

io-port 06107034

Lin, Ming-Yen; Tu, Tzer-Fu; Hsueh, Sue-Chen

**High utility pattern mining using the maximal itemset property and lexicographic tree structures.**

Inf. Sci. 215, 1-14 (2012).

Summary: The problem of high utility mining is discovering all of the high utility itemsets in a transactional database. Most algorithms find high utility itemsets in two steps. The first step identifies all of the potential itemsets. The second step then determines the high utility itemsets from the set of potential itemsets. The large number of potential itemsets in the first step is generally the mining bottleneck. If we can reduce the number of potential itemsets, the mining performance can be improved significantly. In this paper, we use a maximal itemset property and propose an algorithm called UMMI (high Utility Mining using the Maximal Itemset property) to significantly reduce the number of potential itemsets in the first step. In the second step, UMMI uses an effective lexicographic tree structure to determine all of the high utility itemsets. In general, UMMI outperforms all three of the previously used algorithms, including CTU-PRO, an optimized TWU-mining algorithm, and Two-Phase, in our experiments using synthetic datasets. On average, UMMI is 5, 3, and 7 times faster than CTU-PRO, TWU-mining, and Two-Phase, respectively. In a real data experiment, UMMI is 6 times faster than Two-Phase. The other two algorithms are not capable of completing the mining step in a reasonable amount of time. UMMI uses an approximately fixed amount of memory, which is generally less than the other algorithms for each mining. The experimental results show that the proposed algorithm can mine the high utility itemsets efficiently. In addition, UMMI is linearly scalable with respect to the number of transactions.

*Keywords:* high utility mining; maximal itemset mining; data mining; frequent pattern; lexicographic tree structure

doi:10.1016/j.ins.2012.05.015