed high utility item sets (CHUIs) is produced, which serves as a reduced form of high utility item sets without loss of information and quantity. CHUD (Closed High Utility Item sets Discovery) is used to reduce the HUI with combination of DAHU (Derive All High Utility). The number of high utility item sets 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 item sets into three group's high, medium and low then again discover the high utility item sets from the classification. Final set of high utility item sets is generated.

KEY WORDS: Closed high utility item sets, utility mining, data Extraction, Frequent item sets Extraction.

1. INTRODUCTION

Data Extraction in KDD: 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. 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.

Recurrent Item set Extraction: Frequent item set mining (FIM) is a process to find the item set which is appears frequently in the transaction database. It is a usual model to discover the frequent item sets. 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 item set 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.

Frequent Utility Item set Extraction: The utility of the item sets is greater than the user given minimum utility, where the utility represents the cost, quantity, and other basic needs depending on the user preferences. The profit of an item set depends not only on the support (total number of items an item set appear in a transactional database out of the total number of transactions) of the item set, but also on the prices of the items in that item set. Our goal must be in identifying item sets that have better utilities within the database, regardless of whether or not those item sets are common item sets or not.

Related Work: In market basket analysis, Frequent Item set Mining (FIM) is used to discover the frequent item set in transactional database. It produce more number of item sets 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 item set, 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. The search space for mining HUIs cannot be directly reduced as it is done in FIM because a superset of a low utility item set can be a high utility item sets. Many studies, they produce large number of high utility item sets. 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 item sets but they found in frequent item sets not in HUI. Many algorithms are used to generate the high utility item sets; it may slow the mining process and also not suitable for big databases.

2. PROPOSED METHOD

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

RHUI (Redundant High Utility Item set Mining): RHUI (Redundant High Utility Item set Mining) is a method used to avoid the repeated mining process of same item sets during the discovery of closed high utility item set. The

final results of high utility items and closed high utility item set are arranged in ascending order and then the set of closed high utility item sets is classified into three groups based upon their utilities like high, medium and low.

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.

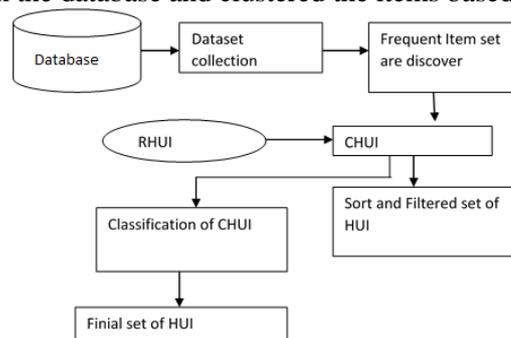


Fig.1. Flow diagram RHUI

Frequent Item Set Mining: Mining the item sets which is frequently purchased by the customer using FIM (Frequent Item set 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 item sets so it degrades the mining process.

Closed High Utility Item Sets: Closed high utility item sets are discovering in this module using Apriori HC-D (Apriori HC algorithm with discarding unpromising and isolated items) and CHUD (Closed High Utility Item set Discovery) algorithms. In Apriori HC-D they use two effective strategy DGU (Discarding Global Unpromising items) and IIDS (Isolated Items Discarding Strategy) 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) 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 item sets. IIDS strategy eliminates the isolated items and their exact utilities from the Transaction Utility.

Redundant High Utility Item set is method used to avoid the repeated combination of item sets in (Fig.1) closed high utility item set and their utilities are removed.

High Utility Item Set Process: Mining the high utility item sets from the closed high utility item sets then sorting into ascending order. By using the Closed High Utility Item sets Discovery (CHUD) and DAHU (Derive All High Utility. Now classify the closed high utility item sets into three classes high, medium and low then again discover the high utility item sets from the classification. Final set of high utility item sets is generated as shown in Fig.1.Flow diagram.

3. EXPERIMENTAL RESULTS

The RHUI (Redundant high utility Item sets) technique is removing the repeated combinations of closed high utility item sets. Compare with existing set of closed high utility item sets we discover accurate utility of an item sets 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.

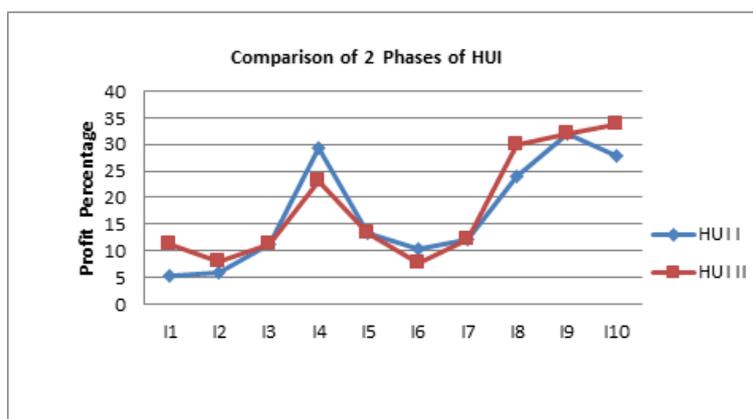


Fig.2. comparison of 2 phases of HUI

4. CONCLUSION

In this paper we have proposed redundant high utility item sets (RHUI) method to avoid the repeated mining process of same item set in closed high utility item sets (CHUI). First phase of High utility item sets is generated by using Closed High Utility item sets 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 includes the applications of Website click stream analysis, Business promotion in chain hypermarkets, Cross marketing in retail stores, online e-commerce management.

REFERENCES

- Agrawal R, and Srikant R, Fast algorithms for mining association rules, in Proc. 20th Int. Conf. Very Large Data Bases, 1994, 487–499.
- Ahmed CF, Tanbeer SK, Jeong BS, and Lee YK, Efficient tree structures for high utility pattern mining in incremental databases, *IEEE Trans. Knowl. Data Eng.*, 21 (12), 2009, 1708 – 1721.
- Calders T, and Goethals B, Mining all non-derivable frequent item sets, in Proc. Int. Conf. Eur. Conf. Principles Data Mining Knowl. Discovery, 2002, 74–85.
- Chan R, Yang Q, and Shen Y, Mining high utility item sets, in Proc. IEEE Int. Conf. Data Min., 2003, 19–26.
- Chuang K, Huang J, and Chen M, Mining top-k frequent patterns in the presence of the memory constraint, *VLDB J*, 17, 2008, 1321–1344.
- Erwin A, Gopalan R.P, and Achuthan N.R, Efficient mining of high utility item sets from large datasets, in Proc. Int. Conf. Pacific- Asia Conf. Knowl. Discovery Data Mining, 2008, 554–561.
- Gouda K, and Zaki M.J, Efficiently mining maximal frequent item sets, in Proc. IEEE Int. Conf. Data Mining, 2001, 163–170.
- Han J, Kanber M, and Pei J, *Data mining: Concepts and Techniques*, 3rd ed., Burlington, MA, USA: Morgan Kaufmann, 2011.
- Han J, Pei J, and Yin Y, Mining frequent patterns without candidate generation, in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2000, 1–12.
- Lan GC, Hong TP, and Tseng VS, An efficient projection based indexing approach for mining high utility item sets, *Knowl. Inf. Syst.*, 38 (1), 2014, 85–107.
- Li YC, Yeh JS, and Chang CC, Isolated items discarding strategy for discovering high utility item sets, *Data Knowl. Eng.*, 64 (1), 2008, 198–217.
- Lin CW, Hong TP, and Lu WH, An effective tree structure for mining high utility item sets, *Expert Syst. Appl.*, 38 (6), 2011, 7419–7424.
- Liu Y, Liao W, and Choudhary A, A fast high utility item sets mining algorithm, in Proc. Utility-Based Data Mining Workshop, 2005, 90–99.
- Tseng VS, Wu CW, Shie BE, and Yu PS, UP Growth: An efficient algorithm for high utility item set mining, in Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining, 2010, 253–262.
- Vo B, Nguyen H, Ho T.B, and Le B, Parallel method for mining high utility item sets from vertically partitioned distributed databases, in Proc. Int. Conf. Knowl.-Based Intell. Inf. Eng. Syst., 2009, 251–260.