

Simple transformations

User manual

Author: Jaroslav Tykal

Library: SimpleTransforms.dll

Content:

1	INTRODUCTION	3
1.1	FILL NULL VALUES	3
1.1.1	How to add into application	4
1.2	SPLIT TABLE.....	4
1.2.1	How to add into application	4
1.3	NORMALIZATION	4
1.3.1	How to add into application	5

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”.

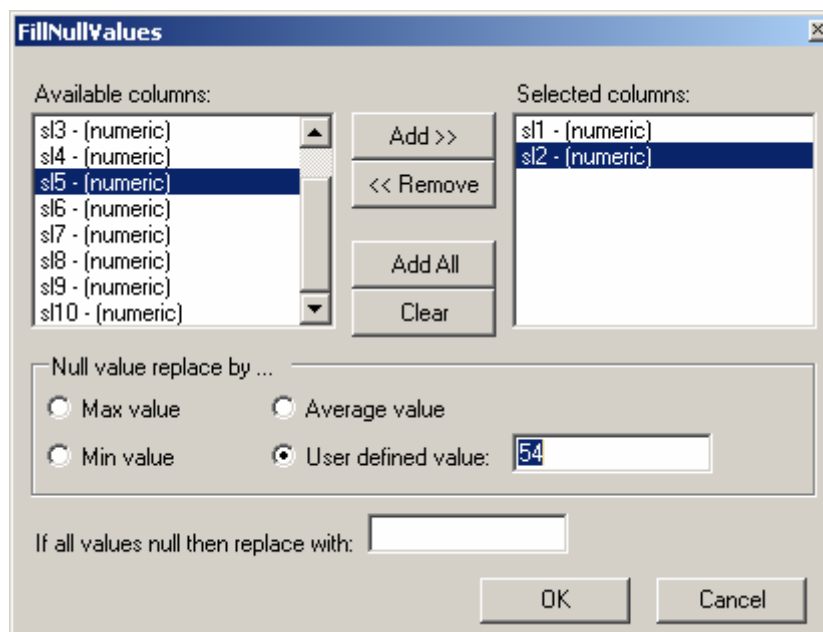
Each of simple transformation has a specific role. Knocker standard distribution is supplied with three simple transformations.

- **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 with 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.

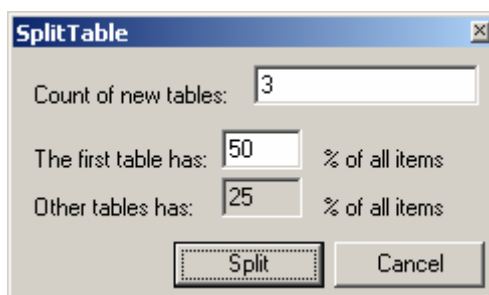
1.1.1 How to add into application

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 of rows.



The SplitTable dialog box has a title bar with a close button. It contains three input fields: 'Count of new tables:' with the value '3', 'The first table has:' with the value '50' and the label '% of all items', and 'Other tables has:' with the value '25' and the label '% of all items'. At the bottom, there are two buttons: 'Split' and 'Cancel'.

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.

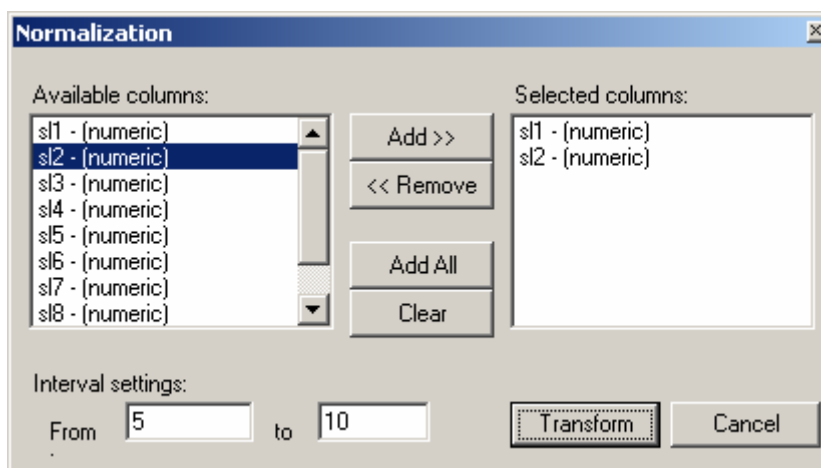
1.2.1 How to add into application

DLL: SimpleTransforms.dll

Class to run: FormSplitTable

1.3 Normalization

This transformation normalizes selected columns into user specified interval.



The Normalization dialog box has a title bar with a close button. It is divided into two main sections. The left section, 'Available columns:', contains a list box with eight items: 'sl1 - (numeric)', 'sl2 - (numeric)', 'sl3 - (numeric)', 'sl4 - (numeric)', 'sl5 - (numeric)', 'sl6 - (numeric)', 'sl7 - (numeric)', and 'sl8 - (numeric)'. The right section, 'Selected columns:', contains a list box with two items: 'sl1 - (numeric)' and 'sl2 - (numeric)'. Between these sections are four buttons: 'Add >>', '<< Remove', 'Add All', and 'Clear'. At the bottom, there is an 'Interval settings:' section with 'From' and 'to' input fields containing the values '5' and '10' respectively. To the right of these fields are two buttons: 'Transform' and 'Cancel'.

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
