

Knowledge Discovery

Übungsblatt 5

Sommersemester 2005

Aufgabe 1

- a) Beweisen Sie die Korrektheit der folgenden zur Regelgenerierung verwendeten Abhängigkeit für häufige Itemmengen X und deren Teilmengen $X' \subset X$:

$$\text{confidence}((X - X') \longrightarrow X') = \frac{\text{support}(X)}{\text{support}(X - X')}$$

Aufgabe 2

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

- a) 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.
- b) Bestimmen Sie nun aus den berechneten Itemmengen alle Assoziationsregeln mit einer Mindestkonfidenz von 66%.
- c) 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?

- d) Wie können die erzielten Ergebnisse interpretiert werden? Wie sähe eine Interpretation aus, wenn statt dessen 20.000 Transaktionen als Grundlage des Association-Mining gedient hätten?