The Architecture of Emergent Self-Organizing Maps to Reduce Projection Errors

Alfred Ultsch and Lutz Herrmann

Philipps University - Dept. of Mathematics and Computer Science
Marburg - Germany

Abstract. There are mainly two types of Emergent Self-Organizing Maps (ESOM) grid structures in use: hexgrid (honeycomb like) and quadgrid (trellis like) maps. In addition to that, the shape of the maps may be square or rectangular. This work investigates the effects of these different map layouts. Hexgrids were found to have no convincing advantage over quadgrids. Rectangular maps, however, are distinctively superior to square maps. Most surprisingly, rectangular maps outperform square maps for isotropic data, i.e. data sets with no particular primary direction.

1 Introduction

Emergent Self-Organizing Maps (ESOM) may be regarded as a non-linear projection technique using neurons arranged on a map. Preservation of topography of the high-dimensional input data onto the map is a primary aim of ESOM projections. There are mainly two types of ESOM grid structures in use: hexgrid (honeycomb like) and quadgrid (trellis like) maps. In addition to that, the shape may be chosen as square or rectangular. The performance of these map types are evaluated with Zrehen's Measure [12] for backward projection errors and a new defined measure for forward projection errors: Minimal U-Ranking.

2 Emergent Self Organizing Maps

This section reports the basic notations of SOM and ESOM [4]. Readers familiar with SOM may skip this section. A data space $D \subset \mathbb{R}^n$ is a metric subspace with distance measure $d : D \times D \rightarrow \mathbb{R}^+$. The training set $E = \{x_1, ...x_k\} \subset D$ consists of input samples presented during the SOM training algorithm. The map space M is a low-dimensional manifold embedded in $\mathbb{R}^l$, $l < n$, with a map distance

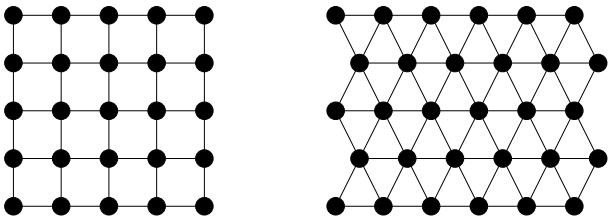

Fig. 1: Quadgrid vs. hexgrid map structures.

measure md. The Self-Organizing Map (SOM) is regarded as a set of neurons I. A neuron $i \in I$ is a tuple (w_i, p_i) consisting of a reference vector $w_i \in W$ and a position $p_i \in P$. The reference vectors $W \subset D$ are used for vector-quantization purposes, whereas the positions $P \subset M$ are used for vector-projection purposes $m : D \to P,\ m(x) = p_{bm(x)}$ with $bm(x) = argmin_{i \in I}\ d(x, w_i)$. The positions of the neurons are chosen such that for each neuron $i \in I$ there is a set of equidistant neighbours $N^M(i) \subset I$. This gives P the form of a regular grid in M. The most popular variants of grids are two-dimensional hexgrids (six immediate neighbours) and quadgrids (four immediate neighbours), i.e. see Figure 1.

The average of all data distances of a neuron's reference vector w_i to the reference vectors $\{w_j : j \in N^M(i)\}$ of its immediate neighbours is called U-height. A visualization of all U-heights for a given SOM is called U-matrix [7]. Most popular visualizations for U-matrices in two-dimensional map spaces are grey level or landscape pictures (see [10]). In the published applications of such SOM, two main types of SOM can be distinguished: first, SOM in which each neuron represents a cluster of input samples. These SOM can be thought of as a variant of the k-means clustering algorithm (see [3]). In these SOM, the number of neurons corresponds to the number of clusters assumed in the input data. Usually, this number is very small (≤ 20). In contrast to that, SOM may be used as tools for visualization of structural features of the data space. The structural features of the data space are usually visualized using U-Matrix or P-Matrix techniques [10]. A characteristic of this paradigm is the large number of neurons, usually several thousands (≥ 4000) of neurons. These SOM allow the emergence of intrinsic structural features of the data space on the map. They are called *Emergent Self-Organizing Maps* (ESOM) [8].

3 Measurements of topography preservation

There are two kinds of topographical errors with regard to ESOM projections: first, a pair of similar data points (x, y) is assigned to a distant pair of positions $(m(x), m(y)) = (p_i, p_j)$ on the map. This means that $d(x, y)$ is small and $md(m(x), m(y))$ is large. This type of error is called a forward projection error (FPE). Second, a pair of close neighbouring positions p_i, p_j is the image of a pair of distant data points $(x, y) = (m^{-1}(p_i), m^{-1}(p_j))$ in the data space. This is called a backward projection error (BPE). In the following, *topography preservation* refers to *preservation of topography by the bijective mapping* $m : W \to P$, $m(w_i) = p_i$ and $m^{-1}(p_i) = w_i$ that connects reference vectors and map space positions (see also [1]). There are regions in the data space where the probability density function becomes very small or even vanishes. These regions are called *gaps*. Gaps divide a data set into several classes of coherent elements. Gaps of the data space usually lead to mismatching distance-based neighbourhoods in data and map space, $N^M(i) \neq N^D(i)$ for $i \in I$, even if the topography is preserved perfectly. Therefore, topography preservation measures should distinguish between gaps and BPE. The most popular published measurements of topography preservation of ESOM are: Topographic Product, Topographic Function, C-

Measure, Minimal Pathlength, Minimal Wiring, Zrehen's Measure, Topological Index and Kaski's Trustworthiness Measure. For an overview see [1, 11]. All measures, except Topographic Function and Zrehen's Measure [12], are based on evaluation of distances or rankings of distances between pairs of neurons in the data space and on the map space. Therefore, these measures are subjective to location, scaling and variance of the input data. The Topographic Function works only if the training set E is sufficiently dense compared to the set of reference vectors W (see [1]). Zrehen's Measure is invariant to gaps but measures only BPE. In addition to Zrehen's Measure, a new gap-invariant FPE measure, called Minimal U-Ranking in analogy to Minimal Wiring, is defined in the following section.

4 The Minimal U-Ranking Measure

In order to avoid mismatching distance-based neighbourhoods of neurons in data and map space, the so-called *U-distance* is proposed as distance measure for the map space. Let $PATH_{ij}$ be the set of all arbitrary map space paths between neurons $i, j \in I$. The data space length of such a path is denoted *pathdistance*. The $udistance(i, j)$ for neurons $i, j \in I$ is defined as follows:

$$PATH_{ij} = \{\ (i_1, ..., i_n)\ :\ n \in \mathbb{N} \setminus \{0, 1\},\ i_1 = i,\ i_n = j,\ i_{k+1} \in N^M(i_k)\ \}$$

$$pathdistance(i_1, ..., i_n) = \sum_{k=1}^{n-1} d(w_{i_k}, w_{i_{k+1}})$$

$$udistance(i, j) = \min_{q \in PATH_{ij}} pathdistance(q)$$

U-distances are minimal-length paths in terms of data distances on the low-dimensional flexible net that is formed by the neurons in the data space. U-distances are used to define a rank-based topography on the set of neurons: Let $(udistance(i, i_1), ..., udistance(i, i_n))$ be the ordered sequence of all U-distances towards an arbitrary neuron $i \in I$ for all neurons in $\{i_1, ..., i_n\} = I$ and $udistance(i, i_k) \leq udistance(i, i_{k+1})$ for $k = 1, ..., n-1$. Then $urank_i(j) = r \in \{1, ..., n\}$ is the position of $udistance(i, j)$ in this sequence, which means $i_r = j$. The Minimal U-Ranking measure is defined as follows:

$$mur(i) = \sum_{j \in N^D(i)} urank_i(j)$$

Obviously, the Minimal U-Ranking rates the vector-projection's scattering of neigbourhood $N^D(i)$ onto the map space[1]. An additional normalization scheme may be used to scale the resulting values between 0 and 1, where 0 points to perfect neighbourhood preservation (see [2]). Gaps of the data space would influence the Minimal U-Ranking only if *md* was used instead of *udistance*. The Minimal U-Ranking rates the FPE of ESOM and combines well with Zrehen's Measure for BPE with regard to gaps in the data space.

[1] Usually the neigbourhood $N^D(i)$ is chosen as a $k \in \mathbb{N}$ sized data space neighbourhood $N_k^D(i)$ around neuron i, e.g. let k be $\approx 5\%$ of the size of I.

5 Data Sets

In order to demonstrate effects on a wide variety of different types of training set topographies, we have used four synthetic training sets and two real life data sets. The synthetic data sets are called Hexa, Atom, Ball and Chainlink. All data sets may be obtained on `http://www.mathematik.uni-marburg.de/~databionics`. **Hexa**: Three-dimensional points in 6 well separated clusters of equal size. **Chainlink**: 1000 three-dimensional points in two separated clusters of equal size; the clusters are not separable by a linear manifold. **Atom**: 800 three-dimensional points that are arranged in two separated classes of equal quantity: nucleus and shell. **Ball**: 800 three-dimensional points that are arranged in an *isotropic* set of points. **Iris**: the well known data set of Fisher. **Oliveoils**: 8 fat acids of 572 olive oils produced in nine different regions of Italy.

6 Experiments

Borderless ESOM with a toroid topology [10] are used for all experiments. Four different kinds of maps with nearly the same number of neurons are considered:

hexsquare	$64 \times 64 = 4096$ neurons with a hexgrid map
hexrectangular	$50 \times 82 = 4100$ neurons with a hexgrid map
quadsquare	$64 \times 64 = 4096$ neurons with a quadgrid map
quadrectangular	$50 \times 82 = 4100$ neurons with a quadgrid map

The learning rate is kept constant at 0.1 and the neighbourhood kernel has a regular *bubble* shape (see [2]). The neighbourhood kernel sizes were chosen such that the number of modified neurons are the same for each ESOM. All data sets are standardized. For each data set and ESOM type, the training was repeated 200 times. The resulting distributions were analyzed using Pareto Probability Density estimation (PDE) [9]. It was found that the median of the BPE and FPE of the repeated trainings can serve as an appropriate description of the average error. To rate the significance of differences, the Kolmogorov-Smirnov (KS) test on a 5% alpha level was used.

7 Results

Figure 2 shows the relative improvement by using different map architectures. For hexgrid vs. quadgrid maps, there is no general effect on the number or intensity of BPE. Six out of twelve KS tests show that quadgrid maps lead to significantly bigger error values whereas four tests show that quadgrid maps lead to smaller error values. In contrast, hexgrid maps lead to less or less intense FPE in nearly all cases (eleven of twelve tests). Seven out of twelve KS tests show a significantly decreasing number or intensity of BPE errors if one uses rectangular instead of square maps. Five tests show opposite results. This means that there is no significant effect on BPE that can be attributed to the shape of the ESOM.

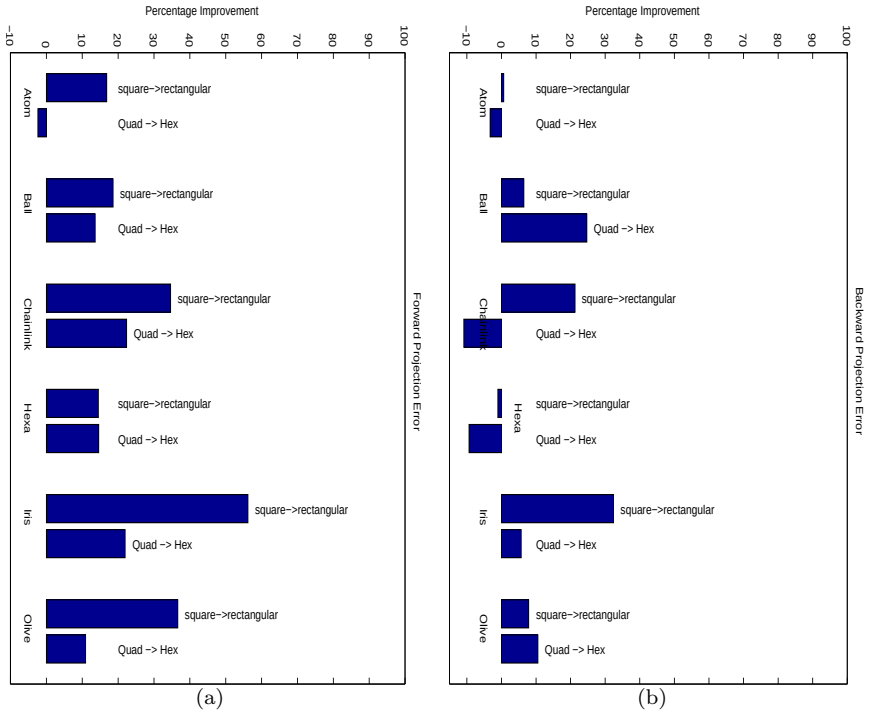

Fig. 2: Changes of median for (a) forward and (b) backward projection errors.

8 Discussion

Most measures for topography preservation misinterpret gaps of the data space as backward projection errors (BPE) of the ESOM. Therefore, we used Zrehen's Measure for BPE since it is gap invariant due to its definition based on the voronoi-tessellation of the data space. In order to measure forward projection errors (FPE) without a distortion by gaps in the data space, the Minimal U-Ranking measure was introduced here. Comparing hexgrid and quadgrid maps, there is no significant difference in BPE. Regarding FPE, hexgrids are slightly better. This may be attributed, however, to the Minimal U-Ranking itself: grids with more connections between adjacent nodes usually lead to shorter minimal-length paths between arbitrary nodes. Therefore, U-distances are, in general, smaller on hexgrid maps than they are on quadgrid maps. Despite of this, hexgrids showed no consistent advantage over quadgrids. The advantage of an elongation of one of the sides of the map is, however, convincing. This *elongation-effect* may be explainable by the so-called automatic dimension selection (see [5]). Scott compared the effect of using a hexgrid instead of a quadgrid for histograms in two dimensions [6]. He found that the effect is rather small: quadgrids are 2% less effective than hexgrids. This coincides with our observation that the error

reduction is much bigger using rectangular vs. square maps than using hexgrid vs. quadgrid maps.

9 Summary

In our experiments, ESOM built on rectangular map spaces produced smaller projection errors compared to ESOM with square map spaces. The usage of hexgrids vs. quadgrid maps, however, showed no or even worsening effects. The simplicity of subsequent display and processing with of-the-shelf programs therefore favours the usage of quadgrid maps. Still, there is an open question concerning isotropic data. Isotropic data sets have no principal component axis with prominent variance. Despite of this, topographic errors occur less often or less intense on rectangular shaped ESOM compared to square maps. Further research is necessary to understand the reasons behind this effect. Our upcoming research aims at determining hints for the right edge-to-edge ratio for different data sets. The isotropic data demonstrated that using the ratio of the first principal components is not sufficient for all data sets.

References

[1] H.-U. Bauer, M. Herrmann and Th.Villmann, Neural Maps and Topographic Vector Quantization, in *Neural Networks*, Edition 12, Elsevier, 1999.

[2] L. Herrmann, Selbstorganisation in Gitterstrukturen, Diploma-Thesis, Philipps-University of Marburg, 2003.

[3] S. Kaski, Data Exploration Using Self-Organizing Maps, PhD-Thesis, Helsinki University of Technology, 1997.

[4] T. Kohonen, Self-Organizing Maps, Springer, Berlin, 1997.

[5] H. Ritter and K. Schulten, Convergence Properties of Kohonen's Topology Conserving Maps: Fluctuations, Stability and Dimension Selection, in *Biological Cybernetics*, Edition 60, Springer, 1988.

[6] D. W. Scott, A Note on Choice of Bivariate Histogram Bin Shape, in *Journal of Official Statistics*, Vol. 4(1), Sweden, 1988.

[7] A. Ultsch and H. P. Siemon, Kohonen's Self Organizing Feature Maps for Exploratory Data Analysis, in *Proceedings of International Neural Network Conference (INNC'90)*, Dordrecht (Netherlands), Kluwer, 1990.

[8] A. Ultsch, Data Mining and Knowledge Discovery with Emergent Self-Organizing Feature Maps for Multivariate Time Series, in E. Oja and S. Kaski, editors, *Kohonen Maps*, 1999.

[9] A. Ultsch, Density Estimation and Visualization for Data containing Clusters of unknown Structure, In Proc. GfKl, Dortmund, 2004.

[10] A. Ultsch, Maps for the Visalization of high-dimensional Data Spaces, in *Proc. Workshop on Self organizing Maps*, pp 225 - 230, Kyushu, Japan, 2003.

[11] J.Venna and S. Kaski, Neighborhood preservation in nonlinear projection methods: An experimental study, in *Artificial Neural Networks - ICANN 2001*, Springer, Berlin, 2001.

[12] S. Zrehen, Analyzing Kohonen Maps With Geometry, in: *Proceedings of the International Conference on Artificial Neural Networks (ICANN'93)*, Springer, London, 1993.