

7. Übung „Knowledge Discovery“

Sommersemester 2006

1 Hashbaum

1. Konstruieren Sie einen Hashbaum für die folgenden Itemsets für $k = 3$ und maximal 2 Transaktionen pro Knoten.
(1,4,5), (2,3,4), (3,6,8), (1,2,5), (4,5,8), (3,4,5), (1,2,4), (1,3,6), (3,5,6), (4,5,7), (3,5,7), (1,5,9), (6,8,9), (3,6,7), (5,6,7)
2. Welche Knoten müssen besucht werden, wenn die 3-Item Teilmengen von (1,3,7,8,9) gefunden werden sollen?

2 Hierarchische Assoziationsregeln

1. Seien A, A_1, A_2, B Items und A eine Verallgemeinerung von A_1 und A_2 . Beweisen oder widerlegen Sie (durch Angabe eines Gegenbeispiels) die folgenden Behauptungen:
 - a) $\text{support}(A \rightarrow B) = \text{support}(A_1 \rightarrow B) + \text{support}(A_2 \rightarrow B)$
 - b) Falls $\text{support}(A_1 \rightarrow B) > \text{minsupport}$, dann ist $\text{support}(A \rightarrow B) > \text{minsupport}$.
 - c) Falls $\text{support}(A \rightarrow B) > \text{minsupport}$, dann ist $\text{support}(A_1 \rightarrow B) > \text{minsupport}$.
2. Wie berechnet sich die erwartete Konfidenz einer Regel $X \rightarrow Y$?
3. Gegeben folgende Menge von Transaktionen:

TID	Items
1	D,F,G
2	C,D,E,G
3	A,D,E
4	C,D,H
5	B,C,D,F,G
6	C,F,G
7	C,D,F
8	E,G

- a) Geben Sie für einen Support von 25% die vom Apriori-Algorithmus generierten Mengen C_k und L_k an.

- b) Bestimmen Sie Support und Konfidenz der Assoziationsregel $DF \rightarrow G$.
- c) Nehmen Sie an, dass die Items E, F und G zum Item X generalisiert sind. Bestimmen Sie den Support von $X \rightarrow C$. Geben Sie an, ob $X \rightarrow C$ und $E \rightarrow C$ R -interessant mit $R = 2$ sind. Begründen Sie ihre Antworten.

3 Quantitative Assoziationsregeln

1. Erweitern Sie den Apriori-Algorithmus so, dass häufig auftretende quantitative Itemsets berechnet werden können. Passen Sie dazu die Funktion `subset` an, so dass zunächst alle Kandidaten in Gruppen mit gleichen (kategorischen) Attributen eingeteilt werden. Finden Sie dann einen effizienten Weg, nach geeigneten Kandidaten zu suchen.