Cluster Correspondence Views for Enhanced Analysis of SOM Displays

Jürgen Bernard, Tatiana von Landesberger, Sebastian Bremm, and Tobias Schreck

Interactive Graphics Systems Group, Technische Universität Darmstadt, Germany

TECHNISCHE UNIVERSITÄT DARMSTADT

TUD Graphisch-Interaktive Systeme

Motivation and Idea

- The Self-organizing Map Algorithm [1] is a popular and widely used cluster algorithm. It organizes clusters on a grid structure, making it very amenable for visualization
- Application areas: Visual cluster analysis of trajectory data [2]

Problems

- SOM method only provides implicit information about the number of clusters found in the data
- SOM requires setting of significant number of parameters
- Lack of quality assessment of SOM clustering results

Idea

- Compare and validate the output of the SOM against alternative clustering algorithms
- Combine advantages of all methods and, avoid their disadvantages

Method

1. Calculation of alternative clustering results

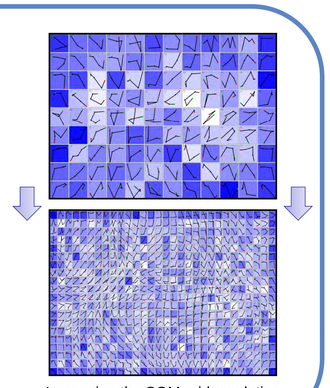
- K-means (partitioning) and DBScan (density-based) [3]

2. Mapping alternative clusters onto the SOM reference grid

- Rising the SOM grid resolution by spline interpolation [4] as a reference for the projection of alternative clustering results

3. Visual representation of alternative clusters

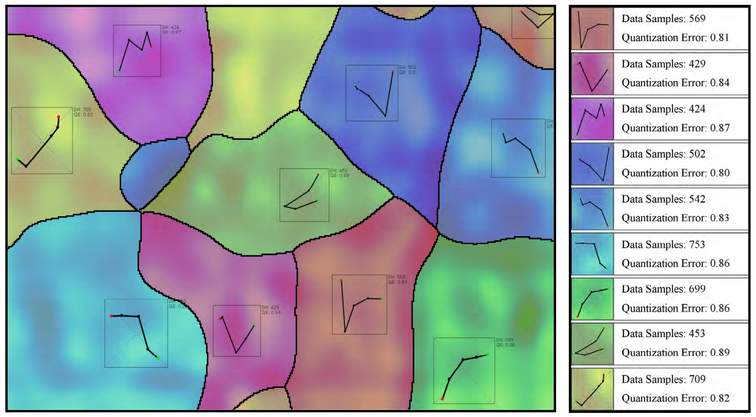
- Glyphs for the cluster representatives based on the underlying data (different size)
- Color coding of corresponding areas on the SOM grid



Increasing the SOM grid resolution

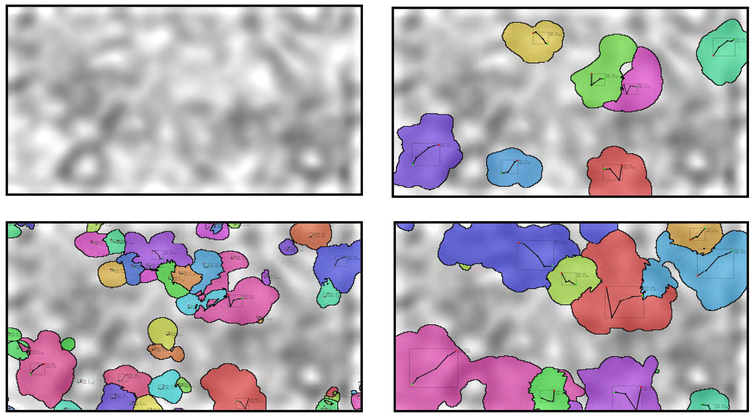
Cluster color coding

Assess the distribution of k-means clusters



Color mapping of a k-means cluster result (k=10) onto the reference SOM. The position of each cluster on the map is shown by a unique colored area (in the nearest neighbor sense). This view allows to assess the distribution of the cluster partition over the map. Each cluster is also represented by a glyph showing the clusters centroid trajectory, and in a list-based caption with statistical information.

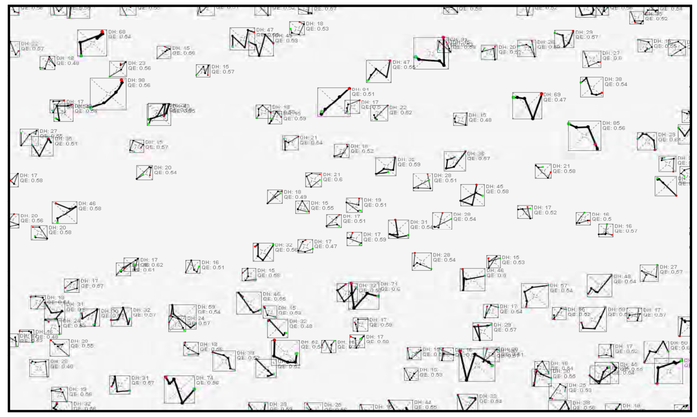
Assess the distribution of DBScan clusters



Color mappings of DBScan cluster results with different parametrizations. Based on the nature of density-based clustering results, not all data items are necessarily also represented by a cluster. The display is used to easily assess the nature and distribution of DBScan clusters in relation to the overall distribution of data elements. An overlay of this view with SOM cell distance view (light regions mean small average distances to neighbor cells) shows a strong correspondence between the density-based clustering and the dense areas of the SOM grid.

Cluster Visualization

Increased number of DBScan clusters



DBScan cluster result with an even larger number of clusters. The underlying SOM grid is used as reference layout for positioning of the DBScan clusters. This visualization utilizes the layout qualities of a SOM to allow an effective exploration of the BDScan clustering result, benefiting on the topology-preserving nature of the SOM [1] .

Future Work

- Integration of additional alternative clustering results
- Design of cluster glyphs to encode additional statistical information about clusters
- Extend combinative approaches for visual and statistical cluster quality assessment
- Visual-interactive cluster detection / cluster refinement (extended work accepted in a Paper at VDA2011)

References

[1] T. Kohonen. Self-Organizing Maps. Springer, 3rd edition, 2001.
[2] T. Schreck, J. Bernard, T. Tekušová, and J. Kohlhammer. Visual cluster analysis of trajectory data with interactive Kohonen maps. Palgrave Macmillan Information Visualization, 8:14–29, 2009.
[3] J. Han and M. Kamber. Data Mining: Concepts and Techniques. Morgan Kaufmann, 2006.
[4] J. Bernard, T. von Landesberger, S. Bremm, and T. Schreck. Micromacro views for visual trajectory cluster analysis. In Eurographics/ IEEE Symposium on Visualization, 2009. Poster.

Acknowledgments

This work was partially supported by the German Research Foundation (DFG) within the project Visual Feature Space Analysis, as part of the Priority Program on Scalable Visual Analytics.

Deutsche Forschungsgemeinschaft DFG