ESANN'1996 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-25-26 April 1996, D-Facto public., ISBN 2-9600049-6-5, pp. 79-84

A Self-Organizing Map for Analysis of High-Dimensional Feature Spaces with Clusters of Highly Differing Feature Density

Stefan Schünemann, Bernd Michaelis

Otto-von-Guericke University Magdeburg
Institute for Measurement and Electronics
P.O. Box 4120, D-39016 Magdeburg, Germany

Abstract. Self-Organizing Feature Maps (SOFM) facilitate a reduction and cluster analysis of high-dimensional feature spaces. One property of these artificial neural nets is a smoothing of the input vectors and thus a certain insensitivity to outliers and clusters of low feature density. While classifying clusters of highly different feature density this property is undesirable. This paper introduces an algorithm which makes a projection of low feature density clusters onto a SOFM possible, too. The resulting weight vectors represent reference vectors of all clusters.

1. Introduction

The theory of Self-Organizing Feature Maps (SOFM) facilitates a reduction and cluster analysis of high-dimensional feature spaces onto two-dimensional arrays of representative weight vectors. The smoothing of input vectors is a basic property of the projection of the feature space onto the SOFM. Therefore the feature map is relatively insensitive to outliers in the learning set.

During classification clusters of different and in some cases low feature density may be encountered. Due to the averaging property, these will not necessarily project onto separate clusters on the SOFM. The weight vectors of the representing neurons cannot be considered as reference vectors for clusters of low feature density. This can be attributed to the strongly dissimilar probabilities during random selection of a feature vector from the individual clusters of differing feature density [2]. In the case that the cluster of low feature density represent significant qualities of the distribution, the smoothing property is highly undesirable. Such feature distributions are well-known from classification tasks in science and medicine.

Classical methods for the detection of these clusters, such as outlier tests, are cumbersome and inefficient in a high-dimensional feature space. This raises the demand for the development of a learning algorithm, which facilitates a separate projection and whose weight vectors converge towards those clusters and therefore, represent reference vectors.

In the following a training algorithm will be introduced that is derived from the basic SOFM method and satifies the requirements mentioned above.

ESANN'1996 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-25-26 April 1996, D-Facto public., ISBN 2-9600049-6-5, pp. 79-84

2. Basic Method of the SOFM

The η-dimensional weight vectors **w**(x,y) of the neurons of the Self-Organizing Feature Map are randomly initialized in the beginning of the training phase. In each training step, a feature vector $\mathbf{v} \in \mathbf{R}^{\eta}$ is chosen from the learning set with a certain probability. Each neuron calculates its Euclidean distance to the presented vector:

$$d_v^w(x,y) = \sqrt{\sum_{k=1}^{\eta}\left[\mathbf{v}(k) - \mathbf{w}_i(x,y,k)\right]^2} \qquad (1)$$

The weight vectors of the winner neuron n_{xy} (the neuron featuring the minimal distance), and its coupled neighbors $n_{x'y'}$ are updated according to the following rule:

$$\mathbf{w}_{i+1}(x,y,k) = \mathbf{w}_i(x,y,k) + \alpha_i \, \varphi_i(x,y) \, [\mathbf{v}(k) - \mathbf{w}_i(x,y,k)] \quad , \ k = 1..\eta \qquad (2)$$

in the direction of the feature vector **v**.
Here, the learning rate α determines the magnitude of the update. It has to be defined as a with increasing training steps monotoneously decreasing function, which asymptotically approaches $\alpha \Rightarrow 0$. Its initial value has to be set at the initialization of the SOFM in the range $\alpha(i=1) \in [1,0)$.
For the following derivation the neighborhood distance D is defined as the Vector- or Manhattandistance between the neuron position x, y. Further the neighborhood radius r is assumed as monotoneously decreasing with increasing learning step. In SOFM with up to several hundred neurons the neighborhood function φ may be defined in a simple way. Each neuron with $r \leq D$ is assigned the value $\varphi=1$. The other neurons are not regarded as coupled with the winner neuron and thus, are assigned $\varphi=0$. In larger SOFM, a Gauss-shaped neighborhood function is commonly used [1].
During the first training epochs the initial order of the map will be established. With progressing training, the effect of the winner neuron on its neighbors decreases and thus, the adaptation in the direction of the feature vector is diminished. At this stage, the training process has reached the phase of asymptotic convergence. The learning rate α and the neighborhood function φ are of essential importance for the convergence of the algorithm, i.e. a suitable representation of the feature space by the weight vectors of the map [3].

3. Modifications of the Basic Method

From the literature, several modifications of the learning algorithm towards an optimization of the convergence are known (e.g. [3], [4]). In this paper, the goal of the modification of the algorithm consists in a change of the neighborhood function φ in such a way that it effects the convergence for fulfilling the task explained in the introduction: the update of the winner neurons in the direction of vectors of low feature density shall be strengthened. The modified neighborhood function φ will be

ESANN'1996 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-25-26 April 1996, D-Facto public., ISBN 2-9600049-6-5, pp. 79-84

noted as φ^*. An idea concerning the initialization of the SOFM will be used for the description of the partial space of the η-dimensional feature space which utilizes the principal components analysis ([5], [1]) and defines a hyper-ellipsoid γ with the eigenvalues λ_k of the feature distribution as its half-axes:

$$\gamma = \frac{v'(1)^2}{\lambda_1} + \frac{v'(2)^2}{\lambda_2} + \cdots + \frac{v'(k)^2}{\lambda_k} \quad , k = 1 \ldots \eta \tag{3}$$

where v' are the feature vectors transformed using the eigenvectors.
The main purpose of the modification of the neighborhood function in Eq.(4) is the definition of a function, which in contrast to the basic algorithm more strongly differentiates the degree of coupling of the neighboring neurons with the winner neuron. If the neuron is the winner neuron n_{xy} then simply results $\varphi^* = 1$. At $\gamma > 1$ (outside the hyper-ellipsoid Eq.(3)) φ^* is additionally affected by the term β (see Eq.(5)).

$$\varphi_i^*(x,y) = \begin{cases} 1 & , n_{xy} : d_v^w(x,y) = \min. \\ \dfrac{\beta}{2\, D_{x'y'}^{xy}}, & n_{x'y'} : r(i) \geq D_{x'y'}^{xy} \\ 0 & , \text{else} \end{cases} \tag{4}$$

Fig.1 schematizes the relations in the feature space. $d(n_{x'y'},M)$ and $d(v,M)$ describe the distance of the weight vectors or the feature vectors, respectively, from the hyper-ellipsoid introduced in Eq.(3). Fig. 2 exemplifies the function φ^* for the coupled neighborhood neurons for the vector distance. Modification occurs depending on the current feature vectors in the training epoch i.
φ^* conjointly with the learning rate α determines the magnitude of the update for the weight of the neurons, where the term β (see Eq. (4)):

$$\beta = \begin{cases} 1 & , \begin{array}{ll} \sqrt{2} \leq r(i) \leq r(i=1) & \\ [r(i)=z] < r(i) < \sqrt{2} & : \ d_M^{n_{xy}} \leq d_M^V \mid \gamma \leq 1 \\ r(i_{max}) \leq r(i) \leq [r(i)=z] & : \ \gamma \leq 1 \end{array} \\ \dfrac{1}{e^{10 \frac{d_M^{n_{x'y'}}}{d_{max}^V}}}, & \text{else} \end{cases} \tag{5}$$

depends on the three stages of the modified training algorithm:

$$\begin{array}{lll} \text{initialization:} & \sqrt{2} \leq r(i) \leq r(i=1) & \\ \text{orientation:} & [r(i)=z] < r(i) < \sqrt{2} & , \quad \sqrt{2} < z \leq r(i_{max}) \\ \text{convergence:} & r(i_{max}) \leq r(i) \leq [r(i)=z] & \end{array} \tag{6}$$

Threshold z sets the beginning of the convergence stage.

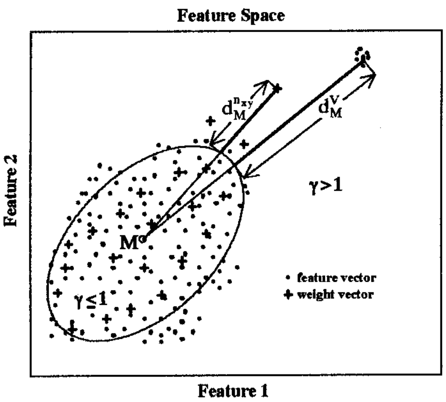

Fig. 1: Feature space

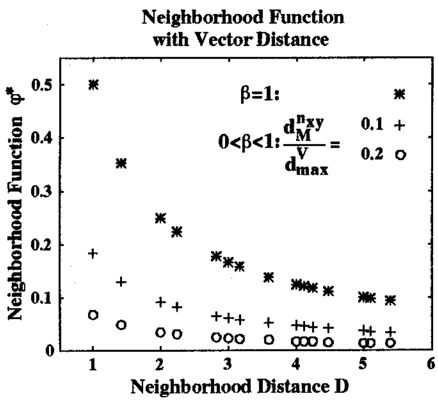

Fig. 2: Neighborhood function φ*

During the initialization of the training ($\beta = 1$) the inital order of the SOFM is established. The relation 1/(2*D) causes a non-linear variation of the neuron's weights in the direction of the feature vector. Using an adequate initial distribution to pre-orientate the neurons, the number of training epochs for this phase can be greatly reduced. There is no differentiation of the partial spaces $\gamma \leq 1$ and $\gamma > 1$ during this stage of training.

The orientation phase serves the further adaptation of the weights to the feature clusters within the hyper-ellipsoid $\gamma \leq 1$ and suppresses a contraction of the map into the ellipsoid. There is no additional limitation on learning in the partial space $\gamma \leq 1$. A second intention for this phase for $\gamma > 1$ is the development of neurons, which are supposed to converge in the direction of the clusters of low feature density. The neighborhood neurons with $\gamma > 1$ and whose winner neurons n_{xy} is pulled by the feature vector **v** towards the hyper-ellipsoid, experience an additional limitation of their adaptability through factor β (see Fig. 1). For a measure, the shortest Euclidean distance of the respective weight vector to the ellipsoid is used, i.e. the length of the normal to the tangential hyper plane of the ellipsoid to the end point of the weight vector. Because of the high computational load, this is simplified to the length of the straight $d(n_{x'y'},M)$ to the point of penetration in the direction of the center of gravity of the distribution. This length is normalized with the maximum distance d(v,max) of the feature vectors. This normalized distance q is the power in the empirically found function $1/e^{10q}$.

In the third phase of the training process the neuron weights converge for a cluster within $\gamma \leq 1$. Outside the hyper-ellipsoid individual neurons will be updated in the direction of clusters of low feature density, they also converge and thus represent reference vectors. This is achieved mainly by means of a limited adaptability of winner neurons with learning rate α in the direction of those clusters and the limitation of the term $\alpha\varphi$ under increasing influence of the factor β for those neurons, which become neighborhood neurons and are found outside and in the direction of the hyper-ellipsoid.

ESANN'1996 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-25-26 April 1996, D-Facto public., ISBN 2-9600049-6-5, pp. 79-84

4. Applications of the Modified Algorithm

The performance of the described neighborhood function φ^* for a cluster analysis and classification is to be tested with an investigation of the self-organization during the growth of crystals of semiconductor nanostructures [6] and the detection of highly complex combination patterns of cell surfaces in the immun system [7][8]. Both projects are characterized by clusters with a low feature density which yield important consequences in physics or medicine.

As the neighborhood radius

$$r(i) = \rho^{bi-b+1}\frac{\rho-1}{\rho}+1 \quad , \quad \rho = \frac{D_{max}}{a}\left(\frac{a}{D_{max}}\right)^{\frac{10}{n}} \tag{7}$$

is applied.

The normalizing parameter *a* and the parameter *b* make a manipulation of the initial value of the function r(i=1) and the steepness, respectivily, possible. At the start of the training the SOFM is initialized in the direction of the two greatest eigenvalues, i.e. variances, of the feature space [5]. So, the initial order of the map can be achieved with a significantly reduced number of training epochs.

Fig. 3 shows the feature (•) and weight vectors (+) of the SOFM in a reduced three-dimensional feature space for the classification of semiconductor nanostructures, side-by-side for both the Basic Method and the modified Algorithm. From the figure it is evident that using the Basic Method, clusters II and III have been suppressed.

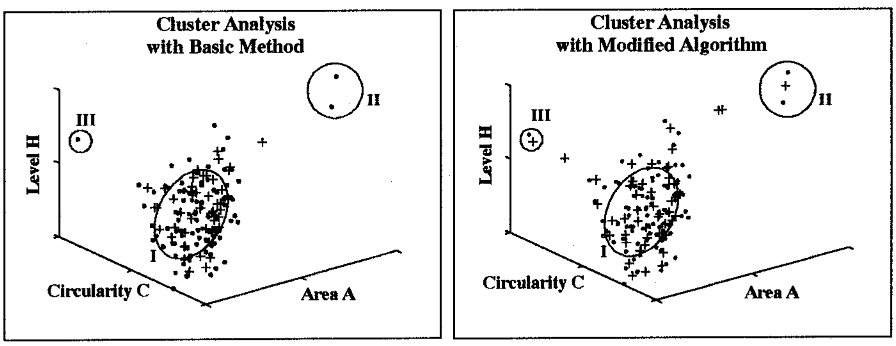

Fig. 3: Feature and weight vector for the classification of semiconductor nanostructures

CD8 28%		CD2/CD8 1%		
			CD2 8%	
CD8/CD7 1.5%		CD2/CD7 3%		CD2/CD4 2%
	CD2/CD7 /CD4 1%		CD4/CD7 1.5%	
	CD7 15%			CD4 50%

Fig. 4: SOFM for several combination patterns of human antigen

ESANN'1996 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-25-26 April 1996, D-Facto public., ISBN 2-9600049-6-5, pp. 79-84

In Fig. 4 the result of the cluster analysis using the Modified Algorithm of a similarly reduced four-dimensional feature space for several combination patterns of extractad human antigen in the blood along with their percentage are presented. Cluster with low percentages are suppressed by the Basic Method analog to Fig. 3.

5. Conclusion and Outlook

The applications of the introduced algorithm show that already with simple modifications of the neighborhood function, i.e. the definition of a hyper-ellipsoid around the center of gravity of the feature distribution and the consequently defined neighborhood function, the convergence of the neuronal weights in the direction of clusters of low feature density can be improved.
The expansion of the algorithm to cluster-dependent multi-layer structures in connection with the definition of several hyper-ellipsoids seems reasonable for a further improvement of the convergence of the SOFM in a high-dimensional feature space. The higher layers increase the resolution of the analyzed clusters.

References

[1] T. Kohonen: Self-Organizing Maps, Springer Series in Information Sciences, Springer, New York 1995

[2] K. Obermayr; G.G. Blasdel; K. Schulten: Statistical-Mechanical Analysis of Self-Organization and Pattern Formation During the Development of Visual Maps, Physical Review A, Vol. 45(10), pp. 7568-7589, 1992

[3] M. McInerney; A. Dhawan: Training the Self-Organizing Feature Map using Hybrids of Genetic and Kohonen methods, Proc. The 1994 IEEE Int. Conf on Neural Networks, pp. 641-644, Orlando 1994

[4] F. Mulier; V. Cherkassky: Learning Rate Schedules for Self-Organizing Maps, Proc. 12th Int. Conf. on Pattern Recognition, pp. 224-228, Jerusalem 1994

[5] H. Bertsch; J. Dengler; H.P. Meinzer: Das selbstlernende System der topologischen Merkmalskarte zur Klassifikation und Bildsegmentierung, Proc. 10. Symposium Deutsche Arbeitsgemeinschaft für Mustererkennung, pp. 298-303, Zürich 1988

[6] S. Dressel; et.al.: Structural Analysis in the Atomic Force Microscopy of Self-Organized Semiconductor Nanostructures, submitted to: Frühjahrstagung der Deutschen Physikalischen Gesellschaft, Regensburg March 1996

[7] W.Schubert: Antigenetic determinants of T lymphocyte $\alpha\beta$ receptor and other leukocyte surface proteins as differential markers of skeletal muscle regeneration: detection of spatially and timely restricted pattern by MAM microscopy, Eur. J. Cell Biol. 58, pp. 395-410, 1992

[8] Ch. Rethfeldt; et.al.: Multi-Dimensional Cluster Analysis of Higher-Level Differentiation at Cell-Surfaces, submitted to: Jahrestagung der Neurowissenschaftlichen Gesellschaft, Berlin February 1996