

Viscovery[®] SOMine 8 – Data Sheet

Extension Module | **Cluster and Classify**

Advanced SOM model exploration, clustering, and classification

Main Functions and Features

General characteristics

- **Cluster and Classify** is an extension module of the Viscovery SOMine Suite. It requires the **Visual Explorer** core module and can optionally be combined with other modules of the suite.
- Provides multiple features for advanced exploration of the self-organizing map (SOM) model, additional cluster algorithms and features for the interactive definition of alternative segmentations.
- The module provides two additional workflows: the **Create Classifier** workflow guides the user through the creation and validation of classifier models, the **Apply Classifier** workflow handles the application of classifier and cluster models to new data and the evaluation of classification results for application and control group.

Group profiling and comparison

- Identification of significant variable deviations using interactive profiling of data groups (pools)
- Creation of groups from clusters, nodes, neighborhoods, or arbitrary selections
- Evaluation of differences between two arbitrary groups using a statistical contrast
- Define reference group from any cluster, node, neighborhood, or arbitrary selection
- Evaluation of non-trivial group descriptives using the “Profile” chart and “Cluster” pane

Cluster characteristics

- Automated comparison of existing clusters regarding significant differences in attribute means
- Display of attributes characterizing each cluster based on cluster profiles
- Option for showing each attribute in the cluster where it is extreme

System state monitoring

- Dynamic simulation of process trajectories in the SOM model
- Sequential reading of records from time-ordered data file
- Selection of velocity and trace length of simulation

Multiple segmentations

- Creation of additional segmentations in a SOM model, each with the option of obtaining compact clusters on the SOM
- Classical K-Means algorithm (in addition to Ward clustering)
- New density-based cluster algorithm (Connectivity) based on multi-linkage fusion and cluster relevance
- Uniform cluster indicator for all cluster methods to identify good segmentations
- Interactive selection of cluster method, number of clusters, and attributes for new segmentation
- Display of existing segmentations in the dedicated “Segmentation” window
- Annotation, comparison, copy, and deletion of segmentations

Interactive cluster definition

- Manual adjustment of cluster boundaries by joining and deleting clusters
 - Definition of a new cluster or extension of an existing cluster with an arbitrary selected area
 - Automatic determination of cluster names from a nominal attribute
 - Annotation and renaming of clusters
 - Assignment of post-processing formulas to clusters
-

Viscovery[®] SOMine 8 – Data Sheet

Classifier models

- Creation of classification models combining SOM data representation and conditional probabilities with the dedicated **Create Classifier** workflow.
- Deterministic and random partitioning of data into a training set and up to two test data sets
- Automatic tuning of the temperature parameter
- Quantitative validation using log loss and classification error on training and test data

Classification

- Application of classifier and cluster models to new data with the dedicated **Apply Classifier** workflow
- Calculation of class probabilities and assignment of new data records to classes from classifier models
- Selection of segmentations as classifier, including evaluation of post-processing formulas
- Automatic random generation and administration of control groups
- Export of classification results and, for each record, class probabilities, and map node values to tab-delimited flat-text files (*.txt) and Microsoft Excel files (*.xlsx)

Evaluation of classification results

- Visualization of applied data distributions over the SOM model
- Specification of actual class names in an additional attribute for test or evaluation purposes
- Additional visualization of classification error over the map if actual class memberships are known
- Confusion matrix for evaluation of predicted vs. actual classes
- Accuracy, precision, sensitivity and specificity for every class and the whole model
- Comparison of campaign results with definable application and control group