ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

A Survey on Mining Algorithm to Archive High Utility Item Set Using TKO With TKU.

Sakshi Karnawat

Department of Information Technology, Maharashtra Institute of Technology, Savitribai Phule Pune University, Pune, India

ABSTRACT: Data mining is computing process of discovering patterns in large data sets involving methods at the intersection of data base system.Data mining popular problem is High utility item set mining (HUI) or more generally utility mining (UI).The problem of HUI (High utility item set) is mainly introduction to frequently item set. Frequent pattern mining is a popular problem in data mining, which consists in finding frequent patterns in transaction databases. To solve the Problem of High utility item set (HUI) with some particular data and state of art of algorithms. To archive the HUI (High utility item set) many popular algorithm have been proposed for this problem such as “Apriori”, FP growth etc. but Now a days most popular algorithms TKO (mining Top-K utility item sets in One phase) and TKU (mining Top-K Utility item sets) here TKO is Top K in one phases and TKU is Top K in utility. In this paper address the above the issues by proposing a new frame work for top k HUI where k is desired number of HUIs to be mined. High utility item set mining is an uncommon term. But we are using it while we are doing online purchases etc. It is a part of business analytics .It Main application area is market basket analysis where when customer purchase the item he can buy another item to maximize profit.so both customer and vendors earn profit.

KEYWORDS: utility mining, high utility item set, top k- pattern mining, top- k high utility item set mining.

I. INTRODUCTION

Data Mining is the efficient discovery of valuable and vivd information from a large collection of data. Frequent item set mining (FIM) discover the only frequent items but HUI High Utility item set. In FIM profile of item set are not considered. This is because the purchase quantity not takes into account. Data mining is the process of analyzing data from different angles and summarizing it into useful data. Data mining is a tool for analyzing data. It allows users to analyze data from different levels or angles, arrange it, and the relationships among the data are found. Data mining is the process of finding patterns among sufficient of fields in large relational database. A classical Top K model-based algorithm consists of two phases. In the first phase, called phase I, the complete setof HTWUIs are found. In the second phase, called phase II, all HUIs are obtained by calculating the exact utilities of HTWUIs with one database scan. Although many studies have been devoted to HUI mining, it is difficult for users to choose an appropriate minimum utility threshold in practice. Depending on the threshold, the output size can be very small or very large. Besides, the choice of the threshold greatly influences the performance of the algorithms. If the threshold is set too low, too many HUIs will be presented to the users and it is difficult for the users to comprehend the results. A large number of HUIs also causes the mining algorithms to become inefficient or even run out of memory, because the more HUIs the algorithms generate, the more resources they consume. On the contrary, if the threshold is set too high, no HUI will be found.

II. BACKGROUND

Frequent item set mining (FIM) problem for Data Mining to archive the High utility item set mainly High utility is problem are solve the two such Algorithms TKO(Top k in one phases) and TKU (Top K in Utility). As paper analysis TKO is most efficient algorithm but one drawback is result of algorithm is not correct. But we disused about the TKU get the utility but Execution time are very high as compared to TKO algorithm If Data base are selected for mining is denser data set then profit of user or product are consider as the High utility item set. Frequently generate a huge set of

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

HUIs and their mining performance is degraded consequently. Further in case of long transactions in dataset or low thresholds are set, then this condition may become worst. The huge number of HUIs forms a challenging problem to the mining performance since the more HUIs the algorithm generates, the higher processing time it consumes. Thus to overcome this challenges the efficient algorithms presented. Top k will not work on the parallel mining.

AIM

1. Set the value of k which is more intuitive than setting the threshold because k represents the number of Item sets that users want to find whereas choosing the threshold depends primarily on database characteristics, which are often unknown to users.
2. The main point of min_util variable is not given in advance in top k HUI mining In traditional HUI mining the search space can be efficiently pruned by the algorithm by using a given the min_util threshold value. In scenario of TKO and TKU algorithm min_util threshold value is provided in advance.

SCOPE

1. The main concepts are when TKO algorithm scans the data set more than ten rounds. In proposed algorithm scans the database only once and hence requires less memory compared to the present algorithms.
2. While TKO generates the item set, without scanning the database the utility-list calculates the utility. Initially the threshold is set to zero.
3. Min_util value is not given in the advances algorithm take as the dynamically values.

III. RELATED WORK

ChowdhuryFarhan Ahmed, Syed KhairuzzamanTanbeer, Byeong-SooJeong, and Young-Koo Lee presented three novel tree structures for efficiently perform incremental and interactive HUP mining[2]. The first tree structure is used to arrange the items according to their lexicographic order. It is known as Incremental HUP Lexicographic Tree (IHUPLTree). It can capture the incremental data without any restructuring operation. The next tree structure is the IHUP Transaction Frequency Tree (IHUPTF-Tree), which arranging items according to their transaction frequency in descending order. To reduce the mining time, the last tree, IHUP-Transaction-Weighted Utilization Tree (IHUPTWUTree) is designed. Structure of this tree is based on the TWU value of items in descending order.

Alva Erwin, Raj P. Gopalan, and N. R. Achuthan, proposed CTU-PROL algorithm for efficient mining of high utility item sets from large datasets[3]. This algorithm finds the large TWU items in the transaction database. If data sets is too large to be held in main memory, the algorithm creates subdivisions using parallel projections and for each subdivision, a Compressed Utility Pattern Tree (CUP-Tree) is used to mine the complete set of high utility item sets. If the dataset is small, it creates a single CUP-Tree for mining high utility item sets.

Shankar S., Purusothaman T., Jayanthi, S., suggested a novel algorithm for mining high utility itemsets[4]. This fast utility mining (FUM) algorithm finds all high utility itemsets within the given utility constraint threshold. The proposed FUM algorithm scales well as the size of the transaction database increases with regard to the number of distinct items available.

R. Chan, Q. Yang, and Y. Shen, suggested mining high utility itemsets[5]. They proposed a novel idea of top-K objective-directed data mining, which focuses the top-K high utility closed patterns. They add the concept of utility to capture highly desirable statistical patterns and present a level wise itemset mining algorithm. They develop a new pruning strategy based on utilities that allow pruning of low utility itemsets to be done by means of a weaker but ant monotonic condition.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

Ramaraju C., Savarimuthu N., proposed a conditional tree based novel algorithm for high utility itemset mining[6]. A novel conditional high utility tree (CHUT) compress the transactional databases in two stages to reduce search space and a new algorithm called HU-Mine is proposed to mine complete set of high utility item sets.

Y. Liu, W. Liao, and A. Choudhary, proposed a fast high utility itemsets mining algorithm[7]. They are present a Two-Phase algorithm to efficiently prune down the number of candidates and can precisely obtain the complete set of high utility itemsets. In the first phase, they propose a model that applies the "transaction-weighted downward closure property" on the search space to expedite the identification of candidates. Latter phase identifies the high utility itemsets

Adinarayanareddy B., O. SrinivasaRao, MHM Krishna Prasad, suggested improved UP-Growth high utility itemset mining[8]. The compact tree structure, Utility Pattern Tree i.e. UP-Tree, maintains the information of transactions and their itemsets. It facilitates the mining performance and avoid scanning original database frequently. UP-Tree scans database only twice to obtain candidate items and manage them in an efficient data structured way. UP-Growth takes more execution time for Second Phase by using UP-Tree. Hence they presents modified algorithm aiming to reduce the execution time by effectively identifying high utility itemsets.

P. Asha, Dr. T. Jebarajan, G. Saranya, presents a survey on efficient incremental algorithm for mining high utility itemsets in distributed and dynamic database[9]. The proposed system employs one master node and two slave nodes. Database is partitioned for every slave node for computation. The slave node counts the occurrence of each item. These data's are stored in their local table. Then each slave node sends these tables to master node. The Master Node maintain global table for storing these data. Based on the minimum utility threshold value it calculates the promising and unpromising itemsets.

IV. EXISTING SYSTEM APPROACH

In existing, frequently search item mining is a research (FIM) topic in data mining. However, the traditional mining may discover a large amount of frequent but low-value item sets and lose the information on valuable item sets having low selling frequencies. Hence, it cannot satisfy the requirement of users who desire to discover item sets with high utilities such as high profits. To address these issues, utility mining emerges as an important topic in data mining and has received extensive attention in recent years. In utility mining, each item is associated with a utility (e.g. unit profit) and an occurrence count in each transaction (e.g. quantity). The utility of an item set represents its importance, which can be measured in terms of weight, value, quantity or other information depending on the user specification. An item set is called high utility item set (HUI) if its utility is no less than a user-specified minimum utility threshold min_util. HUI mining is essential to many applications such as streaming analysis, market analysis, mobile computing and biomedicine. The Two most algorithm are used to archive the HUI (High utility item set) TKO (Top K in one phase) and TKU (Top K in utility) both algorithm are generated the result individually.

Disadvantages:-

1. Value of K is fixed in the algorithms
2. min_util variable value is fixed in both TKO and TKU.
3. TKO algorithm is implement with k value set and the min_util value is stored in advances
4. The Algorithm TKO (Top K in one phases) result are not accurate ,resultconatin the garbage value.

V. PROPOSED SYSTEM APPROACH

In Proposed system, we address the above issues by proposing a new framework efficient hybrid algorithm for high utility item set in parallel mining using TKU and TKO .Two types of efficient algorithms named TKU (mining Top-K Utility Itemsets) and TKO (mining Top-K utility temsets in One phase) are proposed for mining such Itemsets without the need to set min utility. But algorithm TKO have the major disadvantage to result are not accreted mainly the result of TKO given the garbage value in High utility item set. And TKU algorithm result is accreted but Execution time is high then alternative solution is that find the efficient algorithm then In proposed system combination of TKO and TKU algorithms. Can say that result of TKO Top K in one phase is given the input of TKU Top K in utility Result of TKO and TKU is accreted and Execution time is low. In proposed System new algorithm is generated for combination of TKO and TKU name as the TKOWITHTKU or TKMHUI Top k Main High utility item set .

Advantages:

1. The K value is Not fixed its value is user defined.
2. Min_util value is not given the advances
3. Execution time is Low.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

Figure 1: Proposed System Architecture

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

VI. CONCLUSION

In this paper, we have studied the problem of top-k high utility item sets mining, where k is the desired number of high utility item sets to be mined. The Most efficient TKOWITHTKU combination of TKO and TKU algorithms are proposed for mining such item sets without setting minimum utility thresholds. On the other hand, TKO is the first one-phase algorithm developed for top-k HUI mining is called the PHUI (potential high utility item set) and PHUI are given to TKU in utility phases. Empirical evaluationson different types of real and synthetic datasetsshow that the proposed algorithms have good scalabilityon large datasets and the performance of the proposedalgorithms is close to the optimal case of the state-of-thecombination of both phases in one algorithm.

REFRENCES

1. Vincent S. Tseng, Bai-En Shie, Cheng-Wei Wu, and Philip S. Yu, Fellow, IEEE "Efficient Algorithms for Mining High Utility Itemsets from Transactional Databases" IEEE Trans. Knowledge and Data Engineering, vol. 25, no. 8, August 2013.
2. ChowdhuryFarhan Ahmed, Syed KhairuzzamanTanbeer, Byeong-SooJeong, and Young-Koo Lee, Member, IEEE "Efficient Tree Structures for High Utility Pattern Mining inIncremental Databases" IEEE Trans. Knowledge and Data Engineering, Vol. 21, Issue 12, pp. 1708-1721, December 2009.
3. Alva Erwin, Raj P. Gopalan, and N. R. Achuthan, "Efficient Mining of High Utility Itemsets from Large Datasets", In Proc. of PAKDD 2008
4. Shankar, S.; Purusothaman, T.; Jayanthi, S. "Novel algorithm for mining high utility itemsets" International Conference on Computing, Communication and Networking, Dec. 2008.
5. Raymond Chan; Qiang Yang; Yi-Dong Shen, "Mining high utility itemsets" In Proc. of Third IEEE Int'l Conf. on Data Mining ,November 2003.
6. Ramaraju, C., Savarimuthu N. "A conditional tree based novel algorithm for high utility itemset mining", International Conference on Data mining, June 2011.
7. Ying Liu, Wei-keng Liao, AlokChoudhary "A Fast High Utility Itemsets Mining Algorithm" In Proc. of the Utility-Based Data Mining Workshop, 2005.
8. Adinarayanareddy B ,O SrinivasaRao, MHM Krishna Prasad, "An Improved UP-GrowthHigh Utility Itemset Mining" International Journal of Computer Applications (0975-8887) Volume 58-No.2, November 2012
9. P. Asha, Dr. T. Jebarajan, G. Saranya, "A Survey on Efficient Incremental Algorithm for Mining High Utility Itemsets in Distributed and Dynamic Database" IJETAE Journal, Vol.4, Issue 1, January 2014.