Lab: Association Rules

Evgueni N. Smirnov
smirnov@maastrichtuniversity.nl

June 19, 2019

1. Introduction

In this lab we consider a market basket problem. To solve that problem we use an implementation of the `Apriori` algorithm provided in `Weka`. We note that this implementation uses attribute-value representation of items and hence it can produce a bit different association rules (compared with the standard `Apriori` algorithm).

2. Market Basket Problem

We have:

- a set I of 11 items: *fruitveg, freshmeat, dairy, cannedveg, cannedmeat, frozenmeal, beer, wine, softdrink, fish, confectionery*; and
- a data of 1000 transactions T ($T \subseteq I$) associated with different customers from a super market.

We need to find interesting association rules that explain customer behavior. The data is provided in the file `marketBasket.arff`.

The main tasks are as follows:

A. Study the computational complexity of the `Apriori` algorithm. For this purpose set the minimal confidence to 1.0 and plot the frequency distribution of the frequent items sets for minimal support of 0.001, 0.01, 0.1 and 0.2. Based on the plots explain the relationship between the minimal support and the computational complexity of the `Apriori` algorithm.

B. Study the relationship between confidence and support of association rules. For this purpose identify 10 best rules for confidence levels of 0.8, 0.9 and 1.0. Compare their corresponding supports.

C. Find interesting association rules. To do this experiment you will try to find appropriate values of the algorithm options such as support, confidence, lift, and conviction.