

**Efficient Vertical Mining of High Average-Utility Itemsets based on Novel Upper-Bounds**

**Abstract:**

Mining High Average-Utility Itemsets (HAUIs) in a quantitative database is an extension of the traditional problem of frequent itemset mining, having several practical applications. Discovering HAUIs is more challenging than mining frequent itemsets using the traditional support model since the average-utilities of itemsets do not satisfy the downward-closure property. To design algorithms for mining HAUIs that reduce the search space of itemsets, prior studies have proposed various upper-bounds on the average-utilities of itemsets. However, these algorithms can generate a huge amount of unpromising HAUI candidates, which result in high memory consumption and long runtimes. To address this problem, this paper proposes four tight average-utility upper-bounds, based on a vertical database representation, and three efficient pruning strategies. Furthermore, a novel generic framework for comparing average-utility upper-bounds is presented. Based on these theoretical results, an efficient algorithm named dHAUIM is introduced for mining the complete set of HAUIs. dHAUIM represents the search space and quickly compute upper-bounds using a novel IDUL structure. Extensive experiments show that dHAUIM outperforms three state-of-the-art algorithms for mining HAUIs in terms of runtime on both real-life and synthetic databases. Moreover, results show that the proposed pruning strategies dramatically reduce the number of candidate HAUIs.