Viscovery® SOMine 7 – Data Sheet

Basic Module | Visual Clusters

Visual data mining with self-organizing maps

Main Functions and Features

General characteristics

- **Visual Clusters** is the core module of the Viscovery SOMine Suite. It can be operated as a stand-alone system, as well as be flexibly extended with other modules of the suite.
- The two basic **Preprocess** and **Cluster** workflows are available; the first guiding the user through data import and preprocessing, the latter through the creation and exploration of the SOM model.
- Data sets with up to 100,000 records and up to 100 variables can be processed.

Data preprocessing

- Determination of variable names, types, and descriptive statistics
- Optional renaming of attributes
- Cross-reference definition from data records to external documents (links)
- Definition of new variables using built-in formula language
- Definition and automatic management of nominal variables (single-valued and multi-valued)
- Transformation of variables
- Treatment of outliers
- Replacement specification for ranges and special or missing values
- Conditional removal of data records
- Statistical and deterministic sampling and over-sampling of records

Data representation through self-organizing maps (SOM)

- High-performance computation of batch SOM based on classical Kohonen algorithm
- Two-dimensional SOM data representation on a hexagonal grid
- Predefined training schedules with selectable map size, granularity, and tension
- Automatic standardization of variables with additional scaling options
- Definition of the influences from individual attributes on the map ordering by setting attribute priorities
- Automatic compensation of correlations in the data
- Well defined treatment of missing values in all stages of model creation
- Optional setting of default parameters for map creation

SOM visualization and exploration

- Interactive visualization of attribute distributions and characteristic values in a map window
- Automatic color-coding of attributes with transformation-adjusted color scale or black-and-white options
- Annotation of the map with labels
- Manual drawing of trajectories and selections in the map
- Various options for selecting and unselecting map regions (by color-scale, interval, cluster, etc.)
- Display of thumbnails from external documents over the map window
- Display of nearest neighbors from the active node in the map
- Sorting of attributes in the map window according to similarity or priority

Viscovery® SOMine 7 – Data Sheet

Visual cluster analysis

- Automatic execution of agglomerative clustering methods (SOM-Ward, Ward, SOM-Single-Linkage)
- Selection of preferred cluster method and initial number of clusters prior to map creation
- Integrated visualization of cluster boundaries, cluster centers, and inner structures
- Display of separate clusters with optional color-coding of clusters (flat, shaded, or U-matrix)
- Display of cluster means for all attributes in the statistics pane
- Creation of map labels from cluster names

Statistical analysis of data associated with workflow steps

- Descriptive statistics
- Attribute histograms
- Correlation analysis
- Principal component analysis
- Frequency tables
- Box plots
- Scatter plots

Access to original data

- Data-record browser for showing original data from any active workflow step or selected region in the map
- Opening of external documents by clicking a region of the map
- Import and export of map labels, selections, and paths from/to external files

Available data interfaces

- Import and export of tab-delimited flat-text files (*.txt) and Microsoft Excel files (*.xlsx)
- Import of space-separated flat-text files (*.txt) and Excel 97/2000/2003 files (*.xls)
- Intelligent copy-and-paste function between Viscovery internal data and external software
- Export of SOM node values as a table