

Knowledge Discovery in Databases
WS 2017/18

Übungsblatt 4: Frequent Itemset Mining

Besprechung: 30.11. und 01.12.2017

Aufgabe 4-1 Frequent Itemset Mining

Sei D eine Datenbank, die die folgenden vier Transaktionen enthält:

TID	gekaufte Items
T1	{K, A, D, B}
T2	{D, A, C, E, B}
T3	{C, A, B, E}
T4	{B, A, D}

Außerdem sei $min_sup = 60\%$.

- (a) Finden Sie alle frequent itemsets unter Verwendung des FP-growth Algorithmus.
- (b) Bestimmen Sie alle geschlossenen und maximalen frequent itemsets.

Aufgabe 4-2 Assoziationsregeln

Assoziationsregeln können wie folgt hergeleitet werden: Für jedes frequent itemset X und jede nicht-leere Teilmenge $Y \subset X$, erstelle eine Regel $Y \Rightarrow X \setminus Y$ wenn sie die minimale confidence erfüllt.

- (a) Beweisen Sie das folgende Anti-Monotonie Lemma für starke Assoziationsregeln:
Sei X ein frequent itemset und $Y \subset X$. Wenn $Y \Rightarrow X \setminus Y$ eine starke Assoziationsregel ist, dann ist $Y' \Rightarrow X \setminus Y'$ auch eine starke Assoziationsregel für alle $Y \subseteq Y' \subseteq X$.
- (b) Leiten Sie alle starken Assoziationsregeln aus der Datenbank D aus der vorherigen Aufgabe her, die eine minimale Confidence von $min_conf = 80\%$ und $min_sup=60\%$ haben. Welche Kandidaten für eine Regel können auf Grund der Anti-Monotonie gepruned werden?