Extracting information from regional data using Self-Organizing Maps

Markus Casper[1], Marcus Herbst[1]

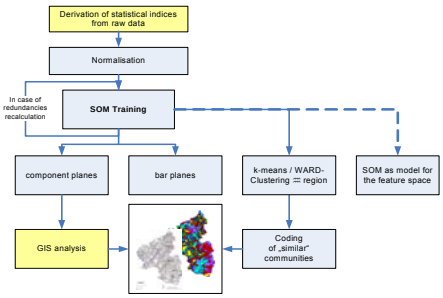
forum stat
Forschungszentrum für Regional- und Umweltstatistik

Problem

The extraction of information from regional statistical data involves processing high-dimensional data sets (i.e. with many variables) in such a way that similarities between the individual spatial units and relationships in their characteristics become accessible to our cognitive abilities.

Data and processing approach

- birth rate per 1000
- population density
- age ratio (≥60 / ≤18)
- population dynamics
- commuter balance
- per capita income
- per capita expenses
- hotels per 1000
- agricultural holdings
- fraction of agricultural areas
- business enterprises per 1000

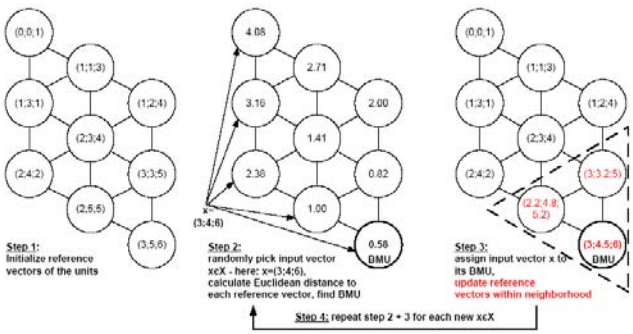


Training of a Self-Organizing Map

A Self-Organizing Map (SOM), consisting of nodes *i=1…k* which are characterized by a vector

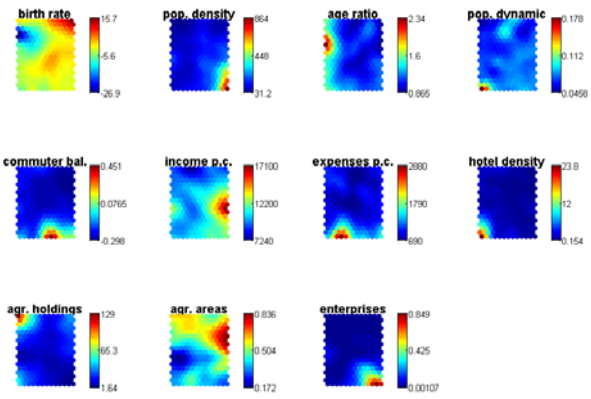
$$\boldsymbol{m}_i = [\mu_{i1}, \mu_{i2}, \ldots, \mu_{in}]^T \in \Re^n$$

with the same dimensionality as the data sets $\boldsymbol{x} \in X$ is trained:

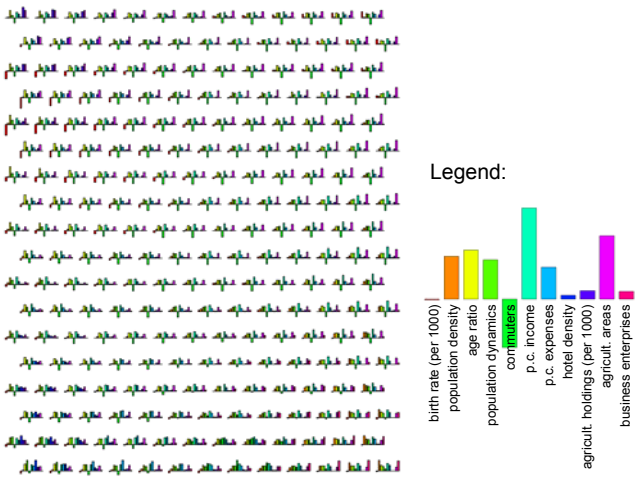


Results

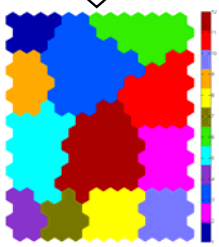
„Component planes" (i.e. distribution for individual variables)

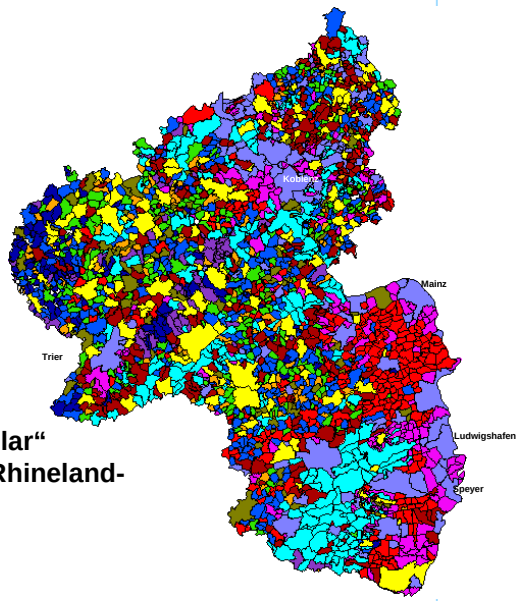


„Bar plane"



K-means clustering of SOM mapping (12 clusters)





Mapping of „similar" communities in Rhineland-Palatinate

Conclusions

- SOM allows topologically ordered projection of high-dimensional data by similarity (i.e. similarity in data characteristics translates into distance on the SOM)
- this allows an intuitive exploration of regional statistical data, e.g. „correlation hunting"
- k-means clustering helps delineating „similar" regions on the SOM for subsequent mapping in space
- the clusters can be interpreted as urbanized regions, forested regions, rural areas, peripheral communities etc.
- it can be seen that spatial vicinity does not always imply similarity of socio-economic characteristics
- the SOM also allows extrapolating new data sets

Further Information on Self-Organizing Maps and their application

This work was carried out using the SOM-Toolbox for Matlab by the "SOM Toolbox Team", Helsinki University of Technology (http://www.cis.hut.fi/projects/somtoolbox)

Kohonen, T.: Self-Organizing Maps,3rd ed., Berlin, Heidelberg, New York, 501pp., 2001.
Herbst, M., Casper, M.C.: Towards model evaluation using Self-Organizing Maps, Hydrol. Earth Syst. Sci., 12, 657-667, 2008.
Herbst, M., Gupta, H.V. and Casper, M.C.: Mapping model behaviour using Self-Organizing Maps, Hydrol. Earth Syst. Sci.,13, 395-409, 2009.
Herbst, M., Casper, M.C., Grundmann, J. and Buchholz, O.: Comparative analysis of model behaviour for flood prediction purposes using Self-Organizing Maps, Nat. Hazards Earth Syst. Sci., 9, 373-392, 2009.

Universität Trier

[1]Physical Geography, University of Trier, 54286 Trier, Germany
e-mail: casper@uni-trier.de, Tel.: +49 651 2014518, Fax: +49 651 2013976

Physische Geographie