Diplomarbeitspräsentationen der Fakultät für Informatik

Visualisations for Comparing Self-organising Maps

Baum, Doris

Technische Universität Wien
Institut für Softwaretechnik und interaktive Systeme
Arbeitsbereich: Information and Software Engineering
Betreuer: Ao. Univ.-Prof. Dr. Andreas Rauber

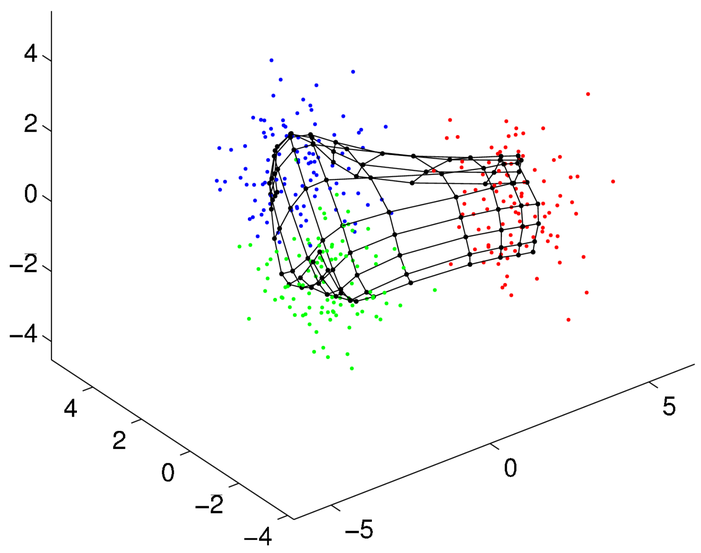
Magisterstudium:
Intelligente Systeme

Goal of the Master's Thesis

The goal of this Master's thesis was to create SOM visualisations that facilitate the comparison of two or more SOMs to gain information on the differences between them, the quality and errors of their projections and the influence of the parameters which control the training. Three new visualisations were developed, implemented and tested on artificially constructed and real-world data: The Data Shifts Visualisation, the Cluster Shifts Visualisation, and the Comparison Visualisation

Self-organising Maps (SOMs)

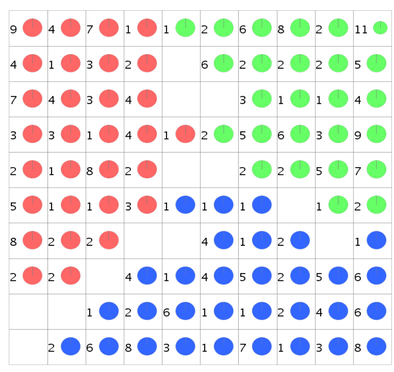
Self-organising Maps (SOMs) are a very useful method for exploring and analysing large data collections: They project high-dimensional data into a low-dimensional output space so that it is easier to analyse for humans than the original data.
The projection is done by training a 2-dimensional grid of artificial neurons to fit the data. The neurons are represented by a vector in the input space which is a prototype for the input vectors surrounding it. Each data vector is mapped onto the neuron with the closest prototype vector.



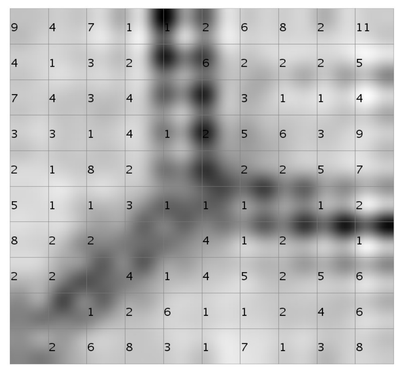
2-dimensional SOM-Grid that was fitted to 3-dimensional data; the data consists of three different classes which are marked in different colours.

SOM Visualisations

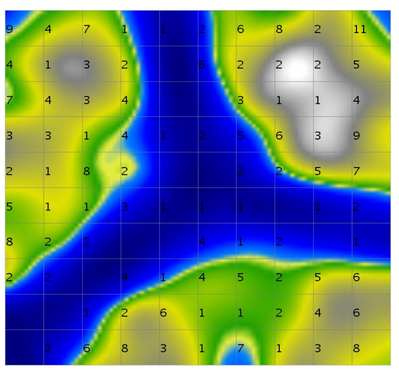
For the purpose of data exploration with SOMs, plenty of visualisations exist which display different aspects and properties of a SOM and the data.
Examples are the Class Information Visualisation, the U-Matrix and the Smoothed Data Histograms.



Class Information



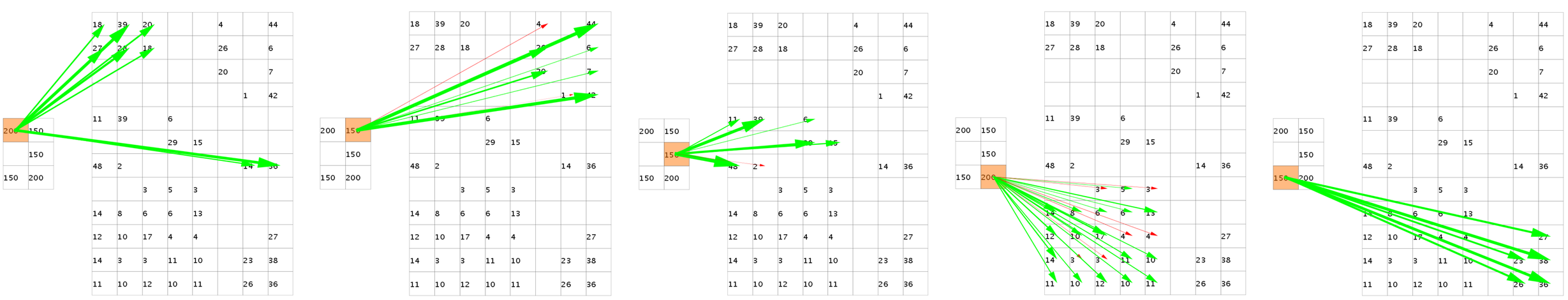
U-Matrix



Smoothed Data Histograms

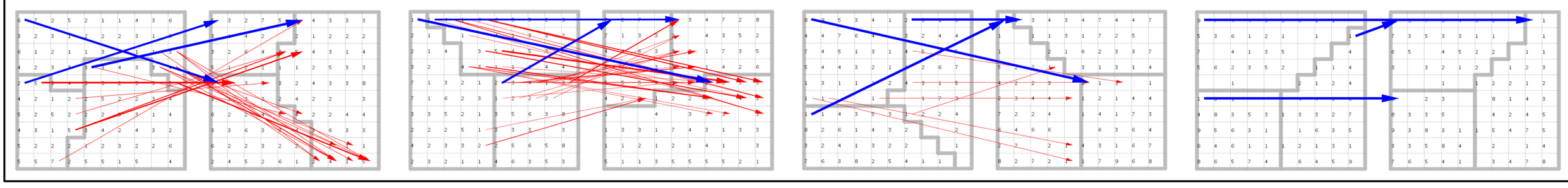
Data Shifts Visualisation

The Data Shifts Visualisation compares two SOMs trained on the same data. It displays changes in the position of data vectors relative to their neighbours on the SOMs. Depending on whether the members of a data vector's neighbourhoods coincide in both SOMs or not, the neighbourhood is considered “stable” (green arrow) or the data vector is called an “outlier” (red arrow). The number of coinciding neighbours and the size of the neighbourhood can be adjusted by the user.



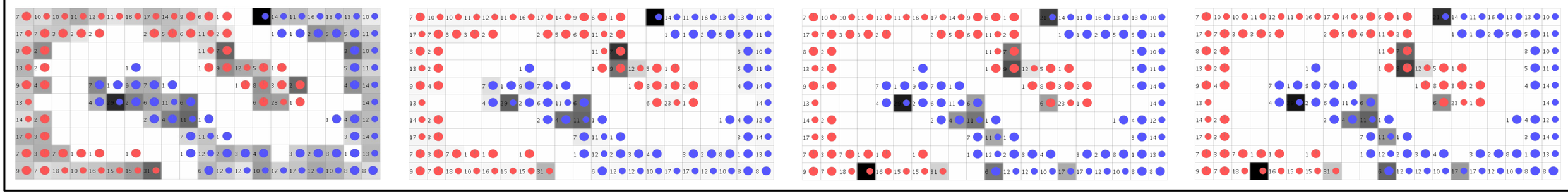
Cluster Shifts Visualisation

The Cluster Shifts Visualisation compares two SOMs trained on the same data. It uses hierarchical clustering methods to cluster the SOMs; the resulting clusters of both SOMs are then identified with each other using a confidence measure. The confidence is computed for each pairing of clusters from the agreement of data vectors in both clusters; then the pairs with the highest confidences are assigned to each other. Cluster assignments are indicated by blue arrows, data vectors that weren't mapped to assigned clusters are considered “outliers” (red arrows).



Comparison Visualisation

The Cluster Shifts Visualisation compares multiple SOMs trained on the same data. For a given SOM it displays for each neuron the mean or variance of the pairwise distance the data vectors mapped to it have in other SOMs. Different distance functions can be used, currently euclidean distance and a cluster-based distance measure similar to single-linkage are implemented.



Kontakt: dorisbaum@gmx.net