ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 13-18

A Hierarchical Self-Organizing Feature Map for Analysis of Not Well Separable Clusters of Different Feature Density

Stefan Schünemann , Bernd Michaelis

Otto von Guericke University of Magdeburg
Institute for Measurement Technology and Electronics
P.O. Box 4120, D - 39016 Magdeburg, Germany
stefan.schuenemann@zenit.uni-magdeburg.de

Abstract. This paper introduces a hierarchical Self-Organizing Feature Map (SOFM). The partial maps consist of individual numbers of neurons, which makes a cluster analysis with different degrees of resolution possible. A definition of a special Mahalanobis space of the data set during the learning improves the properties concerning clusters with low density.

1. Introduction

The Self-Organizing Feature Map (SOFM) [2] is an efficient method for cluster analysis of a high-dimensional feature space onto 2D arrays of reference vectors.

Frequently, there exists no a-priori knowledge about the distributions of the features. In this case it is difficult to choose a useful number of neurons for the SOFM. Many authors successfully deal with this problem, e.g. [1]. But, a correct interpretation of the cluster analysis often makes different SOFMs with comparable orientation and distinct degrees of resolution necessary. Such demands are well known especially from analysis tasks in biomedical research.

Another important problem arises from the smoothing property of the SOFM which leads to a relative insensitivity to clusters of low feature density. As the approximation errors of the representing neurons increase, the weight vectors cannot be considered typical reference vectors any longer.

The main idea of a modification described in previous papers ([3], [4]) was to improve the learning ability of individual neurons on the SOFM in direction of regions with low feature density. This could be achieved by a special definition of partial spaces with a hyper-ellipsoid. A disadvantageous restriction consisted in the definition of only *one* ellipsoid and its *fixed* border.

In the following a learning algorithm of a hierarchical SOFM will be introduced that surmounts the above-mentioned problems by means of a Mahalanobis space which contains a variable number of centers of high feature density.

This research is supported by a grant from the Deutsche Forschungsgemeinschaft (MI377/5-2) and a grant from the Kultusministerium Sachsen-Anhalt (1897A/0025).

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 13-18

2. General structure of the hierarchical SOFM

Figure 1 shows an overview of the structure for the hierarchical SOFM, which contains a number of partial maps $k \in [1, l]$. The different maps are generated successively in a regime depending on pre-defined learning steps t_k. For an improvement of the resolution the number of neurons in a new map is increased.

The introduced learning algorithm is based on the original SOFM algorithm presented by Kohonen in [2]. The weight vector $\boldsymbol{m}$ of the winner neuron c, the neuron featuring the minimal Euclidean distance to the presented feature vector $\boldsymbol{x}$, and its neighbors i in the learning step t are updated according to:

$$\boldsymbol{m}_{ki}(t_k + 1) = \boldsymbol{m}_{ki}(t_k) + \alpha_k(t_k)\ \beta_{ki}(t_k, D_c)\ h_{kci}(t_k)\ [\boldsymbol{x}(t_k) - \boldsymbol{m}_{ki}(t_k)] \qquad (1)$$

where k describes the respective partial map. The learning rate α_k determines the magnitude of the update and the neighborhood function h_{kci} describes the effective training region of the SOFM. In addition, a term β_{ki} controls the update depending on the learning step and the Mahalanobis distance D, which describes the feature density within the N-dimensional feature space $\boldsymbol{x} \in \mathbf{R}^N$.

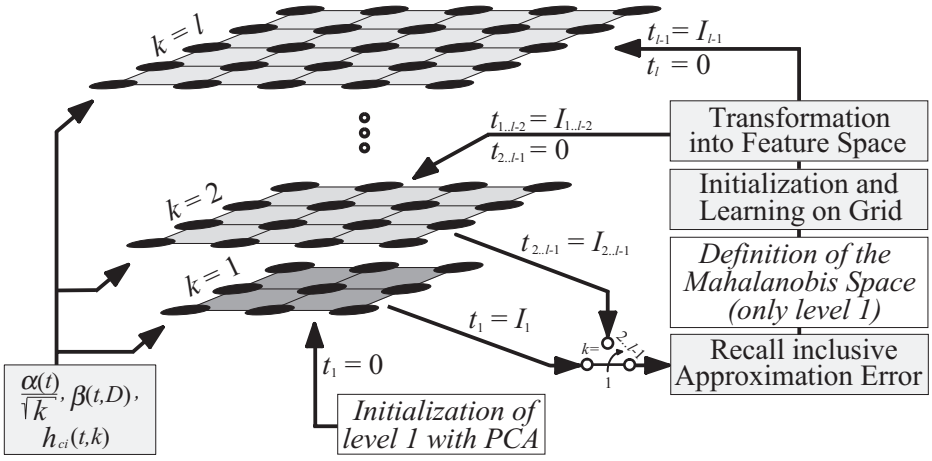


Figure 1: Overview of the hierarchical SOFM

3. Learning of the smallest partial map

In the beginning of the learning process $t_1 < I_1$ the smallest map of the hierarchical SOFM $k = 1$ is pre-orientated in direction of the two greatest eigenvalues of the whole feature distribution, calculated by the principal components analysis (PCA). During the learning steps $t_1 < I_1$ the term β_{1i} is assigned the value $\beta = 1$ (see Eq. 6).

During the first steps the initial order of the SOFM is established. At $t_1 = I_1$ a recall of the map is realized. In the investigated artificial example distributions (e.g. Fig. 2) the step $t_1 = I_1$ was higher than ten percent of the maximal number of learning steps. The feature vectors $\boldsymbol{x}$ projected onto each neuron i

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 13-18

serve for a calculation of regions with high feature density in the feature space. For the definition of a density space the Mahalanobis distance D_i is used:

$$D_i(\boldsymbol{x}) = [\boldsymbol{x}(t_k) - \overline{\boldsymbol{x}}_i]^T \;\; \boldsymbol{C}_i^{-1} \;\; [\boldsymbol{x}(t_k) - \overline{\boldsymbol{x}}_i] \tag{2}$$

where $\overline{\boldsymbol{x}}_i$ describes the mean vector of a partial distribution, that means the center i of high local feature density. The term $\boldsymbol{C}_i$ represents their covariance matrix. In contrast to the previous algorithm in [3], [4] with one fixed hyper-ellipsoid the Eq. 2 leads to equipotential lines with several centers of high density in the N-dimensional Mahalanobis space which describes the estimated density in continuous space. The accuracy of the estimation depends on the number of neurons in the smallest partial map and the quality of learning at $t_1 = I_1$. Notice that an expansion of the smallest map results in a remarkable increase in the computation time for the following learning phase.

After the described process a formation of a new partial map is possible. The continuation of the learning phase for the map k=1 including adaptation towards not well separable clusters with low feature density is described in section 5.

4. Initialization of the higher partial maps

At learning step $t_k = I_k$ a generation of another partial map k+1 with a greater number of neurons than the map k is possible. For the algorithm it is not necessary that the neuron numbers of the new map increase in each dimension by a step one. Especially, it can be useful to define the size of the map $k = 2$ significantly greater than the map $k = 1$.

The averaged and maximum approximation errors:

$$\overline{E}_{ki} = \frac{1}{q\ \sqrt{N}} \sum_{j=1}^{q} \parallel \boldsymbol{x}_j - \boldsymbol{m}_{ki} \parallel \qquad , \qquad E_{ki}^{(m)} = \frac{\max_{ki} \parallel \boldsymbol{x}_{1\ldots q} - \boldsymbol{m}_{ki} \parallel}{\sqrt{N}} \tag{3}$$

of each neuron of the recall at $t_k = I_k$ are used for the initialization of the map k+1. Using the standardization term $\sqrt{N}$ a comparison independent of the dimension of the feature space N is possible yielding $E \in [0, 1]$. The term q indicates the number of states of excitations for each neuron and $\boldsymbol{x}_j$ describes the feature vectors projected onto it. In consideration of an adaptation in the direction of clusters with low feature density the compound error function $E_{ki}^{(I)}$:

$$E_{ki}^{(I)} = \overline{E}_{ki}\ (E_{ki}^{(m)} - \overline{E}_{ki}) \tag{4}$$

is used for the definition of the data set for the initialization learning of map k+1. As its data the learning set contains two-dimensional, equidistant, interpolated coordinates of the topological neighborhood function of map k. The interpolated error function $E_{ki}^{(I)}$ of Eq. 4 determines the frequency of occurrence of each coordinate in the data set. The initialization learning is realized by means of small learning rates and a small number of steps. The neuron weights of the map $k+1$ learn coordinates of the topological neighborhood function which minimized the calculated approximation errors. After that the trained neuron positions on the neighborhood grid are transformed into N-dimensional feature space (see Fig. 2). For this task a bi-linear interpolation is used.

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 13-18

5. Learning with an additional limitation

After that the training of all partial maps $t_k < t_{max}$ of the hierarchical SOFM can be continued with $t_{1\ldots k-1}+1$, $t_k = I_k + 1$ and $t_{k+1} = 1$. During the steps $t_{k+1} < I_{k+1}$ the term $\beta_{(k+1)i}$ of the map k+1 is assigned the value 1 (see Eq. 6).

The learning process with $t_{1\ldots k} > I_{1\ldots k}$ realizes a further adaptation of the neuron weights to the feature distribution.The algorithm also supports an adaptation of several neurons in the direction of clusters with low feature density. The learning rate is limited to $\alpha_k/\sqrt{k}$. A similar restriction on α is used for the standard deviation of h_{kci}. In each learning step of this stage the minimal Mahalanobis distance D_c of the presented feature vector $\boldsymbol{x}$ is calculated according to:

$$D_c(\boldsymbol{x}) = \min_i [D_i(\boldsymbol{x})] \tag{5}$$

The center of high density which is characterized by the minimal distance through Eq. 5 describes the density relationships in the feature space by itself. That means, no superposition between several local centers exists. A monotonously decreasing function D_B controls the effective region of the selected high density center in the term β_{ki}:

$$\beta_{ki}(t_k, D_c) = \begin{cases} \dfrac{1}{[D_c(\boldsymbol{m}_{ki}) - D_B(t_k)]^3} & \begin{array}{l} t_k > I_k \text{ and } i \neq c \text{ and} \\ D_c(\boldsymbol{m}_{kc}) > D_B(t_k) \text{ and} \\ D_c(\boldsymbol{m}_{ki}) > D_c(\boldsymbol{x}) > D_B(t_k); \end{array} \\ 1 & \text{else} \end{cases} \tag{6}$$

of the modified learning rule in Eq. 1. The term β_{ki} is an empirically found function. In the region of lower feature density it is a decreasing function with increasing distance of the neighborhood neurons to the selected center of density.

If the feature vector $\boldsymbol{x}$ and the winner neuron c are located outside the region with high feature density, i.e. $D_c(\boldsymbol{m}_{kc}) > D_B(t_k)$ and $D_c(\boldsymbol{x}) > D_B(t_k)$, then the positions in feature space of all neighborhood neurons are tested: Those neighborhood neurons which are also situated in the region with lower feature density $D_c(\boldsymbol{m}_{ki}) > D_B(t_k)$ and meet the requirement $D_c(\boldsymbol{m}_{ki}) > D_c(\boldsymbol{x})$, respectively, experience an additional limitation on the learning ability through the term β_{ki}. This way, a contraction of the map can be suppressed. Therefore, it is possible to adapt some neurons in the direction of not well separable clusters with low feature density.

The performance of β in the previous algorithm are demonstrated in [5], [7] on artificial and real data sets. Correspondingly, the Figures 2 and 3 show a comparison between a hierarchical SOFM trained with a basic method, that means without limitation through β, and the modified algorithm. The used artificial example distribution contains clusters with different feature density. The clusters I and II are not completely separable. In contrast to the basic method the 4*4-SOFM of the modified algorithm projects the cluster VI onto a separate neuron and the weight vector represents a reference vector of this cluster. Moreover, a decrease of the approximation errors is detectable. Figure 3 also demonstrates the different resolution of the distinct partial maps of the hierarchical SOFM. The increase of the resolution facilitates a detection of smaller subclusters.

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 13-18

6. Conclusion

The used artificial example distribution shows that the described expansion of the learning algorithm to hierarchical structures of SOFM in connection with the definition of a Mahalanobis space which defines several regions of high feature density can improve the convergence of neuron weights in direction of clusters with low feature density. The increasing number of neurons in the higher partial SOFMs raises the resolution of the analyzed feature clusters. Besides, in contrast to SOFMs without pre-defined structures the introduced hierarchical SOFM makes an interpretation of the analysis without in-depth expert knowledge about artificial neural networks possible.

The training of the hierarchical SOFM, in particular with a high number of neurons and levels respectively requires a lot of computation time. However, the algorithm is well suited for a realization on parallel processing hardware.

The introduced algorithm is applied to first investigations of new and complex high-dimensional feature distributions without a-priori knowledge about the number, position and form of the sought after clusters (see also [6]).

In the foreseeable future the results of a cluster analysis with the described algorithm in a special field of biomedical research will be published.

References

[1] B. Fritzke: A Growing Self-Organizing Networks – Why? *Proc. 4^{th} Europ. Symposium on Artificial Neural Networks*, D facto, Bruges 1996, pp. 61-72.

[2] T. Kohonen: *Self-Organizing Maps*. Springer Series in Information Sciences, Springer, Berlin 1995.

[3] St. Schünemann, B. Michaelis: A Self-Organizing Map for Analysis of High-Dimensional Feature Spaces with Clusters of Highly Differing Feature Density. *Proc. 4^{th} Europ. Symposium on Artificial Neural Networks*, D facto, Bruges 1996, pp. 79-84.

[4] St. Schünemann, B. Michaelis, W. Schubert: Analysis of Multi-Fluorescence Signals Using a Modified Self-Organizing Feature Map. *Proc. 6^{th} Int. Conf. on Artificial Neural Networks*, Lecture Notes in Computer Science, Vol. 1112, Springer, Berlin 1996, pp. 575-580.

[5] St. Schünemann, U. Seiffert, B. Michaelis: Two More Modifications of SOMs to Handle Signals with Special Properties. *Proc. 1^{st} Workshop on Self-Organizing Maps*, Espoo 1997, pp. 292-297.

[6] St. Schünemann, B. Michaelis: Data Analysis of not well Separable Clusters of Different Feature Density with a Two-Layer Classification System Comprised of a SOM and an ART2-A Network. *Proc. IEEE Int. Joint Conf. on Neural Networks*, Anchorage 1998, pp. 707-712.

[7] St. Schünemann, B. Michaelis, W. Schubert: Cluster Analysis with SOFM for the Detection of Different Diseases in the Human Immune System. *Proc. 6^{th} Europ. Congress on Intelligent Techniques and Soft Computing*, Aachen 1998, Vol. 2, pp. 1320-1324.

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 13-18

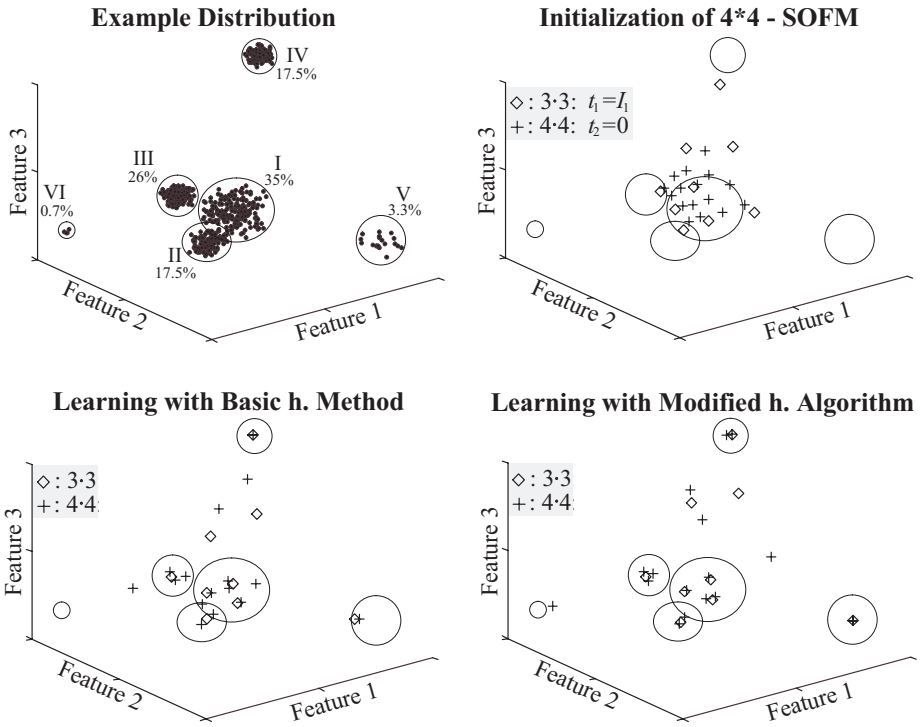
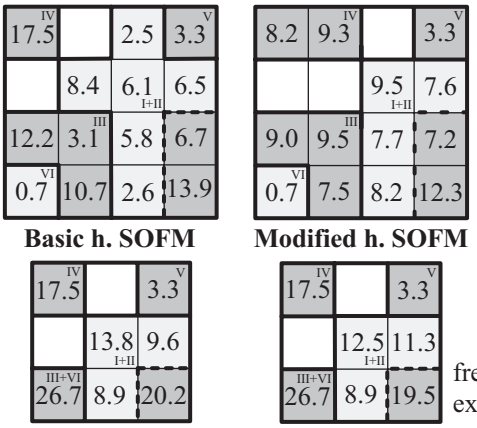


Figure 2: Learning results of the hierarchical SOFM with 3*3 and 4*4 neurons in comparison with the basic method (without limitation through the term β).



Approximation Error

		$\bar{E}_{ki}$	$E_{ki}^{(m)}$
Basic hierar. SOFM	3*3	0.0328	0.2791
	4*4	0.0273	0.1828
Modified hierar. SOFM	3*3	0.0316	0.2711
	4*4	0.0252	0.0650

$\bar{E}_{ki}, E_{ki}^{(m)} \in [\,0\,,\,1\,]$

Figure 3: Recalls of the trained hierarchical SOFMs (frequencies in [%]) and their approximation errors, side-by-side for both the basic method (without limitation through the additional term β) and the modified learning algorithm.