

**Selective Database Projections Based Approach for Mining High-Utility Itemsets**

**Abstract:**

High-utility itemset mining (HilIM) is an emerging area of data mining and is widely used. HilIM differs from the frequent itemset mining (FIM), as the latter considers only the frequency factor, whereas the former has been designed to address both quantity and profit factors to reveal the most profitable products. The challenges of generating the HilI include exponential complexity in both time and space. Moreover, the pruning techniques of reducing the search space, which is available in FIM because of their monotonic and anti-monotonic properties, cannot be used in HilIM. In this paper, we propose a novel selective database projection-based HilI mining algorithm (SPHilI-Miner). We introduce an efficient data format, named HilI-RTPL, which is an optimum and compact representation of data requiring low memory. We also propose two novel data structures, viz, selective database projection utility list and Tail-Count list to prune the search space for HilI mining. Selective projections of the database reduce the scanning time of the database making our proposed approach more efficient. It creates unique data instances and new projections for data having less dimensions thereby resulting in faster HilI mining. We also prove upper bounds on the amount of memory consumed by these projections. Experimental comparisons on various benchmark data sets show that the SPHilI-Miner algorithm outperforms the state-of-the-art algorithms in terms of computation time, memory usage, scalability, and candidates generation.