# PREFACE

High utility itemsets (HUIs) is a subfield of frequent itemsets mining (FIM) which is one of the fundamental research topics in data mining. Compared to the FIM, utility mining provides more informative and actionable information. HUIs mining discovers itemsets whose utility is not less than user-defined minimum utility threshold. Several studies have been carried out on this topic, which often discovers a very large number of itemsets which reduce not only the efficiency but also the effectiveness of HUIs mining. In order to increase the efficiency and discover more interesting HUIs, constrained-based mining plays an important role. So, we propose an algorithm to discover HUIs with length constraints named EHIL (**E**fficient **H**igh utility **I**temsets mining with **L**ength constraints) to decrease the number of HUIs by removing very small itemsets. EHIL utilizes two new upper bounds for pruning search space. We modify upper-bounds by incorporating length constraints. To reduce the dataset scans, the proposed algorithm uses transaction merging and dataset projection techniques. To speed up the utility counting, an array-based technique is used. The experimental results show that the execution time improvements ranged from a modest five percent to two orders of magnitude across benchmark datasets. The memory usage is up to twenty-eight times less than the state-of-the-art algorithm FHM+.

The existing HUIs mining algorithms can discover all the itemsets satisfying a given minimum utility threshold. It is often difficult for users to set a proper minimum utility threshold. A smaller minimum utility threshold value may produce a huge number of itemsets, whereas a higher one may produce a few itemsets. Specification of minimum utility threshold is difficult and time-consuming. To address these issues, top-k HUIs mining has been defined where $k$ is the number of HUIs to be found. We propose a pattern growth-based efficient algorithm (named TKEH) for finding top-k HUIs. TKEH utilizes transaction merging and dataset projection techniques to reduce the dataset scanning cost. TKEH employs three minimum utility threshold raising strategies. TKEH utilizes *EUCP* and *sup* strategies to prune search space efficiently. We carried out extensive experiments on real datasets. The results show that TKEH outperforms the state-of-the-art algorithm kHMC. Moreover, TKEH always performs better for dense datasets.

Most of the HUIs mining algorithms work only for itemsets with positive utility values. However, in the real-world, items are found with both the positive and the negative utility values. To address this issue, we propose an algorithm named EHIN (**E**fficient **H**igh-utility **I**temsets Mining with **N**egative Utility) to find all HUIs with negative utility. EHIN utilizes two new upper bound named revised sub-tree and revised local utility for pruning. To reduce the dataset scans, the proposed algorithm uses transaction merging and dataset projection techniques. EHIN utilizes various properties and pruning strategies to mine HUIs with the negative utility. The experimental results show that proposed algorithm is up to 28 times faster in execution time and it consumes up to 10 times less memory than the state-of-the-art algorithm FHN. Moreover, a key advantage is that it always performs better for dense datasets.

HUIs mining algorithms incur with the problem of generating a large number of itemsets and most of the generated itemsets are very small in size which degrade mining performance and action-ability. Apart from these problems, most of the algorithms work only with positive utility value. In order to overcome these issues, we propose an algorithm named EHNL (**E**fficient **H**igh utility itemsets mining algorithm with **N**egative utility and **L**ength constraints). Although negative utility and constraint-based mining are commonly seen in real-world applications, mining HUIs with negative utility and length constraints have not yet been proposed in literature. Most of the algorithms suffer from the problem of repeated dataset scanning. To reduce the scanning cost, dataset projection and transaction merging techniques are utilized. To enhance the performance of proposed algorithm, we utilize sub-tree based pruning technique. In order to check the efficiency of utilized techniques, the variation of the proposed algorithm named EHNL(RSUP) and EHNL(TM) are introduced. The experimental results show that variants of algorithm mine the HUIs efficiently.

Traditional HUIs mining algorithms mine a large number of itemsets, but most of the mined itemsets are redundant. In order to overcome this issue, closed HUIs mining algorithms have been proposed which avoids duplicate itemsets. However, the state-of-the-art closed HUIs mining algorithms work only with positive utility value. Mining closed HUIs mining algorithm with negative utility has not yet been proposed, although negative utility mining is commonly seen in the real-world applications. To address this issue, we propose an efficient algorithm named CHN (**C**losed **H**igh utility itemsets mining with **N**egative utility). To our best knowledge, this is the first work

addressing the issue of closed HUIs with negative utility value. It relies on pattern-growth approach. It utilizes transaction merging and dataset projection techniques to reduce the dataset scanning cost. In order to enhance the performance of proposed algorithm, we edit and utilize sub-tree based pruning technique. Moreover, an array-based utility counting technique is also utilized to speed up the utility counting process. A bi-directional extension technique is also utilized to check the closure and prune the search space. In order to check the efficiency of utilized techniques, variation of the proposed algorithms named CHN(RSU-Prune) and CHN(TM) are introduced. The experimental results on dense and sparse datasets show that the variants of algorithms mine the closed HUIs efficiently.