Efficient Algorithms for Mining Top-K High Utility Itemsets

Technofist,

YES Complex, 19/3&4, 2nd Floor, Dinnur Main Road, R.T.Nagar,Bangalore-560032

Ph:080-40969981, Website:www.technofist.com. E-mail:technofist.projects@gmail.com

Abstract:

High utility item sets (HUIs) mining is an emerging topic in data mining, which refers to discovering all item sets 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 item set 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 item sets 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.

Introduction:

FREQUENT itemset mining (FIM) [1, 3, 8, 9, 18, 19, 20, 28, 29] is a fundamental research topic in data mining. However, the traditional FIM 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 [2, 4, 7, 10-18, 22, 23, 25, 26, 27, 29, 34, 35, 36] 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 itemset (HUI) if its utility is no less than a user-specified minimum utility threshold min_ util.