Cartographic Selection Using Self-Organizing Maps

Bin Jiang[1] and Lars Harrie[2]

[1]Division of Geomatics, Institutionen för Teknik
University of Gävle, SE-801 76 Gävle, Sweden
e-mail: bin.jiang@hig.se

[2]GIS centre, Lund University
Sölvegatan 13, SE-223 62 Lund, Sweden
e-mail: lars.harrie@lantm.lth.se

Abstract
We propose selection of cartographic objects based on the technique of self-organizing map (SOM), an artificial neural network algorithm for data clustering and visualization. Using the SOM training process, the approach derives a set of neurons by considering multiple attributes including topological, geometric and semantic properties. The set of neurons constitutes actually a SOM, with which each neuron corresponds to a set of real objects with similar properties. Our approach also sets up an exploratory linkage between the SOM and an object set. The method is evaluated on a street network.

Keywords: cartographic generalization, selection, street networks, self-organizing map, neural networks

1. Introduction

A good method for selection of cartographic objects is essential in the generalisation process. Ideally, the selection should be based on all available information about the objects. In this paper we propose using self-organizing map (SOM) (Kohonen 2001) for the selection process. The advantage of using SOM is that the method can take several attributes into account simultaneously and it also contains visualisation tools.

This paper is structured as follows. The remainder of Section 1 contains previous work on cartographic selection and on the applications of SOM in cartography and GIS. Section 2 briefly introduces the basic principle and algorithms of SOM and Section 3 contains a case study.

1.1 Previous studies of selection

An algorithmic approach for selection consists of two steps: identifying regions of high density of objects and removing some of the objects within these regions. The first step is performed by a clustering technique. For example, Regnauld (1996) used a minimum spanning tree of building objects. When the minimum spanning tree was computed, some edges were removed according to criteria such as length of edges, alignment of building objects, etc. Finally, clusters were defined as connected building objects (by the remaining edges), and a density analysis was enabled.

To decrease the density of objects in the identified regions, a selection procedure is applied based on object size and/or location. Töpfer and Pillewizer (1966) investigated the number of objects that ought to be shown at different scales (using reverse engineering). They proposed *radical law*: the ratio of the number of objects in two maps should equal the square root of the ratio of the map scales (this is the general rule; further details are given in Töpfer and Pillewizer 1966). The radical law does not give any information about which objects should be selected. An approach that considers both the number of objects and which objects should be selected is Shannon information theory (Shannon and Weaver 1964). Bjørke (1996) proposed information theory for cartographic selection. He started by creating several proposals for object selection and then evaluated all these proposals. The proposal that communicated the message best (according to the information model) was then chosen.

The selection of road objects in a network is a major problem in generalisation. The most common method is to base the selection on road object types. However, this method may cause loss of important connectivity properties in the road network; to circumvent this problem graph theory can be used. Edges and nodes may be weighted by connection properties (Mackaness and Beard 1993) or by relative importance in linking a given set of locations (Thomson and Richardson 1995). Edges and nodes with low weights are then eliminated. It is, however, questionable whether the segments can be handled individually. Richardson and Thomson (2001) group segments based on the *good continuation principle*. This principle states that if two edges are connected and aligned they are perceptually group by humans. In their work, Richardson and Thomson, base their selection on these segment groups, rather than individual segments. Recently Jiang and Claramunt (2002) proposed a set of algorithms based on streets (that consists of several links) rather than on street segments. The approach is performed at a topological level with a representation, which takes named streets as nodes and street intersections as links of a connectivity graph. Based on the graph-theoretic representation, each street is assigned by two structural measures namely connectivity and average path length on which the selection of important streets is based. Eventually the selection of streets is applied to an entire named street rather than a street segment using the two structural measures respectively.

1.2 Previous studies using SOM in cartography and GIS

SOM has been used in many fields such as data classification, pattern recognition, image analysis, and exploratory data analysis (for an overview, see Oja and Kaski 1999). In the domain of GIS and cartography, relatively few applications have been made. Openshaw and his colleagues have used SOM in spatial data analysis (Openshaw 1994, Openshaw et al. 1995). Recently some new proposals have been made in using SOM for spatial data exploration (Li 1998) and image classification (Luo and Tseng 2000). SOM has been used for building typification in cartographic generalization (Højholt 1995, Sester 2001). In the typification process, a number of building objects are set to represent a larger set of objects. A major issue here is that the new objects should reflect the original pattern of objects. In the approach introduced by Hojholt and Sester, new building objects are placed randomly on the map. Then the locations of the new building objects are changed using SOM. In this training process the original building objects are used for attracting the new building objects. In this way the location of the new building objects will give a similar pattern as the location of the original building objects (but properties such as parallelism is not maintained). The use of

SOM in this paper is rather different. Here it is used for attribute clustering as a pre-process for selection; the locations of the streets are not altered.

2. Self-organizing map

SOM is a well-developed neural network technique for data clustering and visualization. It can be used for projecting a large data set of a high dimension into a low dimension (usually one or two dimensions) while retaining the initial pattern of data samples. That is, data samples that are close to each other in the input space are also close to each other on the low dimensional space. Herewith we provide a brief intuitive introduction to the SOM; readers are encouraged to refer to more complete descriptions in literature (e.g. Kohonen 2001).

2.1 Basic principles

The SOM training algorithm involves essentially two processes, namely vector quantization and vector projection (Vesanto 1999). Vector quantization is to create a representative set of vectors, so called output vectors from the input vectors. In general, vector quantization reduces the number of vectors. This can be considered as a classification, or clustering, process. The other process, vector projection, aims at projecting output vectors (in d-dimensional space) onto a regular tessellation in lower dimensions (i.e., a SOM), where the regular tessellation consists of an arbitrary number of neurons. In the vector projection each output vector is projected into a neuron where the projection is performed as such, “close” output vectors in d-dimensional space will be projected onto neighbouring neurons in the SOM. This will ensure that the initial pattern of the input data will be present in the neurons.

The two tasks are illustrated in figure 1, where usually the number of input vectors is greater than that of output vectors, i.e. $n \succ m$, and the size of SOM is the same as that of output vectors. It should be emphasized that for an intuitive explanation of the algorithm, we separate it as two tasks, which are actually combined together in SOM without being sense of one after another.

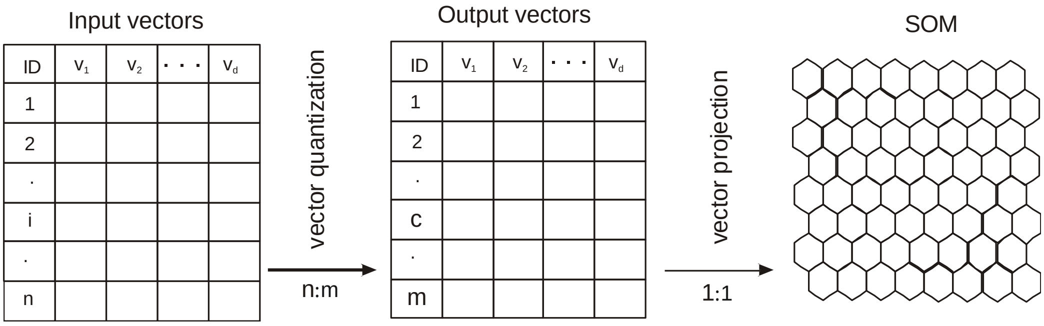

Figure 1: Illustration of SOM principles

2.2 The algorithm

The above two steps,vector quantization and vector projection, constitute the basis of the SOM algorithm. Vector quantization is performed as follows. First the output vectors are chosen randomly or linearly by some values for its variables. In the following training step, one sample vector x from the input vectors is randomly chosen and the distance between it and all the output vectors is calculated. The output vector that is closest to the input vector x is called the Best-Matching Unit (BMU), denoted by c:

$$\| x - m_c \| = \min_i \{ \| x - m_i \| \} , \qquad [1]$$

where $\| . \|$ is the distance measure, usually Euclidian distance. Then the BMU and other output vectors in its neighbourhood are updated to be closer to *x* in the input vector space. The update rule for the output vector *i* is:

$$m_i(t+1) = m_i(t) + \alpha(t) h_{ci}(t)[x(t) - m_i(t)] , \qquad [2]$$

where $x(t)$ is a sample vector randomly taken from input vectors, $\alpha(t)$ is a learning rate function, and $h_{ci}(t)$ is a neighbourhood kernel function, and *t* denotes time step. In figure 2 one step of the update rule is visualised, and the BMU and its neighbours are updated. Vectors that are regarded as neighbours are decided by the neighbourhood kernel function ($h_{ci}(t)$). As seen from the update rule the vector is moving towards the randomly chosen input vector $x(t)$. The distance of this movement is determined by the learning rate function ($\alpha(t)$). The updating process will proceed until a threshold criterion is fulfilled; after that the vector projection process is performed.

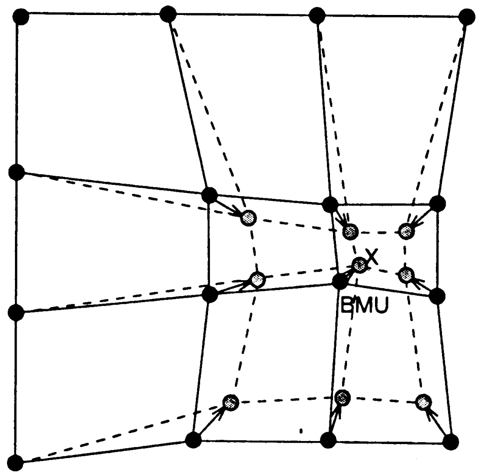

Figure 2: Updating the BMU and its neighbours. The solid and dashed lines represent to situation before and after updating, respectively. Copied from Vesanto et. al (2000, p. 9).

In the vector projection process the output vectors are projected on to a regular 2 dimensional grid, where each neuron corresponds to an output vector (that is the representative of some input vectors; see Figure 1). For a better visual effect of SOM, hexagonal rather than rectangular grid is often adopted.

Case study

Below we evaluate SOM as a technique for cartographic selection in a case study. The case study is performed on a Munich street network with totally 785 streets.

The selection of streets is based on the following attributes: degree, closeness, betweenness, length, number of lanes, speed limit and functional class (the three first attributes here are topological measures of the streets, see Jiang and Claramunt 2002 for details). To base the cartographic selection on these attributes we need to transform the numerical values; in this case study this is performed in a two-stage process. Firstly, the variation of the values of the individual attributes is very large, so we transform the dataset into a unit interval [0, 1] to guarantee that all variables have the same variation. Secondly, the seven attributes should not be considered at equally important in terms of street selection. We adopt the weight vector [1, 1, 1, 2, 2, 2, 3] for the seven attributes in the order of [degree, closeness, betweenness, length, lanes, speed, class]. After this transformation we create one input vector (cf. Figure 1) for each street in the network.

Based on the equation (1) and (2) one hundred output vectors are created; these vectors are then projected on a SOM (cf. Figure 1). The SOM is illustrated on the left side of Figure 3. In this Figure each neuron are connected to a set of streets in the map. By selecting neurons (black cells in figure 3) a selection of the streets are performed. The selection of neurons is done using additional maps that describe the attribute value variations in the SOM.

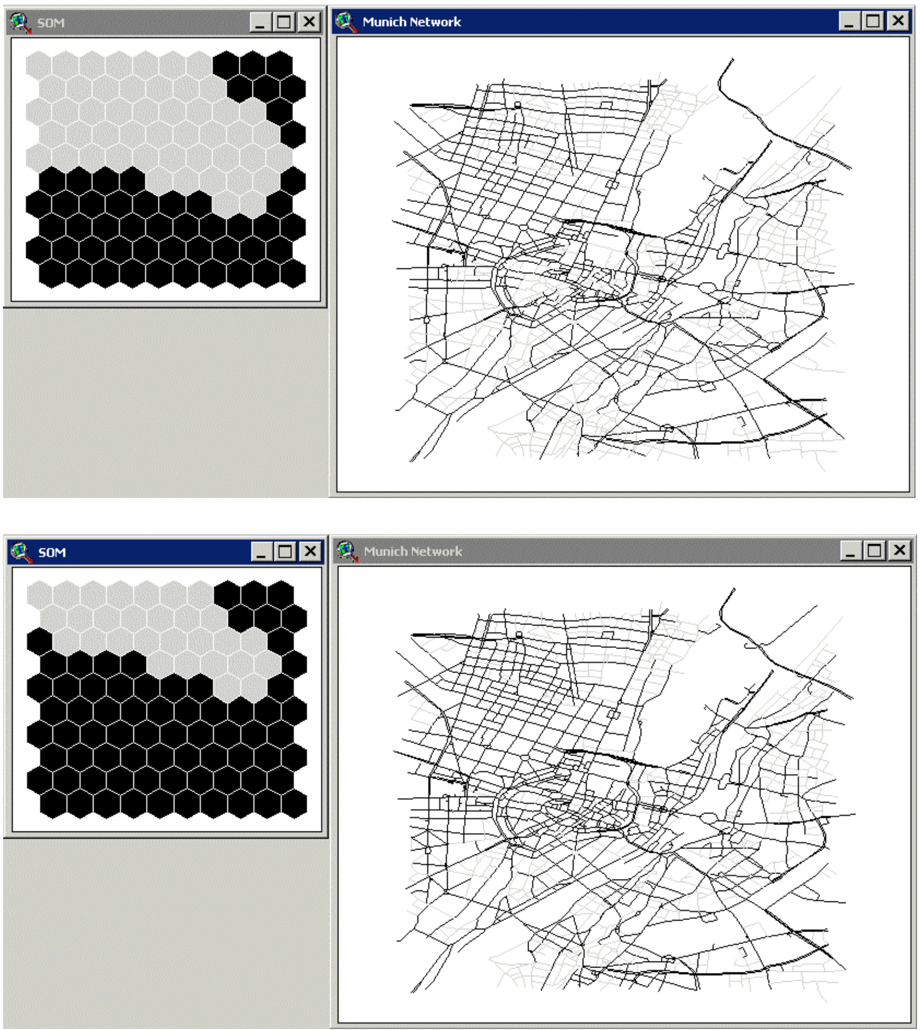

Figure 3: Two levels of detail of streets network. The streets that are related to the black neurons are selected.

From the case study, we have seen how the SOM-based approach can be used for selection of streets from a network. The approach is robust and flexible, as it considers multiple properties, and a dynamic linkage has been set up between the SOM and the corresponding network.

Acknowledgement

The dataset of Munich was provided by NavTech from the year 2000.

References:

Bjørke, J. T. (1996). Framework for Entropy-Based Map Evaluation. *Cartography and Geographic Information Systems*, Vol. 23, No. 2, pp. 78-95.

Hojholt P. (1995). Generalization of build-up areas using Kohonen-networks, *Proceedings of Eurocarto XIII*, 2-4 October, Ispra, Italy.

Jiang B. and Claramunt C. (2002), A Structural Approach to Model Generalisation of an Urban Street Network, *Proceedings of the 4th AGILE Conference on Geographic Information Science*, 24-27 April, Mallorca, Spain.

Kohonen T. (2001), *Self-Organizing Maps* (third edition), Springer, Berlin, Heidelberg, New York.

Li B. (1998), Exploring Spatial and Temporal Patterns with Self-Organizing Maps, *GIS/LIS'98*, Bethesda, Maryland: American Society for Photogrammetry and Remote Sensing, pp. 89-96.

Mackaness W. A. and Beard M. K. (1993), Use of Graph Theory to Support Map Generalisation, *Cartography and Geographic Information Systems*, Vol. 20, pp. 210 – 221.

Oja E. and Kaski S. (editors. 1999), *Kohonen Maps*, Elsevier.

Openshaw S., Blake M. and Wymer C. (1995), Using Neurocomputing Methods to Classify Britain's Residential Areas, (html paper), available at http://www.geog.leeds.ac.uk/papers/95-1/

Openshaw, S. (1994), Neuroclassification of spatial data, in Hewitson, B. C. and Crane, R. G. (eds), *Neural Nets: Applications in Geography*, Dordrecht: Kluwer Academic Publishers, 53-70.

Regnauld, N. (1996). Recognition of Building Clusters for Generalization. *Proceedings of the 7th Spatial Data Handling Symposium*, Delft, the Netherlands, pp. 185-198.

Richardson, D. E., and R. C. Thomson (2001). The 'Good Continuation' Principle of Perceptual Organization Applied to the Generalization of Road Networks. *Cartographica*.

Sester, M. (2001). Kohonen Feature Nets for Typification. *Fourth Workshop on Progress in Automated Map Generalization*, Beijing.

Shannon, C. E., and W. Weaver (1964). *The Mathematical Theory of Communication*, The University of Illinois Press.

Thomson R. C. and Richardson D. E. (1995), A Graph Theory Approach to Road Network Generalisation, in: *Proceeding of the 17th International Cartographic Conference*, pp. 1871 – 1880.

Töpfer, F., and W. Pillewizer (1966). The Principles of Selection. *The Cartographic Journal*, Vol. 3, No. 1, pp. 10-16.

Vesanto J., Himberg J., Alhoniemi E. and Parhankangas J. (2000), *SOM Toolbox for Matlab 5*, Report A57, Helsinki University of Technology, Libella Oy, Espoo.

Lars Harrrie
Has a Master of Science in Geodesy from Royal Institute of Technology, Stockholm, Sweden (1993). Since 1994 he has been active in Cartography and GIS at Lund University, Sweden. He completed his PhD 2001, with a thesis about optimisation techniques in cartographic generalisation. Is currently a post doctoral fellow at Lund university and has a part-time research position at Lantmäteriet (NMA in Sweden).

Bin Jiang
He is currently senior lecturer at the University of Gävle, Sweden. From 1997-1999, he was senior research fellow at the Centre for Advanced Spatial Analysis (CASA), University College London. His research interest involves generalization, geovisualization and spatio-temporal modelling or analysis in geographic information science.