

Maschinelles Lernen und Data Mining

Übungsblatt für den 9. 2. .2006

Aufgabe 1

Gegeben sei folgender String: ADFASASDASFASDFSDFAS

1. Finden Sie mit einem Apriori-ähnlichen Algorithmus alle Teilstrings, die mindestens 2x vorkommen.
2. Bilden Sie aus den gefundenen Subsequenzen mit $\text{minimum support} = 2$ alle Regeln mit $\text{minimum confidence} = 0.5$

Aufgabe 2

Ein on-line Buchgeschäft möchte eine Datenbank mit 10,000 Kunden analysieren, die jeweils eines oder mehrere von 500 verschiedenen Büchern gekauft haben. Zur Entdeckung von Assoziationsregeln wird der Algorithmus Apriori mit einem Minimum Support von 3% und einer minimalen Konfidenz von 75% verwendet.

1. Es wird festgestellt, daß die beiden häufigsten Verkäufe "Harry Potter und der Stein der Weisen" (HP1) und "Harry Potter und die Kammer des Schreckens" (HP2) sind. HP1 wurde von 6,000 Kunden und HP2 von 8,000 Kunden gekauft. 4,000 Kunden kauften beide Bücher.

Welche der beiden Assoziationsregeln findet sich im Output des Assoziationsregellerners?

- HP1 \rightarrow HP2
- HP2 \rightarrow HP1
- beide
- keine von beiden

Geben Sie Support und Konfidenz für beide Regeln an.

2. Wenn man annimmt, daß alle Kunden, die beide Bücher gekauft haben, zuerst HP1 und später HP2 gekauft haben: Wie interpretieren Sie den Einfluß des Kaufs von HP1 auf den Kauf von HP2?
3. Die längste Assoziationsregel, die gefunden wurde, wurde aus einem Itemset der Größe 20 konstruiert. Geben Sie eine möglichst große untere Schranke für die Anzahl der gefundenen Frequent Itemsets an.