

Trainmining - Result analysis

Adrián Pérez Orozco

October 17, 2012

1 cSPADE

2 Analysis of first results

In order to have a first insight on the results that cSPADE is able to provide, we will first run it with default parameters. This does not apply any limit to the maximum time span of our rules, and sets the maximum number of elements in a sequence and the number of items in an element to 10.

A maximum number of elements in a sequence of 10, means that we can obtain frequent sequences containing up to 10 different time periods being concatenated. For our purposes, this might not be the most appropriate approach, as we are more interested in sequences containing no more than 2 elements: an *antecedent* and a *consequent*. There might be cases in which different events happening subsequently might be a better indicator of the events which are likely to happen in the future. However, selecting adequately the timespan comprehended in each event group, we might already include these without the need of creating rules of higher complexity.

Also, not putting any limit in the maximum timespan of our sequence is completely inappropriate. First of all, it highly increases computational complexity for obtaining the rules, as the search space is enormously higher. Also, it doesn't look much adequate to try to predict events happening three months if your time resolution is of one hour.