Data mining laboratory

Lab2 – sequential rules

Basic notions

- Transaction – a finite subset of items belonging to a certain domain with a time stamp.
- Sequence – an ordered set of transactions concerning one object.
- Data base (BD) – a set of sequences.
- Event - occurrence of a set of items in some transaction.

Sequential rules

A sequential rule (SR) is an expression in the following form:

$$A \Rightarrow B$$

where A and B are sequences

A - antecedent of a rule

B - consequent of a rule

A sequential rule expresses certain time consecution of events.

Parameters of sequential rules 1

Parameters of SR available in arulesSequences package

- Support
- Confidence
- Lift

Parameters of sequential rules 2

Parameters of SR available in arulesSequences package

- **maxsize** – the maximal number of items of an element of a sequence, default 10
- **maxlen**– the maximal number of elements of a sequence, default 10
- **mingap** – the minimum time difference between consecutive elements of a sequence, default none
- **maxgap** – the maximum time difference between consecutive elements of a sequence, default none
- ***maxwin –*** *the maximum time difference between any elements of a sequence, default none*

Support 1

The support of a sequential rule is calculated as a number of sequences including that rule.

The important is only a fact of occurrence of that rule in a given sequence not a number of occurrences.

Support 2

absolute support of a rule ($A \Rightarrow B$) =

number of sequences in DB including $A \cup B$ with a given time constrains.

Relative support – a frequency of occurrence of a given rule in DB

relative support of a rule ($A \Rightarrow B$) =

absolute support($A \Rightarrow B$) / number of sequences in DB

Support, Confidence, Lift

Interpretation of support, confidence, and lift parameters for sequential rules is analogous to interpretation of these parameters for association rules.

Hierarchy

Usage of hierarchy gives the same effect as in case of association rules - it allows discovering more general rules, which refer to different kind of hierarchy.

Sequential rules discovery with R

Package ***arulesSequences*** - selected methods:

- ***read_baskets –*** *read data with sequences*
- ***itemFrequency*** *– calculation of frequency of occurrence of items*
- ***ruleInduction*** *– an sequential rules generation based on prior discovered sequences.*
- ***cspade*** – frequent sequences discovery,
- ***inspect*** – showing sequences and rules
- ***subset*** *– selection of sequences or sequential rules meeting a user's requirements*