

Übung zur Vorlesung „Data Mining“ im Sommersemester 2015

Übungsblatt 1

Aufgabe 1 (*Entdeckung von Frequent Itemsets mit dem Apriori-Algorithmus*)

Gegeben sei folgende Datenbank mit fünf Transaktionen:

Nr.	Transaktion
1	{ S, O, M, M, E, R }
2	{ S, O, N, N, E }
3	{ S, O, N, N, E, N, B, R, A, N, D }
4	{ S, T, R, A, N, D }
5	{ M, E, E, R }

- Jede Transaktion ist eine *Menge* von Items – und keine *Multimenge*. Sorgen Sie dafür, dass jede Transaktion ein Item nur höchstens einmal enthält. Sortieren Sie außerdem die Items jeder Transaktion alphabetisch. Geben Sie die „aufgeräumte“ Datenbank an.
- Wieviele sinnvolle Assoziationsregeln sind für diese Datenbank maximal möglich (gemäß Vorlesung; inklusive Regeln mit Support = 0)?
- Sei der Mindest-Schwellwert für den Support $\text{minsup} > 0$. Wieviele Items kann ein Frequent Itemset für die angegebene Datenbank maximal beinhalten?
- Für den Support sei der Mindest-Schwellwert $\text{minsup} = 0.6$ gegeben.
Führen Sie nun den Apriori-Algorithmus gemäß Vorlesung durch um alle Frequent Itemsets zu finden. Geben Sie für jeden Iterationsschritt k die Menge der Frequent Items F_k und die Menge der Kandidaten C_k an. Die Kandidatenmenge soll nicht über ein Kartesisches Produkt erzeugt werden, sondern über einen „Join“-Schritt mit anschließendem Pruning (wie in der Vorlesung angegeben).
- Wie hoch sind hier Confidence und Support der Assoziationsregel „ $\{O\} \Rightarrow \{S, E\}$ “?
- Wenn allgemein die Menge F_1 im Apriori-Algorithmus n Elemente hat, wieviele Elemente hat dann die Kandidatenmenge C_2 ?

Aufgabe 2 (*Maximal & Closed Frequent Itemsets*)

- Nennen Sie (mit Hilfe der Ergebnisse aus Aufgabe 1) alle Maximal Frequent Itemsets und alle Closed Frequent Itemsets.
- Wenn alle Maximal Frequent Itemsets inklusive Support bekannt sind, ist dann die Menge *aller* Frequent Itemsets samt zugehöriger Supportwerte rekonstruierbar? Wie verhält es sich mit Closed statt Maximal Frequent Itemsets?

Abgaben zu Beginn der nächsten Vorlesung am Mittwoch, **29.04.2015**.