Data Mining 2: Finden häufiger Muster Sommer 2018
Christian Borgelt ab 11.06.2018

6. Übungsblatt

Aufgabe 22 Maximale und geschlossene Itemmengen

a) Finden Sie die häufigen / maximalen / geschlossenen Itemmengen für die rechts gezeigte Transaktionsdatenbank und einen minimalen Support $s_{\min} = 3$ (Sie können einen konkreten Algorithmus verwenden oder auch nicht):

$a\ d\ f$	$a\ b\ d$
$a\ c\ d\ e$	$b\ d\ e$
$b\ d$	$b\ c\ e\ g$
$b\ c\ d$	$c\ d\ f$
$b\ c$	$a\ b\ d$

b) Finden Sie ein Beispiel einer (kleinen) Transaktionsdatenbank für die die Anzahl der maximalen Itemmengen abnimmt, wenn der minimale Support verringert wird; oder erklären Sie auf andere Weise, warum dies möglich ist!

c) Kann die Anzahl aller häufigen Itemmengen oder die Anzahl aller geschlossenen Itemmengen abnehmen, wenn der minimale Support verringert wird? Wenn ja, geben Sie ein Beispiel an! Wenn nein, zeigen Sie, warum dies nicht möglich ist!

d) Wie stehen maximale und geschlossene Itemmengen zueinander in Beziehung? Kann die eine Menge in Begriffen der anderen ausgedrückt werden? Welche Informationen erhalten/verlieren maximale/geschlossene Itemmengen? Wie?

e) Definieren Sie geschlossene Itemmengen in Begriffen perfekter Erweiterungen!

Aufgabe 23 Maximale und geschlossene Itemmengen

a) Charakterisieren Sie die Menge $M_T(1)$ für beliebige Transaktionsdatenbanken T (d.h., gültig für alle möglichen Transaktionsdatenbanken T).

b) Angenommen, es gilt $\forall s_{\min} : F_T(s_{\min}) = C_T(s_{\min}) = M_T(s_{\min})$.
Wie sieht die Transaktionsdatenbank T aus?

c) Angenommen, es gilt nur $\forall s_{\min} : C_T(s_{\min}) = M_T(s_{\min})$.
Wie sieht die Transaktionsdatenbank T aus?

d) Angenommen, es gilt nur $\forall s_{\min} : C_T(s_{\min}) - \{\emptyset\} = M_T(s_{\min}) - \{\emptyset\}$.
Wie sieht die Transaktionsdatenbank T aus?

Aufgabe 24 Finden geschlossener Itemmengen

a) Die Menge der geschlossenen (häufigen) Itemmengen kann definiert werden als

$$C_T(s_{\min}) = \{I \subseteq B \mid s_T(I) \geq s_{\min} \wedge \forall J \supset I : s_T(I) > s_T(J)\}.$$

Wie kann man geschlossene Itemmengen alternativ mit Hilfe eines Abschlußoperators definieren? Warum sind diese beiden Definitionen äquivalent?

b) Wie verhalten sich geschlossene Itemmengen und geschlossene Mengen von Transaktionsbezeichnern/-indizes zueinander? Was bedeutet es für eine Menge von Transaktionsbezeichnern, daß sie geschlossen ist?

c) Welche Möglichkeit der Suche nach geschlossenen (häufigen) Itemmengen legt die Beziehung aus b) nahe?