

Business Analytics

Exercise Sheet 9

Josif Grabocka (josif@ismll.de)
Information Systems and Machine Learning Lab (ISMLL)
Universität Hildesheim

02 July 2013
Submission Until 14 July 2013 23:59

NOTE: NEXT TUTORIAL ON TUESDAY, 16/07/2013

Question 1: Frequent Itemset Mining (5 points)

- (a) What is the pruning criteria of the apriori breadth-first search algorithm with respect to the support values of supersets? **(1 point)**
- (b) Consider the transactions dataset shown in Table 1, which is recorded in an imaginary local store, which sells only four products, namely bread, beer, bratwurst and grunkohl. Find the frequent item sets having a minimum support value of three, by using the NAIVE breadth-first search techniques. **(2 points)**
- (c) Repeat the computations of Section 1.b) with the pruning trick of Section 1.a). How many support calculations do you spare? **(2 points)**

Question 2: Frequent Itemset Mining (Programming) (5 points)

- (a) What is the confidence of an association rule? Mention the aspect that confidence assures. **(1 point)**
- (b) Implement the apriori breadth-first search algorithm with support value pruning and briefly describe your source code. **(2 points)**
- (c) The file name "retail.txt" contains a list of 88162 purchase transactions from a real supermarket, where products are represented by their id (Brijs et al. [1999]). Each row is a list of product ids and represent a single transaction. Run your implementation of Section 2.b) and report the frequent item sets for varying values of minimum support, ranging from 5 to 15. **(2 points)**

Table 1: Local Store - Transaction Database

Customer ID	Product 1	Product 2	Product 3	Product 4	Transaction Time
Customer 1	Bratwurst	Beer			02/07/13 19:04:21
Customer 2	Bread	Bratwurst			02/07/13 11:59:03
Customer 3	Bratwurst	Grunkohl	Beer		02/07/13 12:01:27
Customer 4	Bread	Bratwurst	Grunkohl	Beer	02/07/13 10:34:41
Customer 5	Bread	Bratwurst	Beer		02/07/13 14:01:59
Customer 6	Grunkohl	Beer			02/07/13 18:53:10
Customer 7	Bratwurst	Grunkohl			02/07/13 12:20:46

Submission

- Electronically to josif@ismll.de
- Email title must be *BA2013-NAME-Tutorial-NO*, e.g.: *BA2013- JosifGrabocka-Tutorial-1*
- Report file must be a PDF with file name like the email title *BA2013-NAME-Tutorial-NO.pdf*
- Source codes and other materials must be a ZIP with file name like the email title.

References

Tom Brijs, Gilbert Swinnen, Koen Vanhoof, and Geert Wets. Using association rules for product assortment decisions: A case study. pages 254–260, 1999.