

Übungen zur Vorlesung  
**Wissensentdeckung in Datenbanken**  
Sommersemester 2008

Blatt 4

**Wiederholung** Als Vorbereitung auf die folgenden beiden Aufgaben, sollten Sie sich die Vorlesungsinhalte bezüglich der Assoziationsregeln (Apriori-Algorithmus) und der freien Mengen noch einmal anschauen. Dabei sollten Sie folgende Fragestellungen in den Übungen beantworten können<sup>1</sup>:

- (a) Skizzieren Sie den Apriori-Algorithmus kurz und beschreiben Sie das Prinzip der Kandidatengenerierung.
- (b) Welche Schlüsseigenschaften der häufigen Mengen liegen dem Apriori-Algorithmus zu Grunde? Können Sie diese an einem Beispiel darstellen?
- (c) Was geschieht beim Ansatz, häufige Mengen über *free sets* und *closed sets* zu finden? Geben Sie kurz die grundlegende Idee wieder.

*Hinweis:* Interessierte Leser finden die in der Vorlesung vorgestellte Arbeit von J.-F. Boulicaut mit dem Titel "*Free-Sets: A Condensed Representation of Boolean Data for The Approximation of Frequency Queries*" auf der Web-Seite mit den Übungsblättern:

<http://www-ai.cs.uni-dortmund.de/LEHRE/VORLESUNGEN/KDD/SS08/index.html>

**Aufgabe 4.1**

Gegeben sei die rechts abgebildete Transaktionstabelle und ein minimaler Support von 0.1.

- (a) Bestimmen Sie die 1-freien und 2-freien Mengen.
- (b) Bestimmen Sie die *closed item sets*.

A	B	C	D
0	1	1	0
1	1	0	0
1	1	1	1
0	1	1	1
0	1	1	1
1	1	0	1
0	1	1	1
0	1	0	0
1	1	0	0
0	1	1	1
0	1	1	1

---

<sup>1</sup>Keine Angst - es wird von niemandem in der Übung verlangt, sondern ist lediglich als kleine Hilfe für Sie gedacht.

### Aufgabe 4.2

Die Saison eines Fußballvereins wird anhand folgender Tabelle dargestellt. Für jeden Spieltag wird einerseits die Stimmung in der Mannschaft und das resultierende Ergebnis des Spieltages angegeben.

Spieltag	Stimmung	Ergebnis	Spieltag	Stimmung	Ergebnis
1	neutral	Sieg	18	gut	Unentschieden
2	gut	Sieg	19	neutral	Niederlage
3	gut	Unentschieden	20	schlecht	Niederlage
4	gut	Unentschieden	21	schlecht	Unentschieden
5	neutral	Unentschieden	22	neutral	Unentschieden
6	neutral	Unentschieden	23	neutral	Sieg
7	neutral	Niederlage	24	neutral	Sieg
8	schlecht	Niederlage	25	gut	Sieg
9	schlecht	Sieg	26	gut	Sieg
10	gut	Sieg	27	gut	Unentschieden
11	gut	Unentschieden	28	gut	Unentschieden
12	neutral	Sieg	29	gut	Sieg
13	gut	Sieg	30	gut	Unentschieden
14	gut	Niederlage	31	gut	Niederlage
15	neutral	Unentschieden	32	gut	Niederlage
16	neutral	Unentschieden	33	neutral	Niederlage
17	neutral	Sieg	34	schlecht	Unentschieden

Überführen Sie die Tabelle in die Repräsentation zeitlicher Daten, die in Mannilas WINEPI-Ansatz verwendet wird, d.h. geben Sie die Beobachtungssequenz  $s$  an. Als Ereignis gilt dabei das **Eintreten** eines neuen Stimmungs- **oder** Ergebnisverhältnisses, also können zu einem Zeitpunkt null bis zwei Ereignisse stattfinden.

- Formalisieren Sie die Episoden *“Auf ein Unentschieden folgt neutrale Stimmung, danach folgt eine Niederlage”* und *“Niederlagen entstehen bei guter Stimmung”*.
- Geben Sie die diesem Ansatz entsprechende Fensterung von  $s$  an bei einer Fensterbreite von 6 Spieltagen mit einer Schrittweite von 2 Spieltagen.
- Bestimmen Sie die Häufigkeit der beiden Episoden anhand dieser Fensterung. Bestimmen Sie zusätzlich die Konfidenz der Regel *“Niederlagen entstehen bei guter Stimmung”*.