

**Frequent Itemsets Mining with Differential Privacy over Large-scale Data**

**Abstract:**

Frequent itemsets mining with differential privacy refers to the problem of mining all frequent itemsets whose supports are above a given threshold in a given transactional dataset, with the constraint that the mined results should not break the privacy of any single transaction. Current solutions for this problem cannot well balance efficiency, privacy, and data utility over large-scale data. Toward this end, we propose an efficient, differential private frequent itemsets mining algorithm over large-scale data. Based on the ideas of sampling and transaction truncation using length constraints, our algorithm reduces the computation intensity, reduces mining sensitivity, and thus improves data utility given a fixed privacy budget. Experimental results show that our algorithm achieves better performance than prior approaches on multiple datasets.