

Associative learning in SOMs for Fuzzy-Classification

T. Villmann & F.-M. Schleif
University Leipzig - Medical Dep.
Computational Intelligence Group
K.-Tauchnitz-Str. 25, 04107 Leipzig, Germany
thomas.villmann@medizin.uni-leipzig.de

M. van der Werff & A. Deelder
& R. Tollenaar
Leiden University Medical Center
Department of Surgery
Biomolecular Mass Spectrometry Unit

Abstract

We present a general framework for association learning in self-organizing maps (SOMs), which can be specified for usage for supervised fuzzy classification new approach for fuzzy classification of mass spectrometric data. In this way, we obtain a prototype based fuzzy classification model (FLSOM), which can be easily interpreted and visualized due to the fundamental properties of SOMs. Moreover, the provided extension gives the ability to detect class similarities. We apply this approach to classification and class similarity detection for mass spectrometric data in case of cancer disease and obtain comparable results. We demonstrate that class similarity detection leads to clinically expected class similarities.

1 Introduction

The self-organizing map (SOM) constitutes one of the most popular unsupervised approaches for clustering, visualization and data mining of high-dimensional data [16]. SOMs belong to the prototype based methods of data representation. Due to its inherent regularization abilities SOMs are also applicable in case of sparse data sets. Basically, SOMs map the data nonlinearly onto a low-dimensional regular lattice of neurons. This mapping is under certain conditions topology preserving, i.e. similar data points are mapped onto nearby or identical neurons of the neuron grid [21]. Thereby, adaptation takes place as an unsupervised prototype learning. Several approaches are proposed to extend the unsupervised learning scheme for supervised or semi-supervised adaptation. The approaches ranges from simple post-labeling [16], the utilization of auxiliary data information [19], to combination of SOMs with other supervised neural network models like Learning Vector Quantization (LVQ) [25], the multi-layer perceptron (MLP) [3] (for example the counterpropagation model [14]).

Recently, a supervised counterpart of SOM is developed [23]. It allows the determination of a prototype based fuzzy

classification model (FLSOM). In contrast to the widely applied multilayer perceptron [3], prototype based classification allows an easy interpretation, which is of particular interest for many applications. FLSOM leads to a robust fuzzy classifier where efficient learning of fuzzy labeled or partially contradictory data is possible. Additionally, FLSOM gives the possibility to assess and to visualize *class similarities* by inspection of the generated class map using the visualization abilities of SOM [20]. However, FLSOM differs from other extensions of SOM for classification like counterpropagation [14] or Fuzzy SOM [25] fundamentally: In contradiction to these models, in FLSOM the prototype adaptation is also influenced by the class information of the data and, hence, optimized accordingly.

In this contribution, first we develop a general scheme of *association learning in SOMs*, which leads to a general framework of prototype-based learning balancing unsupervised aspects and additional information. Generally, prototype based methods have the advantage of easy interpretation whereas other models like MLP, for example, work as a black box. Thereby, we restrict the description of the framework to the SOM model because of its excellent visualization abilities. However, a transfer to other prototype based models is straight forward. We show that FLSOM can be seen as a special case of this more general approach. After this theoretic part we apply the FLSOM to the classification cancer diseases based on the analysis of mass spectrometric data obtained from MALDI-TOF. Thereby, relevance learning for classification dependent metric adaptation can be used for the detection of potential biomarkers based on the obtained feature ranking. We demonstrate the abilities of the approach, which are comparable or superior to standard tools for spectra analysis provided by commercial tools as well as previous investigations [5],[6].

2 Association learning in SOM's

Before introduction of the association learning framework for SOMs we briefly introduce the basic concepts and principles of SOMs.

2.1 The Self-Organizing Map

As mentioned above, SOMs can be taken as unsupervised learning of topographic vector quantization with a topological structure (grid) within the set of prototypes (codebook vectors). Thereby, roughly speaking, topology preservation means that similar data points $\mathbf{v} \in V$ are mapped onto identical or neighbored grid locations which have pointers into the data space (weight vectors). An exact mathematical definition is given in [21]. The weight vectors also are called prototypes, because they represent parts of the data space.

There exists a wide range of applications in pattern recognition ranging from spectral image processing to bioinformatics. The mathematics behind the original SOM model as proposed by KOHONEN is rather complicated. In particular, the training process does not follow a gradient descent on any cost function for continuous data distributions [9]. However, HESKES proposed a variant of the original algorithm which, in practice, leads to at least very similar or identical results as the original SOM but for which a cost function can be established [15]. We will base our model on this formulation:

Assume that data $\mathbf{v} \in V \subseteq \mathbb{R}^d$ are given distributed according to an underlying distribution $P(V)$. A SOM is determined by a set A of neurons \mathbf{r} equipped with weight vectors (prototypes) $\mathbf{w}_{\mathbf{r}} \in \mathbb{R}^d$. The neurons are arranged on a lattice structure, which determines the neighborhood relation $N(\mathbf{r}, \mathbf{r}')$ between the neurons \mathbf{r} and \mathbf{r}' . Denote the set of prototypes by $\mathbf{W} = \{\mathbf{w}_{\mathbf{r}}\}_{\mathbf{r} \in A}$. The mapping description of a trained SOM defines a function

$$\Psi_{V \rightarrow A} : \mathbf{v} \mapsto s(\mathbf{v}) = \underset{\mathbf{r} \in A}{\operatorname{argmin}} \sum_{\mathbf{r}' \in A} h_{\sigma}(\mathbf{r}, \mathbf{r}') \xi(\mathbf{v}, \mathbf{w}_{\mathbf{r}'}) \quad (1)$$

where $le(\mathbf{r}) = \sum_{\mathbf{r}' \in A} h_{\sigma}(\mathbf{r}, \mathbf{r}') \xi(\mathbf{v}, \mathbf{w}_{\mathbf{r}'})$ is the local neighborhood weighted error of the distances $\xi(\mathbf{v}, \mathbf{w}_{\mathbf{r}'})$. ξ is an appropriate chosen distance measure, usually the quadratic Euclidean norm $\xi(\mathbf{v}, \mathbf{w}_{\mathbf{r}}) = (\mathbf{v} - \mathbf{w}_{\mathbf{r}})^2$. However, here we only suppose $\xi(\mathbf{v}, \mathbf{w})$ to be arbitrary assuming differentiability and symmetry and assessing some dissimilarity. The function

$$h_{\sigma}(\mathbf{r}, \mathbf{r}') = \exp\left(\frac{N(\mathbf{r}, \mathbf{r}')}{2\sigma^2}\right) \quad (2)$$

determines the neighborhood cooperation with range $\sigma > 0$. In this formulation, an input stimulus \mathbf{v} is mapped onto that position $\mathbf{r} \in A$ of the SOM, the local error $le(\mathbf{r})$ of which is minimum, whereby the average over all neurons according to the neighborhood is taken. We refer to this neuron $s(\mathbf{v})$ as the winner.

During the adaptation process a sequence of data points $\mathbf{v} \in V$ is presented to the map representative for the data distribution $P(V)$. Each time the currently most proximate neuron $s(\mathbf{v})$ according to (1) is determined. All prototypes

are gradually adapted according to the neighborhood degree of the respective neuron to the winning one by

$$\Delta \mathbf{w}_{\mathbf{r}} = -\epsilon h_{\sigma}(\mathbf{r}, s(\mathbf{v})) \frac{\partial \xi(\mathbf{v}, \mathbf{w}_{\mathbf{r}})}{\partial \mathbf{w}_{\mathbf{r}}} \quad (3)$$

with a small learning rate $\epsilon > 0$. This adaptation follows a stochastic gradient descent of the cost function introduced by HESKES [15]:

$$E_{\text{SOM}} = \int \frac{P(\mathbf{v})}{2C(\sigma)} \sum_{\mathbf{r}} \delta_{\mathbf{r}}^{s(\mathbf{v})} \sum_{\mathbf{r}'} h_{\sigma}(\mathbf{r}, \mathbf{r}') \xi(\mathbf{v}, \mathbf{w}_{\mathbf{r}'}) d\mathbf{v} \quad (4)$$

where $C(\sigma)$ is a constant which we will drop in the following, and $\delta_{\mathbf{r}}^{\mathbf{r}'}$ is the usual Kronecker symbol checking the identity of \mathbf{r} and \mathbf{r}' .

One main aspect of SOMs is the visualization ability of the resulting map due to its topological structure. Under certain conditions the resulting non-linear projection $\Psi_{V \rightarrow A}$ generates a continuous mapping from the data space V onto the grid structure on A . This mapping can mathematically be interpreted as an approximation of the principal curve or its higher-dimensional equivalents [12]. Thus, as pointed out above, similar data points are projected on prototypes which are neighbored in the grid space A . Further, prototypes neighbored in the lattice space should code similar data properties, i.e. their weight vectors should be close together in the data space according to the dissimilarity measure ξ . This property of SOMs is called topology preserving (or topographic) mapping realizing the mathematical concept of continuity. For a detailed consideration of this topic we refer to [21]. Successful tools for assessing this map property are the topographic function and the topographic product [21],[1].

2.2 Association learning in SOMs

Now we are ready to establish the general framework of *association learning* in SOMs (AL-SOM). The essential presupposition is that the unsupervised SOM model has a cost function as introduced above. Further we assume that for each data point $\mathbf{v} \in V$ an association object \mathbf{x} is available such that AL-SOM model should learn the association between \mathbf{v} and \mathbf{x} in a prototype based manner. Examples of such association models could be simple single values/vectors or complicate structures like linguistic objects (words, sentences ...) or graphs.

Let $\mathbf{y}_{\mathbf{r}}$ be the association model of the prototype $\mathbf{w}_{\mathbf{r}}$ and $\mathbf{Y} = \{\mathbf{y}_{\mathbf{r}}\}$. For the formal framework, we now introduce a general cost function for AL-SOM as

$$E_{\text{AL-SOM}}(\mathbf{W}, \mathbf{Y}) = (1 - \beta) E_{\text{SOM}}(\mathbf{W}) + \beta E_{\text{AL}}(\mathbf{W}, \mathbf{Y}) \quad (5)$$

Thereby the second term E_{AL} reflects the accuracy of the model with respect to association learning task. It has to be

prototype dependent to connect association learning with the prototype adaptation, i.e. $E_{AL} = E_{AL}(\mathbf{W}, \mathbf{Y})$. The parameter $\beta \in [0, 1]$ balances between both criteria. For $\beta = 0$, ordinary SOM learning is achieved.

To obtain a gradient based adaptation scheme we further assume that $E_{AL}(\mathbf{W}, \mathbf{Y})$ is differentiable in \mathbf{w}_r . Then, formal derivation yields

$$\frac{\partial E_{AL-SOM}}{\partial \mathbf{w}_r} = (1 - \beta) \frac{\partial E_{SOM}(\mathbf{W})}{\partial \mathbf{w}_r} + \beta \frac{\partial E_{AL}(\mathbf{W}, \mathbf{Y})}{\partial \mathbf{w}_r} \quad (6)$$

with non-vanishing term $\frac{\partial E_{AL}(\mathbf{W})}{\partial \mathbf{w}_r}$. The general approach allows a broad variety of task specific choices for $E_{AL}(\mathbf{W}, \mathbf{Y})$.

A strategy, conformal with prototype based models, is to make association learning locally: The prototypes specifically learn their associations in dependence from their similarity to the given data. In agreement with the neighborhood concept in SOMs defined by Gaussians a consistent possible choice of local association learning is

$$E_{AL} = \frac{1}{2} \sum_r \int P(\mathbf{v}) \cdot g_\gamma(\mathbf{v}, \mathbf{w}_r) \cdot \tilde{E}_{AL}(\mathbf{Y}) d\mathbf{v} \quad (7)$$

with the *locality kernel*

$$g_\gamma(\mathbf{v}, \mathbf{w}_r) = \exp\left(-\frac{\xi(\mathbf{v}, \mathbf{w}_r)}{2\gamma^2}\right) \quad (8)$$

and γ determining the locality range in the data space. In this way locality is separated from evaluation of the association accuracy determined by $\tilde{E}_{AL}(\mathbf{Y})$. The choice of the classification accuracy term E_{AL} as a sum of weighted data space distances is based on the continuity assumption, which is that data points, close to a prototype \mathbf{w}_r , have correspondingly similar association objects, i.e. a smooth distribution is assumed.

General approaches for the accuracy term \tilde{E}_{AL} may be based on task specific similarity measures (edit distance for symbolic data, graph distances, generalized Minkowski metric for functional data, etc.). For convenience, differentiability with respect to the association objects is recommended. In this case, the Robbins-Monroe-formalism [13] immediately gives the adaptation rule as stochastic gradient descent with

$$\frac{\partial E_{AL-SOM}(\mathbf{W}, \mathbf{Y})}{\partial \mathbf{y}_r} = g_\gamma(\mathbf{v}, \mathbf{w}_r) \cdot \frac{\partial \tilde{E}_{AL}(\mathbf{Y})}{\partial \mathbf{y}_r} \quad (9)$$

As mentioned above, unsupervised SOMs generate a topographic mapping from the data space onto the prototype grid under specific conditions. If the association objects are consistently distributed in the object space with respect to the varying data, one can expect for topographic AL-SOMs that the association models \mathbf{y}_r of the prototypes become ordered within the grid structure of the prototype

lattice. In this case the topological order of the prototypes should be transferred to the topological order of association models such that we have a smooth change of the models between neighbored grid positions \mathbf{r} in the index lattice A . This is the consequence of the following fact: the neighborhood function $h_\sigma(\mathbf{r}, \mathbf{s})$ of the usual SOM learning (3) forces the topological ordering of the prototypes. In AL-SOM, this ordering is further influenced by the locality kernel $g_\gamma(\mathbf{v}, \mathbf{w}_r)$ (8). Hence, the prototype learning and ordering (15) receives information of both data and association objects distribution. Thereby, for high value of the balancing parameter β the latter term becomes dominant. Otherwise, the kernel $g_\gamma(\mathbf{v}, \mathbf{w}_r)$ also triggers the association learning (16), which, of course, also depends on the underlying learned prototype distribution and ordering. Thus, a consistent ordering of the labels is obtained in AL-SOM.

2.3 Metric adaptation - relevance learning

As mentioned above, the general dissimilarity measure $\xi(\mathbf{v}, \mathbf{w}_r)$ for the data space V is often chosen as squared Euclidean metric such that the derivative $\frac{\partial \xi(\mathbf{v}, \mathbf{w}_r)}{\partial \mathbf{w}_r}$ simply becomes $-2(\mathbf{v} - \mathbf{w}_r)$. Yet, other measures also can be applied, for example correlation measures [24]. However, more flexibility is obtained if $\xi(\mathbf{v}, \mathbf{w}_r)$ is parametrized and the parameters are also subject of optimization according to the given classification task [11],[10].

Generally, we consider a parametrized distance measure $\xi^\lambda(\mathbf{v}, \mathbf{w})$ with a parameter vector $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_M)$ with $\lambda_i \geq 0$ and normalization $\sum_i \lambda_i = 1$. Then classification task depending parameter optimization is achieved by gradient descent, i.e. by consideration of $\frac{\partial E_{AL-SOM}}{\partial \lambda_i}$. Formal derivation yields

$$\frac{\partial E_{AL-SOM}}{\partial \lambda_i} = (1 - \beta) \frac{\partial E_{SOM}}{\partial \lambda_i} + \beta \frac{\partial E_{AL}}{\partial \lambda_i} \quad (10)$$

with

$$\frac{\partial E_{SOM}}{\partial \lambda_i} = \frac{1}{2} \sum_r \int \delta_r^{s(\mathbf{v})} P(\mathbf{v}) \sum_{r'} h_\sigma(\mathbf{r}, \mathbf{r}') \frac{\partial \xi^\lambda(\mathbf{v}, \mathbf{w}_r)}{\partial \lambda_i} d\mathbf{v} \quad (11)$$

and, supposing \tilde{E}_{AL} to be independent from ξ^λ ,

$$\frac{\partial E_{AL}}{\partial \lambda_i} = - \sum_r \int \frac{g_\gamma(\mathbf{v}, \mathbf{w}_r)}{4\gamma^2} \frac{\partial \xi^\lambda(\mathbf{v}, \mathbf{w}_r)}{\partial \lambda_i} \tilde{E}_{AL}(\mathbf{Y}) P(\mathbf{v}) d\mathbf{v} \quad (12)$$

for the respective parameter adaptation.

In case of $\xi^\lambda(\mathbf{v}, \mathbf{w})$ being the *scaled* squared Euclidean metric

$$\xi^\lambda(\mathbf{v}, \mathbf{w}) = \sum_i \lambda_i (v_i - w_i)^2 \quad (13)$$

(with $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$) the derivative becomes $\frac{\partial \xi(\mathbf{v}, \mathbf{w}_i)}{\partial \mathbf{w}_i} = -2 \cdot \boldsymbol{\Lambda} \cdot (\mathbf{v} - \mathbf{w}_i)$ with $\boldsymbol{\Lambda}$ is a diagonal matrix and its i -th diagonal entry is λ_i .

The parameter optimization of the *scaled* squared Euclidean metric allows a useful interpretation. The parameters λ_i weight the dimensions of the data space. Hence, optimization of these parameters in dependence on the classification problem leads to a ranking of the input dimensions according to their classification decision relevance. Therefore, metric parameter adaptation of the scaled Euclidean metric is called *relevance learning*. In case of zero-valued λ_i this can also be seen as feature selection.

2.4 Example - Fuzzy-Labeled SOM (FLSOM)

Here, we give an explicit example for the above introduced general framework of association learning: The recently published Fuzzy-Labeled SOM (FLSOM, [23]) for fuzzy classification can be described in this scenario.

Let $N(c)$ be the number of possible data classes. Now, the association objects are class label vectors $\mathbf{x}(\mathbf{v}) \in \mathbb{R}^{N(c)}$ whereby each component $x_i \in [0, 1]$ determines the soft class assignment of \mathbf{v} to class i for $i = 1, \dots, N(c)$. Accordingly, the association models \mathbf{y}_r of the prototypes are taken as fuzzy class-label vectors $\mathbf{y}_r \in [0, 1]^{N(c)}$ which determines the portion of prototype \mathbf{w}_r associated with the respective classes. We take for the accuracy term $\tilde{E}_{AL} = \vartheta(\mathbf{x}(\mathbf{v}), \mathbf{y}_r)$ describing the dissimilarity of the label vectors \mathbf{x} and \mathbf{y}_r . Usually, the squared Euclidean distance $\vartheta(\mathbf{x}(\mathbf{v}), \mathbf{y}_r) = (\mathbf{x} - \mathbf{y}_r)^2$ is chosen. However, as in the case for the dissimilarity ξ in the data space, other definitions are possible.

The gradient of E_{AL} with respect to \mathbf{w}_r now yields

$$\frac{\partial E_{AL}}{\partial \mathbf{w}_r} = - \int \frac{g_\gamma(\mathbf{v}, \mathbf{w}_r)}{4\gamma^2} \frac{\partial \xi(\mathbf{v}, \mathbf{w}_r)}{\partial \mathbf{w}_r} \vartheta(\mathbf{x}(\mathbf{v}), \mathbf{y}_r) P(\mathbf{v}) d\mathbf{v} \quad (14)$$

Thus the complete prototype update for FLSOM becomes

$$\begin{aligned} \Delta \mathbf{w}_r = & -\epsilon(1 - \beta) \cdot h_\sigma(\mathbf{r}, s(\mathbf{v})) \frac{\partial \xi(\mathbf{v}, \mathbf{w}_r)}{\partial \mathbf{w}_r} \quad (15) \\ & + \epsilon\beta \cdot \frac{g_\gamma(\mathbf{v}, \mathbf{w}_r)}{4\gamma^2} \cdot \frac{\partial \xi(\mathbf{v}, \mathbf{w}_r)}{\partial \mathbf{w}_r} \cdot \vartheta(\mathbf{x}(\mathbf{v}), \mathbf{y}_r). \end{aligned}$$

The gradient of E_{AL-SOM} with respect to the label determines the adaptation rule for the prototype labels. Because E_{SOM} is independent on the prototype labels the respective derivative vanishes and we simply get

$$\Delta \mathbf{y}_r = \epsilon_l \beta \cdot g_\gamma(\mathbf{v}, \mathbf{w}_r) (\mathbf{x} - \mathbf{y}_r) \quad (16)$$

with learning rate $\epsilon_l > 0$. This learning scheme can be seen as a weighted average of the data fuzzy labels of those data \mathbf{v} close to the associated prototype \mathbf{w}_r .

The evaluation of the similarities between the prototype label vectors yields suggestions for the similarity of classes, i.e. similar classes are represented by prototypes in a local spatial area of the SOM lattice. In case of overlapping class

distributions this topographic class processing leads to prototypes with unclear decision, located between prototypes with clear vote. Further, if classes are not distinguish-able, there will exist prototypes responsive to those data which have class label vectors containing approximately the same degree of class membership for the respective classes. In this way FLSOM may be used for class similarity detection.

3 Experiments

We applied the FLSOM in classification of overall 286 proteomic expression profiles generated by MALDI-TOF mass spectroscopy (MS) ranging within a mass-to-charge-ratio of 1000 to 11000Da. The data were used before separately in different investigations [5],[6]. The standardized circumstances for sample collection are described in detail in [7]. Using standardized peak-picking procedure restricted to the range of 1000 to 3500Da the spectra are reduced to 27-dimensional data vectors \mathbf{v} . The scaled, squared Euclidean distance (13) is used as dissimilarity measure.

The data set contains 4 subgroups (A – 63 colorectal cancers, B – 50 controls, C – 76 benign bowels, D – 77 breast cancers), see Fig. 1. It has been postulated that MS

Figure 1. top) Data set of the proteomic profiles (linewise). The darker the pixel the higher the intensity. The 4 subgroups A – D (see text) are separated by lines. bottom) The arrows indicate the most relevant frequencies for distinction of the classes for the 2-class-problem (colorectal cancer versus control) and the full 4-class-problem obtained by FLSOM.

is a suitable tool for detecting presence/absence of multiple low-molecular-weight serum proteins which may serve a potential biomarkers of cancer disease.

To be comparable with the study in [5], we trained in a first investigation a FLSOM with data only of the groups A and B. Thereby we used a 7×3 FLSOM lattice according to non-linear SOM-based PCA [2]. The balancing parameter was $\beta = 0.85$, which emphasizes the classification term in (15) but prevents instabilities for higher values [22]. To be generalizing we used the inherent regularization abilities of SOMs by non-vanishing neighborhood range $\sigma = 0.5$ in the neighborhood function h_σ in (2). We obtained an accuracy (majority vote) of 91.4% in a 5-fold cross-validation, which is approximately the same accuracy as in [5]. The relevance parameters λ_i of the scaled Euclidean metric are adapted parallelly. This leads to a ranking of the input dimensions according to their importance for classification, see Fig. 1. The most important frequencies are indicated by straight arrows, dashed arrows refer to further highly relevant frequencies. The result is also comparable to those obtained by means of the commercial analysis tool ClinProt (standard setting), which is provided with the spectrometer: Both, Support vector machines (SVM) and advanced generalized relevance LVQ (GRLVQ) achieve 94% accuracy.

Subsequently, we additionally proceeded a classification over the full data set, in the same manner as for the 2-class case above. The non-linear PCA suggests a 8×4 lattice for FLSOM, the other parameters remain unchanged. Here, a overall accuracy of 69.5% is achieved. The standard ClinProt tools obtain 69% and 61% for SVM and GRLVQ, respectively. The most important relevances detected by FLSOM are depicted in Fig. 1. The most important frequencies are indicated by straight arrows in Fig. ??, dashed arrows refer to further highly relevant frequencies. These frequencies contribute substantially to classification accuracy and, therefore, are important for distinction of the classes. We see that most relevant parameters are dependent on the classification task.

As mentioned, AL-SOM/FLSOM has the additional feature to detect class similarities because of its underlying topographic mapping property. Therefore, we take a closer look to the distribution of the prototype labels y_r in the FLSOM grid. We could expect that colorectal cancers (A) and benign bowels (C) have more similarities than colorectal cancers and controls (B), because tumor micro-environmental protein expressions are frequently related or analogue to inflammatory diseases [4],[17]. Further, breast cancer patients (D) should have more similarity to class A and C than to the controls. The obtained label distribution determining the similarities is depicted in Fig.2 We observe the fine conformity of the detected class similarities with the clinical expectations. Hence, FLSOM successfully discovered the underlying class structure.

4 Conclusions

We presented a general framework for the extension of the SOM for association learning. We specified this frame-

Figure 2. Label distribution in the FLSOM lattice for the full classification problem. Each square represents a label vector y_r of a prototype w_r . The position is according localization r in the grid. The height of the bars reflects the fuzzy amount for the respective class as indicated above. One recognizes the great overlapping region between the classes A and C (colorectal/benign bowels) and some smaller overlaps between the classes A and D (colorectal/breast) as well as C and D. Otherwise the controls (B) are clearly separated. These findings are in agreement with clinical expectations.

work for fuzzy classification obtaining the FLSOM model, which takes class information of training data explicitly into account for prototype adaptation. The visualization abilities of SOMs based on the topology preservation property of unsupervised SOMs can be used for visual inspection of the class labels of the prototypes, which allows a better understanding of the underlying classification decision scheme. Further, the FLSOM is able to detect class similarities.

The FLSOM is applied to classification of mass spectrometric data (profiles) of different cancer diseases, controls and benign bowels. Beside a comparable classification accuracy the model automatically discovered the class similarities in good agreement to clinical expectation. Thus, FLSOM can be used not only for classification and visualization but also for detecting of class dependencies.

References

- [1] 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.
- [2] 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.
- [3] C. Bishop. *Pattern Recognition and Machine Learning*. Springer Science+Business Media, LLC, New York, NY, 2006.
- [4] H. Chang, J. Sneddon, A. Alizadeh, R. Sood, R. West, K. Montgomery, J.-T. Chi, M. van de Rijn, D. Botstein, and P. Brown. Gene expression signature of fibroblast serum response predicts human cancer progression: Similarities between tumors and wounds. *PLoS Biology*, 2(2):206–214, 2004.
- [5] M. de Noo, A. Deelder, M. van der Werff, A. Özalp, and B. Martens. MALDI-TOF serum protein profiling for detection of breast cancer. *Onkologie*, 29:501–506, 2006.
- [6] M. de Noo, B. Martens, A. Özalp, M. Bladergroen, M. van der Werff, C. van de Velde A. Deelder, and R. Tollenaar. Detecting of colorectal cancer using MALDI-TOF serum protein profiling. *European Journal of Cancer*, 42:1068–1076, 2006.
- [7] M. de Noo, R. Tollenaar, A. Özalp, P. Kuppen, M. Bladergroen, P. Eilers, and A. Deelder. Reliability of human serum protein profiles generated with c8 magnetic beads assisted MALDI-TOF mass spectrometry. *Analytical Chemistry*, 77(22):7232–7241, Nov. 2005.
- [8] R. Der and M. Herrmann. Instabilities in self-organized feature maps with short neighborhood range. In M. Verleysen, editor, *Proc. ESANN'94, European Symp. on Artificial Neural Networks*, pages 271–276, Brussels, Belgium, 1994. D facto conference services.
- [9] E. Erwin, K. Obermayer, and K. Schulten. Self-organizing maps: Ordering, convergence properties and energy functions. *Biol. Cyb.*, 67(1):47–55, 1992.
- [10] B. Hammer, M. Strickert, and T. Villmann. Supervised neural gas with general similarity measure. *Neural Processing Letters*, 21(1):21–44, 2005.
- [11] B. Hammer and T. Villmann. Generalized relevance learning vector quantization. *Neural Networks*, 15(8-9):1059–1068, 2002.
- [12] T. Hastie and W. Stuetzle. Principal curves. *J. Am. Stat. Assn.*, 84:502–516, 1989.
- [13] S. Haykin. *Neural Networks - A Comprehensive Foundation*. IEEE Press, New York, 1994.
- [14] R. Hecht-Nielsen. Counterpropagation networks. *Appl. Opt.*, 26(23):4979–4984, December 1987.
- [15] T. Heskes. Energy functions for self-organizing maps. In E. Oja and S. Kaski, editors, *Kohonen Maps*, pages 303–316. Elsevier, Amsterdam, 1999.
- [16] T. Kohonen. *Self-Organizing Maps*, volume 30 of *Springer Series in Information Sciences*. Springer, Berlin, Heidelberg, 1995. (Second Extended Edition 1997).
- [17] L. Liotta and E. Kohn. The microenvironment of the tumour-host interface. *Nature*, 411:375–379, 2001.
- [18] F.-M. Schleif, T. Elssner, M. Kostrzewa, T. Villmann, and B. Hammer. Analysis and visualization of proteomic data by fuzzy labeled self-organizing maps. In D. Lee, B. Nutter, S. Antani, S. Mitra, and J. Archibald, editors, *19th IEEE International Symposium on Computer-based Medical Systems Salt Lake City (CBMS)*, pages 919–924. IEEE Computer Society Press, Los Alamitos, 2006. 0769525171.
- [19] J. Sinkkonen and S. Kaski. Clustering based on conditional distributions in an auxiliary space. *Neural Computation*, 14:217–239, 2002.
- [20] T. Villmann. Neural maps for faithful data modelling in medicine – state of the art and exemplary applications. *Neurocomputing*, 48(1–4):229–250, 2002.
- [21] 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.
- [22] T. Villmann, B. Hammer, F.-M. Schleif, T. Geweniger, and W. Herrmann. Fuzzy classification by fuzzy labeled neural gas. *Neural Networks*, 19:772–779, 2006.
- [23] T. Villmann, U. Seiffert, F.-M. Schleif, C. Brüß, T. Geweniger, and B. Hammer. Fuzzy labeled self-organizing map with label-adjusted prototypes. In F. Schwenker and S. Marinai, editors, *Proceedings of Conference Artificial Neural Networks in Pattern Recognition (ANNPR) 2006, Ulm, Germany*, LNAI 4087, pages 46–56. Springer Verlag, 2006.
- [24] T. Villmann, M. Strickert, C. Brüß, F.-M. Schleif, and U. Seiffert. Visualization of fuzzy information in fuzzy-classification for image segmentation using MDS. In M. Verleysen, editor, *Proc. Of European Symposium on Artificial Neural Networks (ESANN'2007)*, pages 103–108, Brussels, Belgium, 2007. d-side publications.
- [25] P. Vuorimaa. Fuzzy self-organizing map. *Fuzzy Sets and Systems*, 66(2):223–231, Sept 1994.