

### Exercise Sheet 4

#### Exercise 21 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 22 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 23 Finding Closed Item Sets

- a) How are closed item sets and closed sets of transaction identifiers related?  
What does it mean for a transaction index set that it is closed?
- b) How does the Carpenter algorithm exploit this relationship?
- c) What recursive relationship for closed item sets is the IsTa algorithm based on?
- d) Suppose the divide-and-conquer scheme for finding frequent item sets is enhanced by perfect extension pruning, so that subproblems are described by triplets  $S = (T, P, X)$ . Is  $P \cup X$  always a closed item set? Is it always a maximal item set?

### Exercise 24 Evaluating Item Sets

- a) Find some other measure(s) on the partial order  $(2^B, \subseteq)$  that is/are anti-monotone (or downward closed)!
- b) Show that the quotient of observed and expected frequency of an item set (expected w.r.t. an assumption of full independence) is not anti-monotone! Try to find counterexamples with  $|I| \geq 1$  and with  $|I| \geq 2$ . Show that the reciprocal value of this quotient is also not anti-monotone!
- c) Show that if an item set  $I$  and an item  $a$  are independent,  $\varrho_{\text{fi}}(I \cup \{a\}) = \varrho_{\text{fi}}(I)$ . Show that, in contrast to this,  $\varrho_{\text{ii}}(I \cup \{a\}) = 1$  in this case!
- d) Find an example where  $\varrho_{\text{si}}$  scores an item set  $I$  high, even though  $I$  is composed of two independent subsets!

### Exercise 25 Association Rules

- a) What are association rules? With what measures are they evaluated?  
What is the support of an association rule (two versions)?  
What is the confidence of an association rule?
- b) How are association rules induced? What steps are needed?  
How is the induction related to frequent item set mining?
- c) With what minimum support do we have to find frequent item sets for association rule induction? Is it the minimum support of association rules? Or does it depend on the choice of the rule support definition? If yes, how?
- d) What relationships hold between the confidence values of different association rules formed from the same item set? How can we exploit these relationships in the generation of association rules?

### Exercise 26 Association Rules

- a) What are some measures other than minimum support and minimum confidence with which association rules may be evaluated? From what values are they usually computed?
- b) What is the lift (value) of an association rule and how is it defined?  
What is the information gain of an association rule?  
What is the  $\chi^2$ -measure of an association rule?
- c) Are association rules with more than one item in the consequent useful?  
What additional information do they provide?