

# TRIPLET-Mine-Tree Based Approach for Mining High Utility Itemsets from Quantitative Databases

Suneetha Merugula<sup>1</sup> and Dr. M. V. P. Chandra Sekhara Rao<sup>2</sup>

<sup>1</sup> Research Scholar, Acharya Nagarjuna University.

Assistant Professor, Department of IT, GMRIT, Rajam, Srikakulam, Andhra Pradesh, India.

<sup>2</sup> Professor, Department of CSE, RVR & JC College of Engineering, Guntur, AP, India.

**Abstract:** High Utility Mining is popular technique in deriving profitable itemsets rather than its occurrence from the quantitative databases. Several algorithms for high utility itemsets were proposed and classified into 2-phase and 1-phase approaches. These approaches use UP-tree and vertical mining algorithms which are based on UP-tree, utility list data structures for mining high utility itemsets. These algorithms suffer from either huge candidate itemsets generation or too many recursive calls during mining. Hybrid Framework based approach which is a combination of tree based and list based approach is proposed to overcome the above limitations. Unfortunately it is required to switch from one algorithm to another. In this paper, we propose a new approach, namely TRIPLET-Mine that uses the proposed data structure named TRIPLET-Tree for keeping the database and deriving high utility itemsets. We present the effectiveness of the proposed approach on standard datasets against Hybrid, list and tree algorithms.

**Keywords:** Utility Mining, Frequent Itemset Mining.

## 1. Introduction

High Utility Mining (HUIM) is introduced to overcome the limitations of Frequent Itemset Mining which derives itemsets that occurs together in database. HUIM derives a set of itemsets that carries utility is not less than the given utility threshold. It became so popular because of the consideration of profit rather than its occurrence. The profit of an itemset is calculated from the product of quantity and profit. HUIM [13, 20, 22] has been chosen as emerging field to address many more issues. It has been extended to the applications like cross marketing [5, 9] analysis in bio-medical field, e-commerce in mobile environment, web click analysis [16, 20].

Several algorithms have been proposed for deriving HUI's and they can be classified into two groups such as 1-phase and 2-phase algorithms. The vertical mining algorithms comes under 1-phase, they maintain TIDSET for each item and perform intersection operation to get the list of transaction where a particular itemset is appeared. Further improving the performance, inverted-list data structure is used to reduce the size of TIDSET. Another popular approach that comes under 1-phase is list based approach, it is proposed by Lin et al [12]. It maintains necessary information in utility list using inverted list and performs intersection operation to get itemsets. Later, viger et al. [6] has proposed approach to improve lin approach. These vertical approaches performance depends on the length of a list and number of intersections that are required during mining procedure. Intersection is widely recognized as costly method in any databases. Zida et al. [28] presented EFIM recursive algorithm which is based on depth first search. Projected databases are constructed for each item and its extensions, and transaction merging is used to avoid recursive calls and reduce the projected databases. In 2-phase approaches, tree based approaches [21] are used to build tree for the given database in first phase and derive itemsets in second phase by projecting database into sub-portions. The main idea of building tree is to avoid candidate generation. Hence these kind of algorithms very fast in deriving high utility itemsets. However, these kind of algorithms suitable when the database can be accommodated in memory.

DawarS et al [3] proposed Hybrid framework that is UFH algorithm for deriving HUI's from the sparse databases. The UFH is integrated with tree based approach UP-Growth+ and inverted-list based FHM. UP-Growth+ approach is two phase approach where UP-tree is constructed in 1-phase and Utility itemsets are derived in 2-phase. However, utility itemsets are computed faster than other approaches. FHM is inverted-list based 1-phase approach where inverted TIDSET are maintained for each item and intersection operations are performed to get the high utility itemsets. Here, itemsets are derived in a single phase in a faster manner. UFH is hybrid integrated approach which is a combination of the features of both UP-Growth+ and FSM. Inverted-list is constructed for each item of the given