ESANN'1994 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 20-21-22 April 1994, D-Facto public., ISBN 2-9600049-1-4, pp. 251-255

Improvement of learning results of the selforganizing map by calculating fractal dimensions

Heike Speckmann, Günter Raddatz, Wolfgang Rosenstiel

Lehrstuhl für Technische Informatik,
Sand 13, 72076 Tübingen, Germany
speckman@peanuts.informatik.uni-tuebingen.de

Abstract. In this paper we present a possibility to bring up the learning results for different data sets which earlier were difficult or impossible to learn. We choose the calculated fractal dimension of the data set as the dimension of the selforganizing map for guaranteeing the maps ability of topology preserving. Furthermore we explore different states of the learning process and the final map for its fractal dimension and get interesting results.

1 Introduction

Kohonen's selforganizing map (SOM) which has been introduced by T. Kohonen [1] has many applications. So the SOM's algorithm is an efficient alternative to traditional signal processing.

But there is still one big problem. There exists no common valid prove of convergence for this efficient algorithm. A couple of theoretic works taggle with this problem (e.g. [2]). But they are only valid in special cases:

- The input data have to be equally distributed.
- The map's state must be near an equilibrium state means the adaptation factor has to be already small.
- The map's dimension is one.

In our opinion the input data set's properties have to be considered. So there exists data sets which are very easy to learn using a two dimensional map, e.g. gas spectra [3] or VLSI process data [4]. Others are difficult or impossible to learn, e.g. EEG data or pictures of PCB layouts what we are currently trying. Probably these data sets require a greater dimension of the map.

The SOM's main property is the nonlinear projection to the principal manifolds means finding a n-dimensional layer to approximate the input data. This layer, represented by the geometrical dimension of the map is lower dimensional than the input data set. If the dimensionality of the map is too low, the map tries to approximate the higher dimension by folding itself into the input space. So we have to determine the necessary dimensionality of the input space to avoid this violation of topology preserving. Second a correct dimension of the map

ESANN'1994 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 20-21-22 April 1994, D-Facto public., ISBN 2-9600049-1-4, pp. 251-255

leads to better learning results. So for the most evaluation tools of the SOM like spanning trees and component cards a correct topology preserving is urgent.

We calculate the information dimension, a method of nonlinear dynamics, of the input space and the space represented by the weight vectors of the map in different learning states. Topology preservation is measured with the waber product.

2 Fractal dimension

Root of the matter is a method from the theory of dynamical systems, the calculation of the dimension of attractors, characterizing the geometrical structure of an attractor in phase space [5]. Because this number has often a noninteger value it is called fractal dimension. We use this method to determine the scaling behaviour of a given data set during decreasing partitioning of the phase space. Transmitting this technique of nonlinear dynamics to a real data set M with finite resolution is done by some simplifications [6].

In the embedding space of the data set M a set of volume elements $V_i(l)$ with the characteristic size l is chosen (e.g. l is the length of the edge of a hypercube), which should cover the whole set M. Now the definition of the fractal dimension is based on the following idea. Starting from a suitably defined property of M which depends on $V_i(l)$, we look at the scaling behaviour of this property in respect to the size of the volume elements. The resulting scaling law provides an abstract definition of a dimension. Different kinds of fractal dimensions are available.

Measuring the information dimension the number of points, determined by the vectors, $N_i(l)$ in each volume element $V_i(l)$ are counted. E.g. if the points of a given data set are arranged on a plain $N_i(l)$ is proportional to l^2, leading to an information dimension of 2. Bisection of the length of the edge of the embedding cubes leads to a reduction of factor 4 of points. The scaling behaviour of the information dimension d_i is given as follows. Following the definition of the information $I(l)$ by Shannon,

$$I(l) = -\sum_{i=1}^{A(l)} p_i(l) \log p_i(l) \tag{1}$$

where $A(l)$ is the number of volume elements necessary to cover the whole data set and $p_i(l)$ is the probability to find a point of the data in volume element i. This results in the following definition of the information dimension d_I:

$$d_I = \lim_{l \to 0} \frac{I(l)}{\log(1/l)} \tag{2}$$

3 Topology measurement using the waber product

The SOM's topology preservation is measured with the waber product. The main idea of this algorithm is to compare the neighbourhood relation between the neurons with respect to their position on the map on the one hand ($Q_2(j,l)$) and according to their stored weight vectors on the other ($Q_1(j,l)$). This leads to the following formula,

$$P(j,k) = (\prod_{l=1}^{k} Q_1(j,l)Q_2(j,l))^{\frac{1}{2k}} \tag{3}$$

where k means the k nearest neighbour, j the number of the actual neuron.

If there are no topological defects, $Q_1(j,l)$, $Q_2(j,l)$ as well as P(j,k) have the value 1. As a global criteria for characterizing the organization of the whole map we sum all P(j,k):

$$P_{waber} = \frac{1}{N(N-1))} \sum_{j=1}^{N} \sum_{k=1}^{N-1} \log(P(j,k)) \tag{4}$$

So a value of 0 for P_{waber} characterizes perfect topology preservation, a negative value folding into the input space. N is the number of all neurons.

4 Results

4.1 Exploring different data sets for their information dimension

We explore different data sets of real applications concerning their information dimension. The gas spectra and VLSI process data are easy to learn resulting in good structured spanning trees and component cards. In contrast to that we try many training cycles with a two dimensional map for the EEG data and the PCB layout pictures. Variing the different training parameters like number of processing units, number of training steps, heights and width of adaptation at the beginning, we were not able to get good learning results.

Data set	information dimension
Gas spectra	1.5
VLSI process	1.7
PCB layout pictures	5.4
EEG	5.7

Table 1: Different data sets and their information dimension

Considering table 1 the reason is a higher information dimension than 2. The lower dimensional map tries to approximate the higher dimensional input space by folding resulting in unstructered component cards.

ESANN'1994 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 20-21-22 April 1994, D-Facto public., ISBN 2-9600049-1-4, pp. 251-255

4.2 Different learning states, different map dimensions

Next we learn the EEG data set with SOMs having different dimensions and calculate the waberproduct and the information dimensions in different learning states. The results are shown in figures 1 and 2.

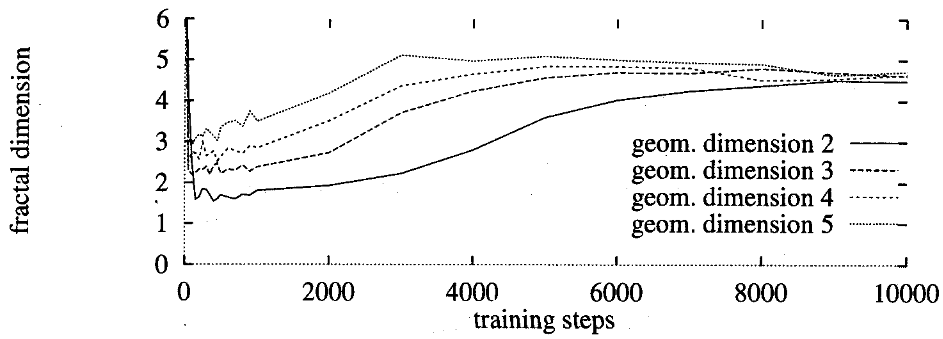

Figure 1: The informationdimension for maps in different dimension during learning

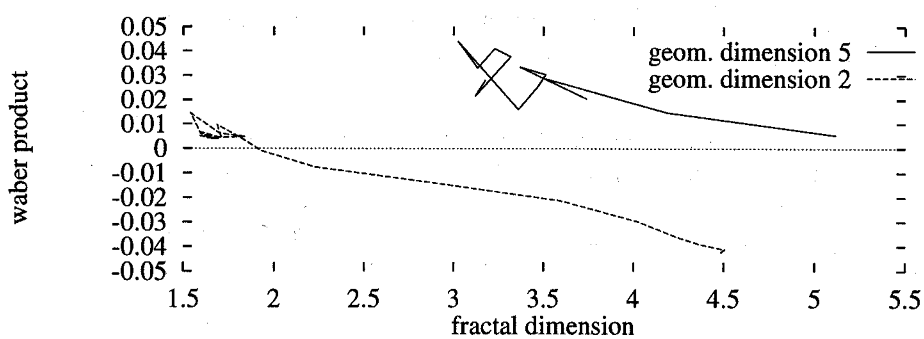

Figure 2: The waberproduct for maps in different dimension during learning

Considering figure 1 though the information dimension of the EEG data set is 5.7 the 5 dimensional map's weight vectors are only able to achieve an information dimension of 5.1 learning this data. Also a lower dimensional map is able to achieve an information dimension of about 4.75 but at the expense of the preservation of topology (figure ??). Considering the dynamics of learning, while the geometrical dimension of the map corresponds to the fractal dimension of the weight vectors the waber product is about 0. If the fractal dimension in-

ESANN'1994 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 20-21-22 April 1994, D-Facto public., ISBN 2-9600049-1-4, pp. 251-255

creases the waber product becomes negative, indicating, that the folding process begins. The behaviour of the curves at the beginning characterizes the ordering of the weight vectors from the random initialisation. Second a map with a lower geometrical dimension takes more time to achieve this higher information dimension. So it is necessary to consider both, the information dimension and the preservation of topology, to discuss the learning results.

5 Conclusion and further work

We look at different data sets and learning states of the SOM calculating their information dimension and determine their ability of topology preservation by using the waberproduct. Both is necessary to discuss a learning result. Lower dimensional maps approximate a higher dimensional input space at the expense of topology preservation.

In further work we will try to control the training phase. Choosing the corresponding map with a dimension equal to the information dimension of the data set we will stop the decreasing of width on the point when the map has achieved the necessary dimension and only decrease the heights of the adaption functions. Simulations have shown, that this strategy leads to the best learning results means the lowest rate of topological defects.

References

[1] T. Kohonen: Selforganization and associative memory. Springer Verlag Heidelberg New York Tokyo. 1984.

[2] H. Ritter, T. Martinetz, K. Schulten: Neuronale Netze. Addison Wesley, 1990.

[3] J. Göppert et.al.: Evaluation of spectra in chemistry and physics with kohonen's selforganizing feature map. In *NEURO NIMES*, 1992.

[4] K.M. Marks, K. Goser: Analysis of VLSI process data based on selforganizing feature maps. In *NEURO NIMES*, 1988

[5] J. Peinke et. al.: Encounter with chaos. Springer-Verlag. 1992.

[6] K. Pawelzik: Topologiererhaltung und Dimensionsbestimmung. In *PI IP*. 1992.