**Efficient Algorithms for Mining Top-K High Utility Itemsets**

**ABSTRACT:**

High utility itemsets (HUIs) mining is an emerging topic in data mining, which refers to discovering all itemsets having a utility meeting a user-specified minimum utility threshold min\_util. However, setting min\_util appropriately is a difficult problem for users. Generally speaking, finding an appropriate minimum utility threshold by trial and error is a tedious process for users. If min\_util is set too low, too many HUIs will be generated, which may cause the mining process to be very inefficient. On the other hand, if min\_util is set too high, it is likely that no HUIs will be found. In this paper, we address the above issues by proposing a new framework for top-k high utility itemset mining, where k is the desired number of HUIs to be mined. Two types of efficient algorithms named TKU (mining Top-K Utility itemsets) and TKO (mining Top-K utility itemsets in One phase) are proposed for mining such itemsets without the need to set min\_util. We provide a structural comparison of the two algorithms with discussions on their advantages and limitations. Empirical evaluations on both real and synthetic datasets show that the performance of the proposed algorithms is close to that of the optimal case of state-of-the-art utility mining algorithms.

**EXISTING SYSTEM:**

* The traditional FIM (Frequent itemset mining) may discover a large amount of frequent but low-value itemsets and lose the information on valuable itemsets having low selling frequencies. Hence, it cannot satisfy the requirement of users who desire to discover itemsets 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 itemset represents its importance, which can be measured in terms of weight, value, quantity or other information depending on the user specification. An itemset is called high utility itemset (HUI) if its utility is no less than a user-specified minimum utility threshold min\_util.
* In recent years, high utility itemset mining has received lots of attention and many efficient algorithms have been proposed, such as Two-Phase, IHUP, IIDS, UPGrowth, d2HUP and HUI-Miner. These algorithms can be generally categorized into two types: twophase and one-phase algorithms.

**DISADVANTAGES OF EXISTING SYSTEM:**

* Although many studies have been devoted to HUI mining, it is difficult for users to choose an appropriate minimum utility threshold in practice.
* The existing studies may perform well in some applications, they are not developed for top-k high utility itemset mining and still suffer from the subtle problem of setting appropriate thresholds.

**PROPOSED SYSTEM:**

* In this paper, we address all of the above challenges by proposing a novel framework for top-k high utility itemset mining, where k is the desired number of HUIs to be mined.
* Major contributions of this work are summarized as follows:
* First, two efficient algorithms named TKU (mining Top-K Utility itemsets) and TKO (mining Top-K utility itemsets in One phase) are proposed for mining the complete set of top-k HUIs in databases without the need to specify the min\_util threshold.
* The TKU algorithm adopts a compact tree-based structure named UP-Tree to maintain the information of transactions and utilities of itemsets. TKU inherits useful properties from the TWU model and consists of two phases.
* In phase I, potential top-k high utility itemsets (PKHUIs) are generated. In phase II, top-k HUIs are identified from the set of PKHUIs discovered in phase I. On the other hand, the TKO algorithm uses a list-based structure named utility-list to store the utility information of itemsets in the database.
* It uses vertical data representation techniques to discover top-k HUIs in only one phase.

**ADVANTAGES OF PROPOSED SYSTEM:**

* Two efficient algorithms TKU (mining Top-K Utility itemsets) and TKO (mining Top-K utility itemsets in One phase) are proposed for mining such itemsets without setting minimum utility thresholds.
* TKO is the first one-phase algorithm developed for top-k HUI mining, which integrates the novel strategies RUC, RUZ and EPB to greatly improve its performance.
* Empirical evaluations on different types of real and synthetic datasets show that the proposed algorithms have good scalability on large datasets and the performance of the proposed algorithms is close to the optimal case of the state-of-the art two-phase and one-phase utility mining algorithms.

**SYSTEM REQUIREMENTS:**

**HARDWARE REQUIREMENTS:**

* System                           :         Pentium Dual Core.
* Hard Disk                      :         120 GB.
* Monitor                         :         15’’ LED
* Input Devices                 :         Keyboard, Mouse
* Ram                               :         1GB.

**SOFTWARE REQUIREMENTS:**

* Operating system            :         Windows 7.
* Coding Language           :         JAVA/J2EE
* Tool                               :         Eclipse
* Database                        :         MYSQL