Ludwig-Maximilians-Universität München
Institut für Informatik
Prof. Dr. Thomas Seidl
Julian Busch, Evgeniy Faerman,
Florian Richter, Klaus Schmid

Knowledge Discovery in Databases
SS 2016

Exercise 3: Frequent Itemset Mining

Regarding tutorials on 11.05.-13.05.2016.

Exercise 3-1   Frequent Itemsets

The Apriori-algorithm makes use of prior knowledge of subset support properties. Let I be the set of all items. Give proofs or counterexamples for the following claims:

(a) Let $S \subseteq I$ be a frequent itemset. Then every non-empty subset $S' \subseteq S$ must also be frequent.

(b) Let $S \subseteq I$ be an arbitrary itemset. Then $support(S') \geq support(S)$ holds for any non-empty subset $S' \subseteq S$.

Exercise 3-2   Frequent Itemset Mining

Let D be a database that contains the following four transactions.

TID	items_bought
T1	$\{K, A, D, B\}$
T2	$\{D, A, C, E, B\}$
T3	$\{C, A, B, E\}$
T4	$\{B, A, D\}$

In addition let *min_sup = 60%*.

(a) Find all frequent itemsets using the Apriori algorithm.

(b) Find all frequent itemsets using the FP-growth algorithm.

(c) Determine all closed and maximal frequent itemsets.

Exercise 3-3   Association Rule Mining

After frequent itemset mining, association rules can be extracted as follows: For each frequent itemset X and every non-empty subset $Y \subset X$, generate a rule $Y \Rightarrow X \setminus Y$ if it fulfills the minimum confidence property.

(a) Proof the following anti-monotonicity lemma for strong association rules:

Let X be a frequent itemset and $Y \subset X$. If $Y \Rightarrow X \setminus Y$ is a strong association rule, then $Y' \Rightarrow X \setminus Y'$ is also a strong association rule for every $Y \subseteq Y'$.

(b) Extract all strong association rules from the database D provided in the previous exercise with a minimum confidence of *min_conf = 80%*. Which candidate rules can be pruned based on anti-monotonicity?