

### Organisatorische Regelungen für die Prüfungszulassung

Zur Vorlesung gibt es Übungsblätter mit Aufgaben. Die Aufgaben werden am Anfang der zugehörigen Übung *votiert*. Durch eine Votierung erklärt man sich bereit, zu einer Aufgabe vorzutragen. (Lösungsvorschläge werden diskutiert, sie müssen nicht gleich richtig sein.) Zur Prüfung wird zugelassen, wer

1. *mindestens* die Hälfte der Übungsaufgaben votiert *und*
2. *mindestens* zweimal in den Übungen vorgetragen hat.

### 1. Übungsblatt

#### Aufgabe 1 Nutzen häufiger Muster

- a) Was versteht man unter “Warenkorbanalyse” (*market basket analysis*)?
- b) Wie könnte man Informationen in Form von häufigen Mustern oder Assoziationsregeln für Marketingzwecke nutzen?
- c) Was sind mögliche Anwendungen des Findens häufiger Itemmengen?  
Versuchen Sie, über die Motivationsfolie der Vorlesung hinauszugehen!
- d) Macht seine Häufigkeit ein Muster oder eine Regel bereits interessant?  
Wenn nicht, warum nicht? Versuchen Sie, Gegenbeispiele zu finden!
- e) Welche Vorteile hat die Darstellung häufiger Muster als Assoziationsregeln im Vergleich zu häufigen Itemmengen? Welche Nachteile hat sie?

#### Exercise 2 Grundbegriffe

- a) Warum werden Transaktionsdatenbanken als Vektoren/Arrays dargestellt?  
Warum können wir nicht einfache Mengen verwenden? Was sind alternative Darstellungen? Was sind die Vor- und Nachteile alternativer Darstellungen?
- b) Warum ist die Itembasis oft nicht explizit gegeben? Wie kann man sie bestimmen?
- c) Wie können absoluter und relativer (minimaler) Support ineinander umgerechnet werden? Welche zusätzliche Information braucht man für die Umrechnung?
- d) Beim Finden häufiger Itemmengen betrachten wir nur, ob ein Item in einer Transaktion enthalten ist oder nicht. In Warenkorbdaten haben die Items aber Vielfachheiten (ein Kunde kauft 3 Packungen Milch, 5 Äpfel, 2 Dosen Suppe etc.).  
Wie kann man damit umgehen? Kann dies durch Vorverarbeitung behandelt werden? Wenn ja, wie?

#### Exercise 3 Eigenschaften des Support

- a) Was ist die *Überdeckung* (*cover*) eines Items oder einer Itemmenge?  
In welcher Beziehung steht sie zum *Support* eines Items oder einer Itemmenge?

- b) Was bedeutet es, daß der Support anti-monoton oder abwärts geschlossen ist? Wie kann man diese Eigenschaft als Formel ausdrücken?
- c) Welche Eigenschaft der Item-Abdeckungen entspricht der Anti-Monotonie?
- d) Was ist die *Apriori-Eigenschaft*? Wie wird sie formal angegeben? Wie folgt sie? Was sagt die Kontraposition der Apriori-Eigenschaft? Was legt sie nahe?

**Exercise 4**    Eigenschaften des Support

- a) Zeigen Sie daß der Support einer Itemmenge und einer ihrer Obermengen gleich sein kann! (D.h., der Support der Obermenge muß nicht echt kleiner sein.)  
Wie kann man die Bedingung beschreiben, unter der dies gilt?  
Kann es mehrere verschiedene Obermengen mit dem gleichen Support geben?
- b) Sei  $I$  eine Itemmenge und  $a$  und  $b$  zwei Items, die nicht in  $I$  enthalten sind.  
Sei  $s(I) = s(I \cup \{a\}) = s(I \cup \{b\})$ . Was ist der Support der Menge  $I \cup \{a, b\}$ ?
- c) Finden Sie ein Beispiel, in dem man, ohne die Transaktionsdatenbank anzuschauen, aus dem Support von (mehreren) Itemmengen  $I_k$  ableiten kann, da  $\bigcup_k I_k$  häufig sein muß bzgl.  $s_{\min} \geq 3$  (Sie dürfen allerdings die Größe  $n$  der Datenbank verwenden). Benutzen Sie *nicht* die Lösung/die Annahmen aus b)!