Data Mining 2: Frequent Pattern Mining Summer 2018
Christian Borgelt from 11.06.2018

Exercise Sheet 6

Exercise 22 Maximal and Closed Item Sets

a) Find the frequent / maximal / closed item sets for the transaction database shown on the right and a minimum support $s_{\min} = 3$ (you may or may not use some concrete algorithm):

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

b) Find an example of a (small) transaction database for which the number of maximal item sets goes down if the minimum support is reduced; or explain in some other way why this may happen!

c) Is it possible that the number of all frequent item sets or the number of closed frequent item sets goes down if the minimum support is reduced? If yes, give an example! If no, argue why this is not possible!

d) How are maximal and closed item sets related to each other?
Can the one be expressed in terms of the other?
What information is preserved/lost with maximal/closed item sets? How?

e) Define closed item sets with the help of the notion of a perfect extension!

Exercise 23 Maximal and Closed Item Sets

a) Characterize the set $M_T(1)$ for an arbitrary transaction database T (that is, your characterization should work for all possible transaction databases T).

b) Suppose we have $\forall s_{\min} : F_T(s_{\min}) = C_T(s_{\min}) = M_T(s_{\min})$.
What does the transaction database T look like?

c) Suppose we have only $\forall s_{\min} : C_T(s_{\min}) = M_T(s_{\min})$.
What does the transaction database T look like?

d) Suppose we have only $\forall s_{\min} : C_T(s_{\min}) - \{\emptyset\} = M_T(s_{\min}) - \{\emptyset\}$.
What does the transaction database T look like?

Exercise 24 Finding Closed Item Sets

a) The set of closed (frequent) item sets may be defined as

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

How may closed (frequent) item sets alternatively be defined with the help of a closure operator? Why are these two definitions equivalent?

b) How are closed item sets and closed sets of transaction identifiers/indices related?
What does it mean for a transaction index set that it is closed?

c) Which possibility of searching for closed (frequent) item sets is suggested by the relationship in b)?