Simple transformations
User manual

Author: Jaroslav Tykal
Library: SimpleTransforms.dll

Content:

1 Introduction

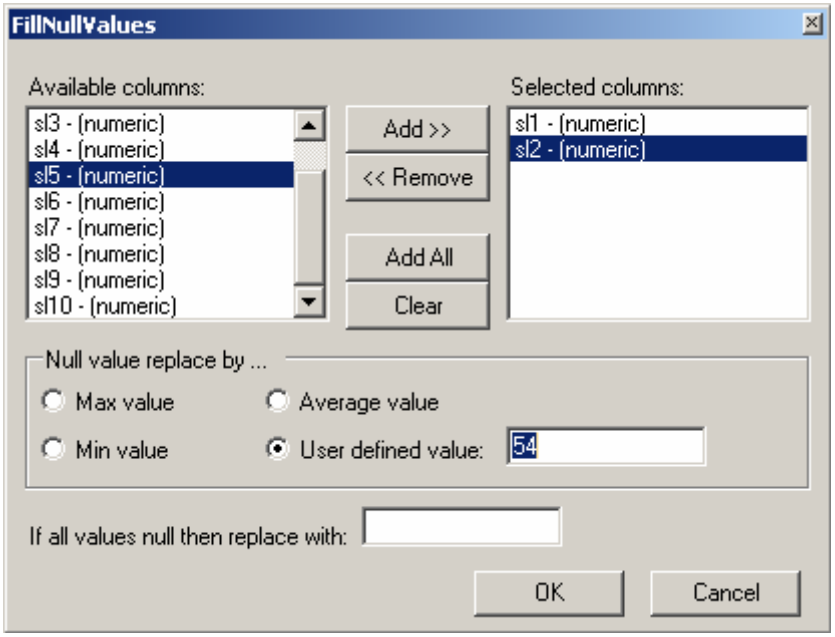
Data in database are very often incomplete or unsuitable for direct using in Data mining methods. Knocker application offers a connection to additional tools (via special dll) for manipulation with data. These tools are called "Simple transformations".
Each of simple transformation has a specific role. Knocker standard distribution is supplied with three simple transformations.

- **Fill null values** – replaces null values in selected column(s) according to user choice.
- **Normalization** – normalizes not null values in selected columns(s) into user specified interval.
- **Split table** – for many methods it is useful to have part of data for learning and part of data for verification if method adequately learned.

Simple transformations are started from main application window. Each method supports launch witch configuration so they can be launched repeatedly on similar data.

1.1 Fill null values

This transformation replaces null values in selected column(s) according to user input.



1. User selects columns where to replace null values
2. User chooses type of replacement.
   - Max value – maximal value is set instead of null value
   - Min value – minimal value is set instead of null value
   - Average value – average value is set instead of null value
   - User defined value – user entered value is set instead of null value
3. If there is selected one of {Max, Min, Average} in "Null value replace by" area, there should be defined value if all values in column are null. This value is useful when you reapply this transformation on similar data, but no all rows have filled value in the specified column.
4. Confirm using "OK" button.

Notes:

- There are only integer or number with decimal point types columns available.

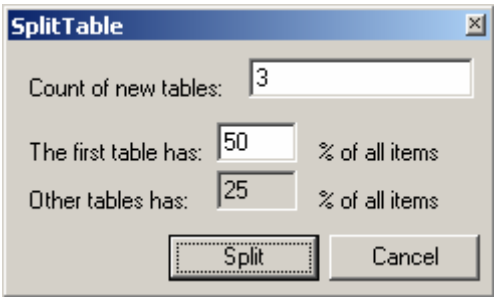
- All processed columns will be retyped to number with decimal point.

1.1.1 How to add into application

DLL: SimpleTransforms.dll
Class to run: FormFillNullValues

1.2 Split table

This transformation splits table into user defined count of tables. The first table has user defined number of rows of original table (in %). Other tables have equally divided the rest or rows.



1. Count of new table (value between 2 and 20)
2. Count of rows original table (in %) (value between 1 and 99)
3. Other values are computed.
4. Confirm using "Split" button.

Note:

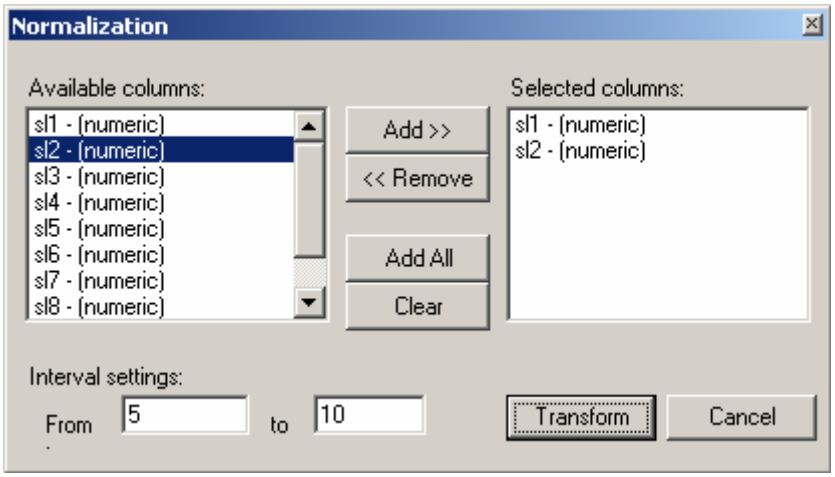
- This transformation creates (count of new tables) plus one derived version. Only one derived version has saved configuration of this transformation.

1.2.1 How to add into application

DLL: SimpleTransforms.dll
Class to run: FormSplitTable

1.3 Normalization

This transformation normalizes selected columns into user specified interval.



1. User selects column(s) to normalize value
2. User defines interval (From, to)
3. Confirm using "Transform" button

Notes:

- Available columns have type of integer or number with decimal point only.
- All processed columns will be retyped to number with decimal point.

1.3.1 How to add into application

DLL: SimpleTransforms.dll
Class to run: FormNormalization