Conquering the Search Space for the Calculation of the Maximal Frequent Set

Abstract: The most time consuming operation in data mining applications is the computation of frequent itemsets. Since the search space is exponential, efficient pruning is necessary. On the other hand, data mining on large data volumes makes the coupling of mining with database systems increasingly important. In this paper, we propose a new operator to efficiently calculate the support of candidate itemsets within the database engine. Based on this operator, we propose a novel approach to reduce search complexity by combining top-down with bottom up pruning in order to obtain an algorithmic complexity that is only proportional to the volume of the maximal frequent itemsets. In contrast to other approaches, this strategy avoids expensive database scans as well as intermediate result materialization.

Stichworte: Data Mining; Frequent Itemset Generation; Search Space

Jahr: 2000

Jahr / Monat: 2000-09-01 00:00:00

Seiten: 16

Occurences:

- Einrichtungen > Fakultäten > Fakultät für Informatik > Technische Berichte > 2000