Association Rules Extractor

1. Goals of the software

Goal of the software is to allow user to extract association rules from Weka output and store these rules in structured form.

The software retrieves association rules from Weka output and extracts these rules to Excel sheet in structured form for further work. It also extracts additional parameters - support of the rule in count of cases, support in percentage, confidence of the rule and ID of the rule.

The software is able to extract metadata concerning source data set from which association rules were mined and can be also used for part of the data set preparation process.

The software is also able to check if association rules are ordered by confidence in descending order.

2. Source data

For association rules extraction is used Weka output. It is assumed that names of attributes are from OT01 to OT99. Kategories are named from A to Y. Category Z is reserved for missing values.

The software can be also used for part of the data set preparation process. This assumes that data set is in "Data set" sheet.

3. Sheets description

3.1. Entry sheet

„Entry sheet" contains metadata about source of association rules. It is required to set count of attributes (B4) and count of cases (B5). These values can be extracted by pressing "Extract count of attributes" button and / or "Extract count of cases" button.

"Association rules check" (B6) shows result of "Check association rules" function. If extracted rules have correct format then it shows "OK" value. Otherwise message box with error message is shown during "Check association rules" function.

"Fill attributes names" button can be used for changing attributes names in "Data sheet". Count of attributes (B4) must be filled.

3.2. Association rules-source

This sheet is used for association rules exported from Weka software. The sheet can be easily filled by copy and paste of Weka output. The software is tested with Weka version 3.6.1. Association rules must be ordered by consequent in descending order.

3.3. Association rules

This sheet contains association rules in structured form. The sheet is filled after pressing "Extract association rules" button. The sheet is filled based on content of "Association rules-source" sheet.

The first row of the sheet conatins the header.

Column „ID" contains ID of the rule. Id from "Association rules-source" is filled.

Column „confidence" contains confidence of the rule. Confidence is in range from 0 to 1 (0=0%, 1=100%).

Column „support [count]" contains support of the rule in count of cases (instances).

Column „support [-]" contains support of the rule in percentage (count of instances supporting the rule divided by count of all instances).

Column „cons. (attribute)" contains name of the attribute in the consequent of the rule.

Column „cons. (value)" contains value of the attribute in the consequent of the rule.

Rest of columns (OT01 - OT99) contains values from antecedent of the rule. If the value of the attribute is present then it means that the attribute is in antecedent and is equal to the filled value. If no value is present then it means that the attribute is not present in the antecedent of the rule.

3.4. Data set

This sheet can be used for preparation of the data set for processing in Weka software. Function "Fill attributes names" works with this sheet. It can be used for changing attributes names in "Data sheet". Count of attributes (B4) must be filled.

4. Software functions

The software has several function for association rules extracting and data set preparation.

4.1. Fill attributes names

"Fill attributes names" button can be used for changing attributes names in "Data sheet". Count of attributes (B4) must be filled. This can be usefull for preparation of the data set for processing in Weka software. This function works with "Data set" sheet.

4.2. Extract count of attributes

This function is used for extraction of attributes count from "Association rules-source" sheet. This function works with "Association rules-source" sheet.

4.3. Extract count of cases

This function is used for extraction of cases (instances) count from "Association rules-source" sheet. This function works with "Association rules-source" sheet.

4.4. Extract association rules

This function is used for extraction of association rules and transformation of these rules to structured forms. This function works with "Association rules-source" sheet and use this sheet as a source of association rules. These rules are extracted and transformed to structured forms. The result is in "Association rules" sheet. The output is described in this manual in chaper 3 sheet – Association rules. The funciton extracts rules with consequent length equal to 1. Lenght of antecedent is not limited (only by maximum count of attributes). Rules with missing values (the "Z" category) in consequent are omitted.

4.5. Check association rules

This function can be used for checking of association rules in structured form and works with "Association rules" sheet. This function checks extracted rules for wrong values and checks order of association rules.

5. Software restrictions

Count of attributes are limited to 99

Count of cases is limited by maximum rows in one Excel sheet. The first row is used for the header.

Length of association rules source is also limited by maximum rows in one Excel sheet.

Count of categories is limited to 25 (letters A to Y). The "Z" category is used for missing values.

Software is tested in MS Excel 2010.