

7. Übung „Knowledge Discovery“

Wintersemester 2008/2009

1 Häufige Itemmengen

Beweisen Sie die Korrektheit der folgenden zur Regelgenerierung verwendeten Abhängigkeit für häufige Itemmengen X und deren Teilmengen $\hat{X} \subset X$. Verwenden Sie dazu die textuelle Beschreibung der Konfidenz im Kapitel 4.2 auf Folie 6.

$$\text{confidence} \left((X \setminus \hat{X}) \rightarrow \hat{X} \right) = \frac{\text{support}(X)}{\text{support}(X \setminus \hat{X})}$$

2 Assoziationsregeln

Gegeben sei die folgende Menge D von Warenkorbdaten (Transaktionen):

Transaktion	Items
t_1	Windeln, Bier, Chips
t_2	Chips, TV-Zeitschrift
t_3	TV-Zeitschrift, Bier, Chips
t_4	Bier, Windeln, Zahnpasta
t_5	Zahnpasta, Chips
t_6	TV-Zeitschrift, Chips, Bier
t_7	Bier, Windeln
t_8	TV-Zeitschrift, Chips

1. Bestimmen Sie zu den gegebenen Transaktionen die häufigen Itemmengen, die einen Mindestsupport von 25% aufweisen. Gehen Sie bei der Bestimmung der häufigen Itemmengen nach dem aus der Vorlesung bekannten Algorithmus vor.
2. Bestimmen Sie nun aus den berechneten Itemmengen alle Assoziationsregeln mit einer Mindestkonfidenz von 66%.
3. Wie können die erzielten Ergebnisse interpretiert werden? Wie sähe die Interpretation aus, wenn statt dessen 20000 Transaktionen als Grundlage des Assoziationsregel-Mining gedient hätten?

4. Ein weiteres Maß für Regeln ist der sogenannte *Lift* einer Regel, welcher definiert ist durch

$$\text{lift}(X \rightarrow Y) = \frac{\text{confidence}(X \rightarrow Y)}{\text{support}(Y)}$$

Was besagt der Lift einer Regel?

3 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 Blattknoten müssen besucht werden, wenn die 3-Item Teilmengen von (1,3,7,8,9) gefunden werden sollen?