ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 111-116

Benefits and Limits of the Self-Organizing Map and its Variants in the Area of Satellite Remote Sensoring Processing

Thomas Villmann*
Universität Leipzig, Klinik für Psychotherapie
D–04107 Leipzig, Karl–Tauchnitz–Str.25, Germany

Abstract. In the contribution we consider advantages and limits of the self-organizing maps in the area of satellite remote sensoring processing. Thereby we concentrate on the topology preservation property as well as the magnification control. We demonstrate the benifits for exemplary examples of LANDSAT-TM satellite images.

1. Introduction

The process of satellite remote sensoring usually is indicated by very large data sets, high-dimensional data spaces and correlated and noisy data. These facts prefer the application of neural maps for investigations. Thereby they may used as preprocessing tools as well as final applications [1, 10].

Self-organizing maps (SOM) [8] as special kind of neural maps project data from some (possibly high-dimensional) input space $\mathcal{V} \subseteq \Re^{D_\mathcal{V}}$ onto a position in some output space, such that a continuous change of a parameter of the input data should lead to a continuous change of the position of a localized excitation in the neural map.[1] This property of *neighborhood preservation* depends on an important feature of the SOM, its output space topology, which has to be specified prior to learning. Usually the output space $\mathcal{A}$ of the SOM is a $D_\mathcal{A}$-dimensional rectangular grid.(hypercube). If the topology, i.e. dimensionality and edge length ratios, of $\mathcal{A}$ does not match that of the data shape, neighborhood violations are inevitable [12]. A higher degree of topology preservation, in general, improves the accuracy of the map [3].

In the present paper we consider the advantages and limits of the SOM for applications in satellite remote sensoring processing systems and give some variants to improve the performance of the SOM. Thereby we emphasize the aspect of a continuous mapping, i.e. the control of the topology preservation.

2. SOMs as Topology Preserving Maps

We require that the output space $\mathcal{A}$ in SOM is a $D_\mathcal{A}$-dimensional rectangular grid of N neurons $\mathbf{r}$ which can in principle be of any dimensionality, or can extend to any dimension along its individual directions. This can be cast in a formal way by writing the output space positions as $\mathbf{r} = (i_1, i_2, i_3, ...)$, $1 < i_j < n_j$ with $N = n_1 \times n_2 \times ...$.[2] Associated to each neuron $\mathbf{r} \in \mathcal{A}$, is a weight vector $\mathbf{w_r}$ in $\mathcal{V}$. The map $\Psi_{\mathcal{V}\to\mathcal{A}}$ is realized by a winner take all rule

$$\Psi_{\mathcal{V}\to\mathcal{A}} : \mathbf{v} \mapsto \mathbf{s} = \operatorname*{argmin}_{\mathbf{r}\in\mathcal{A}} \|\mathbf{v} - \mathbf{w_r}\| \tag{1}$$

* *email: villmann@informatik.uni-leipzig.de*

[1] In this way the SOM determines the non-linear principle components of the data.

[2] Yet other arrangements are also admissible which can be described by a connectivity matrix. Here we only consider hypercubes.

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 111-116

whereas the back mapping is defined as $\Psi_{\mathcal{A}\to\mathcal{V}} : \mathbf{r} \mapsto \mathbf{w_r}$. Both functions determine the map $\mathcal{M} = (\Psi_{\mathcal{V}\to\mathcal{A}}, \Psi_{\mathcal{A}\to\mathcal{V}})$ realized by the network. All data points $\mathbf{v} \in \mathcal{V}$ which are mapped onto the neuron $\mathbf{r}$ perform its (masked) receptive field $\Omega_\mathbf{r}$. To achieve the map $\mathcal{M}$, SOMs adapt the pointer positions with respect to a presented sequence of data points $\mathbf{v} \in \mathcal{V}$ selected according to the data distribution $\mathcal{P}(\mathcal{V})$:

$$\triangle \mathbf{w_r} = \epsilon h_{\mathbf{rs}} (\mathbf{v} - \mathbf{w_r}) \ . \tag{2}$$

$h_{\mathbf{rs}}$ is the neighborhood function depending on the best matching neuron according to (1), usually chosen to be of Gaussian shape: $h_{\mathbf{rs}} = \exp\left(-\frac{\|\mathbf{r}-\mathbf{s}\|^2}{2\sigma^2}\right)$.

Topology preservation in SOMs is understood as preserving of the continuity of the mapping from the input space onto the output space, exactly spoken: it is equivalent to the *continuity* of $\mathcal{M}$ between the *topological spaces* with properly chosen metric in both $\mathcal{M}$ and $\mathcal{V}$. Because of the lack of space we refer to [12] for a detailed consideration. The topology preserving property one can use for immediately evaluations of the resulted map, for instance for interpretation as color space as demonstrated in sec. 3. . On the other hand topology preservation allows the applications of interpolating schemes like the parametrized SOM (PSOM) [11] or interpolating SOM (I-SOM) [5]. Several approaches were developed to judge the degree of topology preservation for a given map. Here we shortly give a variant $\tilde{P}$ of the well known topographic product P [3] which uses, instead of the Euclidean distances between the weight vectors, the respective distances $d^{\mathcal{G}_\mathcal{V}}(\mathbf{w_r}, \mathbf{w_{r'}})$ of minimal path length in the induced Delaunay-graph $\mathcal{G}_\mathcal{V}$ of the $\mathbf{w_r}$. $\mathcal{G}_\mathcal{V}$ corresponds to the Voronoi tessellation of $\mathcal{V}$ by the masked receptive fields $\Omega_\mathbf{r}$. During the computation of $\tilde{P}$ for each node $\mathbf{r}$ the sequences $\mathbf{n}_j^\mathcal{A}(\mathbf{r})$ of j-th neighbors of $\mathbf{r}$ in $\mathcal{A}$ and $\mathbf{n}_j^\mathcal{V}(\mathbf{r})$ describing the j-th neighbor of $\mathbf{w_r}$, have to be determined. These sequences and further averaging over neighborhood orders j and nodes $\mathbf{r}$ finally leads to

$$\tilde{P} = \frac{1}{N(N-1)} \sum_\mathbf{r} \sum_{j=1}^{N-1} \frac{1}{2j} \log\left(\Pi_{l=1}^{j} \frac{d^{\mathcal{G}_\mathcal{V}}\left(\mathbf{w_r}, \mathbf{w}_{\mathbf{n}_l^\mathcal{A}(\mathbf{r})}\right)}{d^{\mathcal{G}_\mathcal{V}}\left(\mathbf{w_r}, \mathbf{w}_{\mathbf{n}_l^\mathcal{V}(\mathbf{r})}\right)} \cdot \frac{d_\mathcal{A}\left(\mathbf{r}, \mathbf{n}_l^\mathcal{A}(\mathbf{r})\right)}{d_\mathcal{A}\left(\mathbf{r}, \mathbf{n}_l^\mathcal{V}(\mathbf{r})\right)} \right) \ . \tag{3}$$

$\tilde{P}$ can take on positive or negative values: If $\tilde{P} < 0$ holds the output space is too low-dimensional, in contrast if we have $\tilde{P} > 0$ the output space too high-dimensional. In both cases neighborhood relations are violated. Only if $\tilde{P} \approx 0$ is valid the output space approximately matches topology of input data. The present variant $\tilde{P}$ of the overcomes the problem of strongly curved maps which may be judged to be neighborhood violating by the original P, even though the shape of the map might be perfectly justified [12].

2.1. Growing SOM for Structure Adaptation

The growing SOM (GSOM) approach is an extension of the usual SOM [4]. Its output $\mathcal{A}$ are hypercubes whereby during the learning procedure additionally to the weight vector learning both the dimension and the respective length ratios are adapted, i.e. we allow a variable overall dimensionality and variable dimensions along the individual directions in the hypercube.

The GSOM starts from an initial 2-neuron chain, learns according the regular SOM-algorithm, adds neurons to the output space with respect to a certain criterion to be described below, learns again, adds again, etc., until a prespecified maximum number $N_{\max}$ of neurons is distributed. During this procedure,

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 111-116

the output space topology remains to be of the form $n_1 \times n_2 \times ...$, with $n_j = 1$ for $j > D_{\mathcal{A}}$, where $D_{\mathcal{A}}$ is the current dimensionality of $\mathcal{A}$.[3] From there it can grow either by adding nodes in one of the directions which are already spanned by the output space or by initilizing a new dimension.

This decision is made on the basis of the receptive fields $\Omega_{\mathbf{r}}$. When reconstructing $\mathbf{v} \in \mathcal{V}$ from neuron $\mathbf{r}$, an error $\theta = \mathbf{v} - \mathbf{w_r}$ remains decomposed along the different directions, which result from projecting back the output space grid into the input space $\mathcal{A}$:

$$\theta = \mathbf{v} - \mathbf{w_r} = \sum_{i=1}^{D_{\mathcal{A}}} a_i(\mathbf{v}) \frac{\mathbf{w_{r+e_i}} - \mathbf{w_{r-e_i}}}{\| \mathbf{w_{r+e_i}} - \mathbf{w_{r-e_i}} \|} + \mathbf{v}' \tag{4}$$

Thereby, $\mathbf{e}_i$ denotes the unit vector in direction i of $\mathcal{A}$.[4] Considering a receptive field $\Omega_{\mathbf{r}}$ and determining their first principle components $\boldsymbol{\omega}_{PCA}$ allows a further decomposition of $\mathbf{v}'$. Projection of $\mathbf{v}'$ onto the direction of $\boldsymbol{\omega}_{PCA}$ then yields $a_{D_{\mathcal{A}}+1}(\mathbf{v})$,

$$\mathbf{v}' = a_{D_{\mathcal{A}}+1}(\mathbf{v}) \frac{\boldsymbol{\omega}_{PCA}}{||\boldsymbol{\omega}_{PCA}||} + \mathbf{v}''. \tag{5}$$

The *criterion for the growing* now is to add nodes in that direction which has on average the largest error (normalized) expected amplitudes $\tilde{a}_i$:

$$\tilde{a}_i = \sqrt{\frac{n_i}{n_i + 1}} \sum_{\mathbf{v}} \frac{\mid a_i(\mathbf{v}) \mid}{\sqrt{\sum_{j=1}^{D_{\mathcal{A}}+1} a_j^2(\mathbf{v})}}, \quad i = 1, ..., D_{\mathcal{A}} + 1 \tag{6}$$

After each growth step, a new learning phase has to take place, in order to readjust the map. For a detailed study of algorithm we refer to [4].

2.2. Magnification Control in SOM

The usual SOM distributes the pointers $\mathbf{W} = \{\mathbf{w_r}\}$ according to the input distribution: $\mathcal{P}(\mathcal{V}) \sim \mathcal{P}(\mathbf{W})^{\alpha}$ with the magnification factor $\alpha = \frac{2}{3}$.[5] BAUER ET AL. in [2] introduced a local learning parameter $\epsilon_{\mathbf{r}}$ with $\langle \epsilon_{\mathbf{r}} \rangle \propto \mathcal{P}(\mathcal{V})^m$ in (2) which now reads as

$$\triangle \mathbf{w_r} = \epsilon_{\mathbf{s}} h_{\mathbf{rs}} (\mathbf{v} - \mathbf{w_r}) \ . \tag{7}$$

This local learning rule leads to a relation $\mathcal{P}(\mathcal{V}) \sim \mathcal{P}(\mathbf{W})^{\alpha'}$ with $\alpha' = \alpha(m+1)$ and, hence, allows a magnification control. Especially, one can achieve a resolution $\alpha' = 1$ which maximizes transinformation [9, 13].

[3] Hence, the initial configuration is $2 \times 1 \times 1 \times ...,\ D_{\mathcal{A}} = 1$.

[4] At the border of the output space grid, where not two, but just one neighboring neuron is available, we use $\frac{\mathbf{w_r} - \mathbf{w_{r-e_i}}}{||\mathbf{w_r} - \mathbf{w_{r-e_i}}||}$ or $\frac{\mathbf{w_{r+e_i}} - \mathbf{w_r}}{||\mathbf{w_{r-e_i}} - \mathbf{w_r}||}$ to compute the backprojection of the output space direction $\mathbf{e}_i$ into the input space.

[5] This result is valid for the one-dimensional case and higher dimenional ones which separate.

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 111-116

3. Application to Satellite Remote Sensoring

Satellites of LANDSAT-TM type produce pictures of the earth in 7 different spectral bands. The ground resolution in meter is 30×30 for the bands 1-5 and band 7. Band 6 (thermal band) has a resolution of 60×60 only and, therefore, it is often dropped. The spectral bands represent useful domains of the whole spectrum in order to detect and discriminate vegetation, water, rock formations and cultural features [12]. Hence, the spectral information, i.e., the intensity of the bands associated with each pixel of a LANDSAT scene is represented by a vector $\mathbf{v} \in V \subseteq \Re^{D_\mathcal{V}}$ with $D_\mathcal{V} = 6$. The aim of any classification algorithm is to subdivide this data space into subsets of data points which belong to a certain category corresponding to a specific feature like wood, industrial region, etc., each feature being specified by a certain prototype data vector. One way to get good results for visualization is to use a SOM dimension $D_\mathcal{A} = 3$ [7]. Then we are able to interpret the positions of the neurons $\mathbf{r}$ in the lattice A as vectors $\mathbf{r} = \mathbf{c} = (r, g, b)$ in the color space $\mathcal{C}$, whereby r, g, b are the intensity of the colors red, green and blue [7]. This assigns colors to winner neurons so that we end up immediately with the pseudo color version of the original picture for visual interpretation. However, since we are mapping the data clouds from a 6-dimensional input space onto a *three*-dimensional color space there may arise dimensional conflicts and the visual interpretation may fail. Usually, for visual interpretation only the band 2, 3, 4 are used which means a loss of information.

In the first example we investigated a picture of the north–east region of Leipzig[6]. For comparison we also trained several usual SOMs with fixed output spaces and determined the respective $\tilde{P}$–values (3) which are depicted in Tab. 1. The topographic product $\tilde{P}$ prefers a output space dimension $D_\mathcal{A}$ between 2 and 3. However a clear decision can not be made. An additional Grassberger-Procaccia-analysis [6] yields $D_\mathcal{A}^{\mathcal{GP}} \approx 1.7$. Here should be mentioned that we applied in this runs, as well as in the further GSOM-simulations, instead the original learning rule (2) the new one (7) to achieve a maximum of transinformation as pointed out in sec.2.2. .

The GSOM algorithm was applied in several runs (10^6 lernsteps) with different values $N_{\max}$. The obtained results are depicted in Tab. 2 and Fig. 1 . The achieved $\tilde{P}$–values are better than the respective values for the fixed lattice structures. Furthermore, for all numbers $N_{\max}$ of maximal allowed neurons, we achieved approximately the same structure and the length ratio's of the edges shows a good agreement with considerations of the usual principle component analysis (PCA) of the data space:

[6] obtained from UMWELT-FORSCHUNGSZENTRUM Halle-Leipzig, Germany

N	lattice structure	$\tilde{P}$
256	256	-0.189 ± 0.00612
256	16×16	-0.0642 ± 0.00031
252	$7 \times 6 \times 6$	$+0.0282 \pm 0.00024$
256	$4 \times 4 \times 4 \times 4$	$+0.0816 \pm 0.00387$

Table 1: Table of the topographic product $\tilde{P}$ for different but fixed output spaces for the LANDSAT satellite image of Leipzig. For each structure over 3 runs was averaged.

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 111-116

$$\mathbf{ev} = (274.06, 76.19, 39.78, 11.92, 8.27, 6.28)^T \tag{8}$$

as the vector of the respective eigenvalues. However, in general, the usual but linear PCA fails, as shown in second example. It is again a LANDSAT-scene, but now from the Colorado-area[7]. The respective PCA yields

$$\mathbf{ev} = (4.93121, 0.6838, 0.29047, 0.05474, 0.02242, 0.01737)^T \tag{9}$$

suggesting an one-dimensional structure, whereas a Grassberger-Procaccia-analysis [6] gives $D_{\mathcal{A}}^{\mathcal{GP}} \approx 3.1414$. The GSOM generates a $12 \times 7 \times 3$ lattice structure ($N_{\max} = 256$) which corresponds to a $\tilde{P}$-value of 0.0095 indicating again a good topology preservation.

References

[1] M. F. Augusteijn, K. A. Shaw, and R. J. Watson. A study of neural network input data for ground cover identification in satellite images. In S. Gielen and B. Kappen, editors, *Proc. ICANN'93, Int. Conf. on Artificial Neural Networks*, pages 1010–1013, London, UK, 1993. Springer.

[2] H.-U. Bauer, R. Der, and M. Herrmann. Controlling the magnification factor of self–organizing maps. *Neural Computation*, 8(4):757–771, 1996.

[3] H.-U. Bauer and K. R. Pawelzik. Quantifying the neighborhood preservation of Self-Organizing Feature Maps. *IEEE Trans. on Neural Networks*, 3(4):570–579, 1992.

[4] H.-U. Bauer and T. Villmann. Growing a Hypercubical Output Space in a Self–Organizing Feature Map. *IEEE Transactions on Neural Networks*, 8(2):218–226, 1997.

[5] J. Goppert and W. Rosenstiel. The continuous interpolating self-organizing map. *Neural Processing Letters*, 5(3):185–92, 1997.

[6] P. Grassberger and I. Procaccia. Maesuring the strangeness of strange attractors. *Physica*, 9D:189–208, 1983.

[7] M. H. Gross and F. Seibert. Visualization of multidimensional image data sets using a neural network. *Visual Computer*, 10:145–159, 1993.

[8] T. Kohonen. *Self-Organizing Maps*. Springer, Berlin, Heidelberg, 1995. (Second Extended Edition 1997).

[7] Thanks to M. Augusteijn (Univerity of Colorado) for providing this image.

$N_{\max}$	lattice structure	$\tilde{P}$
128	$12 \times 5 \times 2$	0.0047
256	$14 \times 6 \times 3$	0.0050
512	$15 \times 6 \times 4$	0.0051

Table 2: Table of the results of several GSOM-runs for the picture of Leipzig.

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 111-116

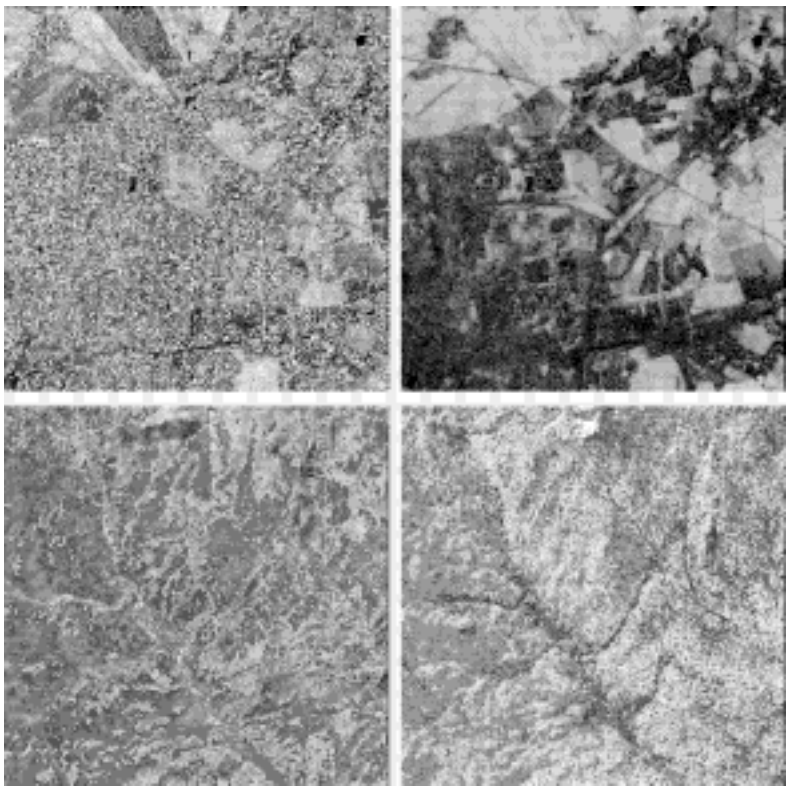

Figure 1: Pseudo-color imgages of LANDSAT-TM six-band spectral images: above - Leipzig image using the $7 \times 6 \times 6$ standard SOM (left) and the $14 \times 6 \times 3$ GSOM-solution; bottom - Colorado image using the standard psuedo-color vizualization using only bands 2, 3, 4 (left) and the $12 \times 7 \times 3$ GSOM-solution;

[9] R. Linsker. How to generate maps by maximizing the mutual information between input and output signals. *Neural Computation*, 1:402–411, 1989.

[10] E. Merenyi. Self-organizing ANN for planetary surface composition research. In M. Verleysen, editor, *Proceedings ESANN'98*, pages 197–202, Brussels, Belgium, 1998. D-facto publications.

[11] H. Ritter. Parametrized self-organizing maps. In S. Gielen and B. Kappen, editors, *Proc. ICANN'93 Int. Conf. on Artificial Neural Networks*, pages 568–575, London, UK, 1993. Springer.

[12] T. Villmann, R. Der, M. Herrmann, and T. Martinetz. Topology Preservation in Self–Organizing Feature Maps: Exact Definition and Measurement. *IEEE Transactions on Neural Networks*, 8(2):256–266, 1997.

[13] T. Villmann and M. Herrmann. Magnification control in neural maps. In *Proc. Of European Symposium on Artificial Neural Networks (ESANN'98)*, pages 191–196, Brussels, Belgium, 1998. D facto publications.