The Metro Visualisation of Component Planes for Self-Organising Maps



Robert Neumayer, Rudolf Mayer, Georg Pölzlbauer, and Andreas Rauber

Department of Software Technology Vienna University of Technology **Favoritenstraße 9-11/188** Vienna, Austria {neumayer,mayer,poelzlbauer,rauber}@ifs.tuwien.ac.at

Experimental Setup

Experiments

o evaluate our method we performed experiments on the Boston Housing benchmark data set describing 506 instances. It contains information collected by the U.S Census Service concerning housing in the area of Boston. The data is described in thirteen continuous and one binary attributes, such as 'Median value of owner-occupied homes' / 'medv' or 'average number of rooms per dwelling' / 'rm'. We trained a Self-Organising Map of 8 x 18 units, discretisation is done for six regions.

One well-known method to get a better understanding of the chamap is the visualisation of Component Planes, i.e. the colour-coding of single components or variables; each component needs its own illustration.



Component Planes

The unified distance matrix (U-Matrix) is a visualisation technique that shows the local cluster boundaries by depicting pair-wise distances of neighbouring prototype vectors. We use the U-Matrix as background visualisation for metro lines.



U-Matrix Visualisation

Metro Maps

Metro maps introduced the concept of skewed distances, as opposed to geographically correct distances, focusing on the structure and relationships (points of connection) between metro lines. This concept was originally developed by Harry Beck in 1931 and then became well-known as the map design of the London Underground transport network.



Geographically correct distances for the London metro http://homepage.ntlworld.com/clivebillson/tube/tube.html

Self-Organising Map

The Self-Organising Map provides a mapping from a high-dimensional input space to a lower-dimensional output space. In the process of this mapping input patterns that are located close to each other in the input space will also be located closely in the output space. The SOM therefore provides a sort of clustering of the data, however, without explicitly assigning data items to clusters.

Related Work





Metro Visualisation Experiments Conclusions and Future Work nterchange Stations

Related Work



Connected region centres for the Boston Housing data set. We performed discretisation of the Component Planes into six bins, resulting in six region centres for all 14 components. Black squares along the differently coloured lines denote the centres. The U-Matrix serves as background visualisation, revealing the overal clusture structure of the map.

Boston Housing Connected Region Centres





Classic 1933 London Metro Map

Binned Components

In order to more easily identify the distribution of each variable on the map, we perform a discretisation step for Component Planes so that components are presented as n partitions of the SOM. This more clearly reveals the structure of each individual component.



Connected Region Centres

Next, we locate the centre of each bin and display these region centres as black dots resembling metro stations. Adjacent centres are linked by lines to emphasise the ordering of the bins across the map. The resultant lines may be depicted as overlay to other visualisations of the SOM such as the U-Matrix (which, when, using a suitable colour space, resembles rivers as cluster boundaries).

of the data set. Discretisation was performed for six bins.

Abstract

The Self-Organising Map (SOM) is a popular unsupervised neural network model which has successfully been used for clustering various kinds of data. To help in understanding the influence of single variables or components on clusterings, we introduce a novel method for the visualisation of Component Planes for SOMs. The approach presented is based on the discretisation of the components and makes use of the well-known metro map metaphor. It depicts consistent values and their ordering across the map for discretisations of various components and their correlations in terms of directions on the map. In our approach Component Lines are drawn for each component of the data, allowing the combination of numerous Component Planes into one plot, possibly overlaying standard visualisation methods like the U-Matrix. We also propose a method to further aggregate these Component Lines, by grouping highly correlated variables, i.e. similar lines on the map. To show the applicability of our approach we performed experiments on for popular machine learning data sets.

The Boston Housing map after the snapping step. This illustration consists of 45 degree angles (and multiples thereof) between lines only. This performs a structuring of the metro lines; readability is thus enhanced. Note that some metro lines are highly correlated like 'rm' and 'medv' in this illustration, which are displayed as pink and dashed red lines, respectively.

Boston Housing Snapped

Component Discretisation and Visualisation

Visual Enhancements

For a more intuitive and smooth representation, and to more closely resemble the metaphor of an actual metro map, the locations of the region centres are adjusted in a way that the lines are drawn only horizontally, vertically, or diagonally, i.e. in multiples of 45 degrees.

To obtain a clearer picture, we combine very similar component lines (i.e. highly correlated components) into a single one. This is achieved by calculating the distances between lines on the map as the pairwise distances between their respective region centres. Lines that show a distance below a certain threshold are subsequently combined.

The resultant lines after the aggregation step was performed. Ward's clustering was applied in connection with the proposed distance measure between lines. Four pairs of components have been aggregated, resulting in ten remaining lines. The 'medv' and 'rm' components, for instance, are displayed by one line only.

Boston Housing Aggregated

Aggregation of **Component Planes**

Line Distances

Future Work

We want to investigate using different distance functions between Component Lines, focusing on local distance functions and edit distances, which are used for snapping and aggregation. We also plan on utilising the resultant graph structure for the comparison of differently trained SOMs which are based on the same data set.

Conclusions

We presented a novel and intuitive method for the visualisation of Component Planes and their correlations, by using the metro map as a metaphor of Component Lines. Driven by the problems arising when trying to plot all components, namely the resulting high number of plots, we proposed the discretisation of single components of the input data. For higher-dimensional input vectors, grouping the Component Lines can reduce the complexity of the visualisation. The main advantages of the metro visualisation are:

- Plotting Component Planes in one single illustration
- Visualisation of correlations between components
- Aggregation of highly correlated components
- The Metro Visualisation can be used as an overlay to other standard visualisations of the SOM, simultaneously displaying different characteristics of the map

Presented at the **20th International Joint Conference** on Neural Networks August 12 - 17 2007 Orlando, Florida, USA