ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 191-196

Magnification Control in Neural Maps

Thomas Villmann[1] and Michael Herrmann[2]

[1]Universität Leipzig, Klinik für Psychosomatische Medizin,
Karl-Tauchnitz-Str. 25, 04107 Leipzig, F.R.G.
[2]MPI für Strömungsforschung, Bunsenstr. 10, 37073 Göttingen, F.R.G.

Abstract. In self-organizing maps reconstruction error minimization and topology preservation have shown to be conflicting goals. For one dimensional maps this dilemma can be alleviated e.g. by locally adaptive learning rates. On the other hand, the neural gas algorithm and its topology representing extension allow for vector quantization at theoretically optimal reconstruction error for arbitrary data dimensionality. Thus, it is possible to modify the neural gas algorithm such as to meet optimality criteria other than mean square error in an exact way for data dimensions greater than one.

1. Introduction

Vector quantization aims at representing data distributions by prototypical data vectors. Once the prototypes are fixed, the data can be addressed and transmitted by referring to the prototype labels as a code. The quality of a vector quantizer can be evaluated by the mean error occurring when reconstructing the data from the code by the receiver. If the code itself is subject to noise during transmission, neighborhood relations in the code space can be exploited to partly preserve information of the data as done in topographic vector quantization.

For self-organizing maps (SOMs) [4] it is known [7] that the ordering process of the reference vectors interferes with the representation quality of the vector quantizer. This leads to data representations that are suboptimal with respect to mean square reconstruction error (MSE). In contrast, the neural gas (NG) algorithm yields quantizations that approximate minimal-error configurations. Whereas for SOM analytical results have been obtained essentially only for one dimensional maps, MSE-optimality has been shown in NG for arbitrary data dimensionality.

In application to signal processing, control and robotics also optimality criteria other than MSE turned out to be relevant. In particular, it may be important to resolve rarely occurring critical states of a controlled system more detailed than by common statistical vector quantizers, in contrast to frequently visited states which may not call for refined control. On the other hand, in data visualization or in representations of positional information an even partition of the data set into subregions may be desirable. Thus, controlling the resolution

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 191-196

properties of a vector quantizer or its *magnification* serves to the algorithm's flexibility.

2. Magnification of neural maps

Neural maps project data vectors $\mathbf{v} \in M$ sampled from a data distribution $P(\mathbf{v})$ onto a set A of neurons i. The configuration of the map is denoted by $\mathbf{w} = \{\mathbf{w}_i\}_{i \in A}$, where $\mathbf{w}_i$ are pointers associated to each neuron. An input vector $\mathbf{v}$ is mapped onto best matching neuron s according to $\mathbf{v} \mapsto s(\mathbf{v}) = \text{argmin}_{i \in A} \|\mathbf{v} - \mathbf{w}_i\|$. The subset of the input space $R_i = \{\mathbf{v} \mid s(\mathbf{v}) = \text{argmin}_{i \in A} \|\mathbf{v} - \mathbf{w}_i\|\}$ forms the receptive field of neuron i. Neural pointers are adapted according to

$$\triangle \mathbf{w}_i = \epsilon h_\lambda (i, \mathbf{v}, \mathbf{w}) (\mathbf{v} - \mathbf{w}_i), \tag{1}$$

where interactions among neurons are included via a neighborhood function

$$h_\lambda (i, \mathbf{v}, \mathbf{w}) = \exp(-D_A(i - s(\mathbf{v}))/\lambda). \tag{2}$$

D_A denotes a distance measure in A. The SOM is characterized by a squared distance of vector indices i and s: $D_A^{SOM}(i - s(\mathbf{v})) = \|i - s(\mathbf{v})\|_A^2$. $2\sigma^2$ is used instead of λ. In the NG algorithm the distance is based on rankings of distances in the input space: $D_A^{NG}(i - s(\mathbf{v})) = k_i(\mathbf{v}, \mathbf{w})$ [6].

The definition of the distance measures in the two models gives rise to learning rules with different properties. The NG adaptation rule follows on average a potential dynamics. In contrast, for the SOM-algorithm no global energy function exists. Further, in [6] the convergence rate of the neural gas has been shown to be faster than that of the SOM.

SOM and NG have different resolution properties. These are specified as the relation of $P(\mathbf{v})$ and the pointer density $\rho(\mathbf{w})$ which can be expressed in many instances as a power law:

$$P(\mathbf{w})^\alpha \propto \rho(\mathbf{w}). \tag{3}$$

α is called *magnification exponent*. In eq. (3) and later on it will be assumed that either the set of pointers form a continuous manifold or that $P(\mathbf{v})$ is approximately constant inside each receptive field. For $\alpha = d/(d+p)$ a vector quantizer minimizes the p-norm distortion error [8]

$$E_p = \int_M \|\mathbf{w}_s - \mathbf{v}\|^p P(\mathbf{v}) \, d\mathbf{v}. \tag{4}$$

This includes the maximum error $p = \infty$ corresponding to $\alpha = 0$. If $\alpha = 1$ then the mutual information between P and ρ is maximized.

For one dimensional SOMs $\alpha_{SOM} = \frac{2}{3}$ holds in the limit $1 \ll \sigma \ll N$ [7]. For higher dimensions this result is valid only for separable cases [7], but a general result is unknown. For very small values of σ the magnification rate

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 191-196

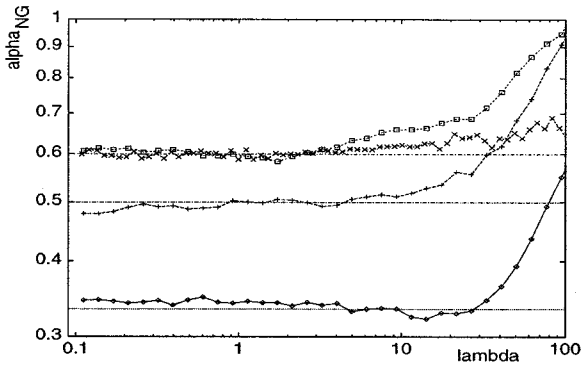

Figure 1: *The magnification exponent α_{NG} for input spaces of different dimension ($d = 1$ ($\diamond$), $d = 2$ ($+$), $d = 3$ ($\square$ and $\times$)) in dependence on the neighborhood range λ. The networks consisted of $N = 512$ neurons, except for the ($\times$) curve, where $N = 4096$ neurons have been trained. The horizontal lines indicate the theoretical values $\alpha_{NG} = 1/3,\ 1/2,\ 3/5$ for $d = 1,\ 2,\ 3$, resp., which can be obtained analytically in the limit $\lambda \to 0$.*

approaches the value $\alpha = \frac{1}{3}$ [3], as the neighborhood ceases to be effective, which in turn causes topological defects for $d > 1$.

For the NG the magnification exponent was shown [6] to obey $\alpha_{NG} = d/(d+2)$ at small λ. This result is valid for arbitrary data dimensions [6] and coincides with the figure obtained for optimal Euclidean vector quantizers [8].

It is crucial for further applications to determine the validity of this result for λ larger than zero. For this purpose, we have performed simulations with network sizes $N = 2^k$, $k = 4, \ldots 14$, input dimensions $d = 1, 2, 3$, and various forms of the input distribution, cf. Fig. 1. When decreasing λ, the optimal magnification rate is approached closely if $\lambda < \nu$, where ν depends on d and on the variation of $P(\mathbf{v})$. Generally, the MSE is minimized by neural gas maps already at nonzero neighborhood parameters.

3. Magnification control

As suggested in [1] we modify the learning parameter ϵ adaptively and locally. The learning rate at the winner $\epsilon_{s(\mathbf{v})}$ is used for all neurons in the current learning step. The new parameters $\epsilon_i = \epsilon(\mathbf{w}_i)$ are intended to effectively modify the input density, and are chosen, hence, dependent on $P(\mathbf{v})$ at the position of the weight vectors $\mathbf{w}_i$ via

$$\langle \epsilon_i \rangle = \epsilon_0 P(\mathbf{w}_i)^m \ , \tag{5}$$

where the brackets $\langle \ldots \rangle$ denote the average in time.

The new parameter m allows to deliberately control the magnification rate

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 191-196

of the evolving map. When considering the short time average of eq. (1)

$$\langle \triangle \mathbf{w}_i \rangle = \epsilon_0 \int_{\mathcal{V}} \epsilon_{s(\mathbf{v})} h(i, \mathbf{v}, \mathbf{w}) (\mathbf{v} - \mathbf{w}_i) \; P(\mathbf{v}) \, d\mathbf{v}, \tag{6}$$

we find that eq. (5) results in a modification of the input space density $P(\mathbf{v})$ $\longrightarrow$ const $P(\mathbf{v})^{m+1}$. The convergence properties of the original algorithms are retained when the common learning rate ϵ_0 decreases algebraically. Following the derivation of the SOM and NG magnification exponents in [7] and [6], resp., we get for the new magnification exponent

$$\alpha' = \alpha \cdot (m+1) \; , \tag{7}$$

which allows an explicit control of α' when inserting on the right hand side the intrinsic exponent of the algorithm α_{SOM} or α_{NG}, resp. For SOM the application of (7) is restricted to the analytically treatable one dimensional case [1], but also typical instabilities of high dimensional SOMs would render such a map to be unstable for some range of the modified learning parameters. For the NG rule (7) is valid for any data dimension.

Since $P(\mathbf{v})$ is assumed to be unknown, the information already acquired by the network is used:

$$P(\mathbf{w}_i) \propto p_i \rho(\mathbf{w}_i) \; . \tag{8}$$

Here, p_i is the probability that i is the winner neuron. Using this ansatz, relation (5) is *approximated* by

$$\epsilon_s(t) = \epsilon_0 \left(\frac{1}{\triangle t_s} \left(\frac{1}{\|\mathbf{v} - \mathbf{w}_s\|^d} \right) \right)^m \; , \tag{9}$$

with s being the best-matching neuron for the present stimulus $\mathbf{v}$ and $\triangle t_s$ is the time difference between the present t-value and the last time when this neuron has been winner.

4. Numerical results

Here we only report numerical results obtained by the NG algorithm. For the SOM we refer to [1]. In our simulations we used the eq. (9) to approximate the relation (5) for $\epsilon_s(t)$. In addition, we prespecified an upper bound $\epsilon_{\max}$. The data were chosen with respect to the density function $g(x) = \sin(\pi \cdot x)$, and $x \in [0,1]$ was equally distributed. The map with $N = 50$ neurons was trained for 10^7 adaptation steps with λ decreasing towards one. The two- and three dimensional data distributions were defined in an analogous way. After the training procedure the entropy of the map

$$H = -\sum_{i=1}^{N} \hat{P}(\mathbf{w}_s) \cdot \log\left(\hat{P}(\mathbf{w}_s)\right) \; . \tag{10}$$

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 191-196

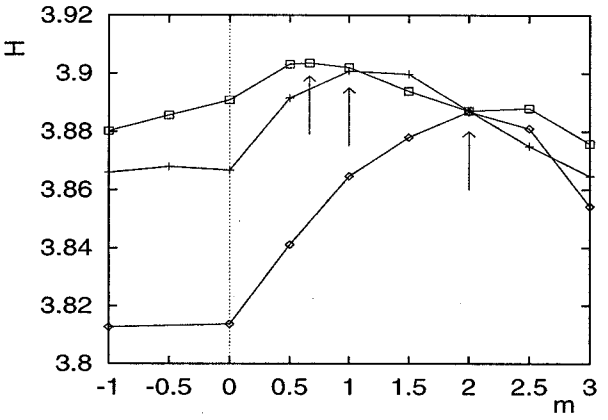

Figure 2: *Plot of the entropy H for maps trained with different magnification control parameters m ($d = 1$ ($\diamond$), $d = 2$ ($+$), $d = 3$ ($\square$)). The arrows indicate the theoretical values of m ($m = 2$, $m = 1$, $m = 2/3$, resp.) which maximize the entropy of the map.*

was determined. H is expected to have a maximum for the $\alpha' = 1$, i.e. for $m_1 = 2$, $m_2 = 1$, $m_3 = \frac{2}{3}$, cf. eq. (7).

The results are depicted in Fig. 2. For an ($\alpha = 1$)-map the maximal entropy value is given by $H_{\max}(N) = -\sum_{i=1}^{N} \frac{1}{N} \cdot \log\left(\frac{1}{N}\right)$, i.e. in the example we have $H_{\max}(50) \approx 3.912$. Hence, in the cases of the optimal magnification correctors the resulted maps approximate the theoretical value $H_{\max}$ best, whereas for other m-values the obtained entropies are significantly smaller. The magnification rates obtained for $d = 1$ and $d = 2$ are presented in Fig. 3.

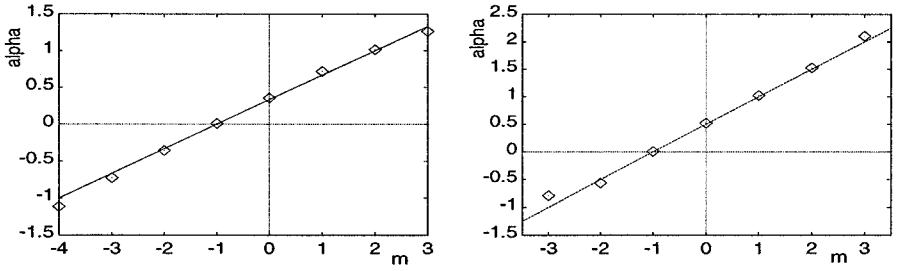

Figure 3: *Effect of magnification control in the neural gas algorithm (left $d = 1$, right $d = 2$).*

5. Discussion

Considering accuracy of the obtained magnification rate, it must be noted that the exponent is numerically difficult to be determined precisely, due to fluctuations of the pointers and boundary effects. Numerical simulations indicate that for decreasing learning rate the accuracy of α improves by a power law with an exponent smaller than one, depending on the algorithm and input di-

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 191-196

mension. Thus, long learning rate are needed to reproduce nicely the predicted values. As this holds already for the original algorithms their controlled versions cause even longer runs due to the necessity of additional averaging, such as in eq. (9). As for non trivial data distributions large learning rates may cause topological distorsions, a further advantage of NG networks compared to SOM becomes apparent, although the reliability of the learned connectivity in topology representing networks may suffer at large ϵ.

Due to the lack of space we cannot compare here to results obtainable by other special algorithms for neural vector quantization. However, compared to the basic k-means algorithm, which yields error-optimal partitions as the NG does, the neighborhood interaction speeds up the global reordering of the pointers necessary for changing α at least by an order of magnitude. Further, effects of approximations concerning the form of the receptive fields implicit in (9) are reduced by λ-induced local averaging.

In the present paper we reported a approach which allows to control neural maps like SOM and NG to produce an optimal vector quantizer. The scheme relies on an effective modification of the input distribution performed by a position dependent adaptation of the learning rate. In neural gas maps the results are valid for arbitrary data dimension, which, however, has to be known in advance, but may be determined even for heterogeneous data sets e.g. by local principal component analysis [2].

References

[1] H.-U. Bauer, R. Der, M. Herrmann (1996) Controlling the Magnification Factor of Self-Organizing Feature Maps. *Neural Computation* **8**, 757-771.

[2] J. Bruske, G. Sommer (1997) Topology Representing Networks for Intrinsic Dimensionality Analysis. *Artificial Neural Networks - ICANN'97*, W. Gerstner e.a. (eds.), Springer, Berlin, 595-600.

[3] E. deBodt, M. Verleysen, M. Cottrell (1997) Kohonen Maps vs. Vector Quantization for Data Analysis. *Proc. ESANN'97*, D facto, Brussels, 211-218.

[4] T. Kohonen (1995) *Self-Organizing Maps.* Springer-Verlag, Berlin.

[5] Th. Martinetz, K. Schulten (1994) Topology Representing Networks. *Neural Networks,* **7(3)**, 507-522.

[6] T. M. Martinetz, S. G. Berkovich, K. J. Schulten (1993) 'Neural-Gas' Network for Vector Quantization and its Application to Time-Series Prediction. *IEEE Transact. Neur. Netw.* **4**:4, 558-569.

[7] H. Ritter, K. Schulten (1986) On the Stationary State of Kohonen's Self-Organizing Sensory Mapping. *Biological Cybernetics* **54**, 99-106.

[8] Zador, P. L. (1982) Asymptotic Quantization Error of Continuous signals and Quantization Dimension. *IEEE Transact. Inform. Th.*, **28**:2, 139-149.