ESANN'2007 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 25-27 April 2007, d-side publi., ISBN 2-930307-07-2.

Visualization of Fuzzy Information in Fuzzy-Classification for Image Segmentation using MDS

T. Villmann[1], M. Strickert[2], C. Brüß[2], F.-M. Schleif[1], and U. Seiffert[2]

1- University Leipzig - Medical Department
K.-Tauchnitz-Str. 25, 04107 Leipzig - Germany

2- Leibniz Inst. of Plant Genetics & Crop Plant Research - Pattern Rec. Group
Corrensstraße 3, 06466 Gatersleben - Germany

Abstract. In this work we introduce a method for visualization of fuzzy label information obtained from prototype based fuzzy labeled self-organizing map (FLSOM) for fuzzy classification. FLSOM returns vectors of fuzyy class labels for the prototypes containing class simlarity information. This information is used for apropriate visualization by an adequate, similarity preserving color space embedding realized by an advanced MDS-approach. Using this embedding, classification results can be visualized by the similarity-based color representation of the data. The method is applied for image segmentation with uncertain (fuzzy) class membership.

1 Introduction

Real world data are often characterized by uncertain and possibly inconsistent information. In particular in medical and biological applications this problem plays an important role. If data are processed by machine learning approaches, this uncertainty has to be taken into account. Prototype based fuzzy classification offers a possible solution for learning of labeled data in the context of classification tasks. Respective algorithms are the fuzzy labeled neural gas (FLNG - [1]) or its self-organizing map (SOM) based counterparts FuzzySOM [2] and fuzzy labeled self-organizing map (FLSOM - [3]), which offer better data visualization possibilities due to its underlying regular grid structure of self-organizing maps (SOM - [4]). This advantage is caused by the topology preserving mapping realized by unsupervised SOMs under certain conditions. Usually, the supervised variants return a probability vector assigned to each prototype, which reflects the class probabilities for each possible data class. Yet, FuzzySOM does not minimize a predefined cost function because both parts of the adaptation, the SOM learning and the subsequent learning vector quantization are not based on a gradient descent approach. FLSOM, however, is based on the Heskes' variant of SOMs, for which a cost function is given. The SOM adaptation scheme is imposed by a gradient descent on classification accuracy. Further, the topology preservation may be lost by the subsequent learning in FuzzySOM, whereas it is not in FLSOM.

Yet, the visualization of fuzzy classification information is difficult. One method is to consider the barplots of the probability vectors as demonstrated for FLSOM [5]. However, we can use a special feature of FLSOM. In contrast to FuzzySOM, FLSOM inherently contains the neighborhood cooperativeness also in label adaptation. Therefore, in case of adequate conditions, topological ordering is obtained for both prototype distribution and label vectors. From the

latter one it is possible to detect class similarities, which can be used for improved visualization: We determine a similarity preserving color representation of label vectors, here done by multi-dimensional scaling (MDS) in a faithful similarity preserving manner. In this way we obtain an appropriate fuzzy visualization. We demonstrate the method for a problem of image segmentation in biological structures.

2 Algorithms, notations, and used data

Assume data $\mathbf{v} \in V$ are given distributed according to an underlying distribution $P(V)$. The prototypes are vectors $\mathbf{w_r} \in \mathbb{R}^d$, whereby $\mathbf{r} \in A$, A is an index set. We assume that each training point $\mathbf{v}$ is equipped with a label vector $\mathbf{x} \in [0,1]^{N(c)}$ whereby the component x_i of $\mathbf{x}$ determines the assignment of $\mathbf{v}$ to class i for $i = 1, \ldots, N(c)$. Hence, we can interpret the label vector as probabilistic or possibilistic fuzzy class memberships. In case of crisp labeling, one label component x_i simply is the unit whereas all other are zero. Accordingly, we enlarge each prototype vector $\mathbf{w_r}$ of the map by a label vector $\mathbf{y_r} \in [0,1]^{N(c)}$ which determines the portion of neuron $\mathbf{r}$ assigned to the respective classes. During training by the above algorithms, prototype locations $\mathbf{w_r}$ and label vectors $\mathbf{y_r}$ are adapted according to the given labeled training data $(\mathbf{v}, \mathbf{x})$ following a gradient descent on the cost function

$$E_{\mathrm{FLSOM}} = (1-\beta)\, E_{\mathrm{SOM}} + \beta E_{\mathrm{FL}} \tag{1}$$

where $\beta \in [0,1]$ is a balance factor to determine the influence of the goal of clustering the data set and the goal of achieving a correct labeling. The classification accuracy is assessed by

$$E_{\mathrm{FL}} = \frac{1}{2} \int P(\mathbf{v}) \sum_{\mathbf{r}} g_\gamma(\mathbf{v}, \mathbf{w_r}) (\mathbf{x} - \mathbf{y_r})^2 \, d\mathbf{v} \text{ with } g_\gamma(\mathbf{v}, \mathbf{w_r}) = \exp\left(-\frac{\xi(\mathbf{v}, \mathbf{w_r})}{2\gamma^2}\right) \tag{2}$$

and the E_{SOM}-term is the cost function of Heskes' SOM approach [6]:

$$E_{\mathrm{SOM}} = \frac{1}{2C(\sigma)} \int P(\mathbf{v}) \sum_{\mathbf{r}} \delta_{\mathbf{r}}^{\mathbf{s}(\mathbf{v})} \sum_{\mathbf{r}'} h_\sigma(\mathbf{r}, \mathbf{r}') \xi(\mathbf{v}, \mathbf{w_{r'}}) d\mathbf{v}, \tag{3}$$

We remark that the gradient $\frac{\partial E_{\mathrm{FL}}}{\partial \mathbf{w_r}}$ contributes to the prototype adaptation, and, hence, it is dependent on the classification accuracy. Further, this fact implies the topological ordering also of the class labels at the grid due to the coupling of the two neighborhood functions $g_\gamma(\mathbf{v}, \mathbf{w_r})$ and $h_\sigma(\mathbf{r}, \mathbf{r}')$.

To investigate the degree of fuzzy information possessing, we applied different types of prototype based supervised vector quantizers. The first one is the generalized relevance learning vector quantization (GRLVQ, [7]), which, however, is only able to handle crisp class information for both data and prototypes. Secondly, we apply FLSOM for the crisp labeled data keeping in mind that FLSOM delivers fuzzy labeled prototypes. In the last step we used fuzzy labeled data which are obtained by special fuzzyfication of the crisp data set [8]. The prototype label vectors are then embedded in the RGB-color space by an advanced MDS-scheme described below. However, other embedding schemes are possible

[9]. In this way we get a color representation for each prototype label. Thereby, for GRLVQ this color representation is fixed apriori because the prototype labels are not subject to adaptation, whereas for the other algorithms the prototype labels are adjusted during learning yielding different color representations in dependence on the learned fuzzy classification.

The data for this application are serial transverse sections of barley grains at different developmental stages. Developing barley grains consist of three genetically different tissue types: the diploid maternal tissues, the filial triploid endosperm, and the diploid embryo. Because of their functionality, cells of a fully differentiated tissue show differences in cell shape and water content and accumulate different compounds. Based on those characteristics, scientists experienced in histology are able to identify and to label differentiated tissues within a given section of a developing grain (segmentation). However, differentiating cells lack these characteristics. Because differentiation occurs along gradients, especially borders between different tissue types of developing grains often consist of differentiating cells, which cannot be identified as belonging to one or the other tissue type. Thus, fuzzy processing is highly desirable. However, since (training) examples, manually labelled by a biological expert, are costly and rarely available, one is interested in automatic classification based on a small training subset of the whole data set. In our example, the training set consists of 4418 data points (vectors) whereas the whole transverse section of the image contains 616×986 samples, which finally have to be classified and visualized as an image for immediate interpretation by biologists. The data vectors are 22-dimensional, the number of classes is $N(c) = 11$.

3 High-Throughput MDS (HiT-MDS-2)

For embedding the prototype labels we use an advanced MDS scheme. Generally, MDS refers to the optimization of N point locations $\mathbf{t}_i = \left(t_i^1, \ldots, t_i^{\check{D}}\right) \in \mathbb{R}^{\check{D}}$ in a *target space* in such a way that their distance relationships faithfully reflect those of the *original data vectors* $\mathbf{o}_i \in \mathbb{R}^D$ [10]. Obviously, in case of dimension reduction with $D > \check{D}$ such optimization will need to find a compromise solution. Let $\delta_{i,k} = \delta(\mathbf{o}_i, \mathbf{o}_k)$ be the input space distances and $d_{i,k} = d(\mathbf{t}_i, \mathbf{t}_k) = \mathbf{D}$ be the distance in target space $\mathbb{R}^{\check{D}}$ and $\mathbf{T}$ be the matrix of all target vectors. For Euclidean input and target distances the minimum of the raw point-embedding stress function $S = \sum_{i<k} (\delta_{i,k} - d_{i,k})^2 = \min$ yields target configurations which are equivalent to the linear projections of principal component analysis (PCA). Although the benefit of MDS over PCA is more flexibility in the choice of distance measures, even many alternative MDS related stress functions suffer from the presence of local minima. One avoidable reason for local minima is a too stringent formulation of the stress function. In most metric approaches reconstructed distances are forced to fall onto the unit slope regression line in the corresponding $\delta_{i,k}$-vs.-$d_{i,k}$ Shepard plot. In contrast to that, HiT-MDS-2 maximizes the more relaxed Pearson correlation $r \in [-1; 1]$

$$r = \frac{\sum_{q<k} (\delta_{q,k} - \mu_\delta) \cdot (d_{q,k} - \mu_d)}{\sqrt{\sum_{q<k} (\delta_{q,k} - \mu_\delta)^2 \cdot \sum_{q<k} (d_{q,k} - \mu_d)^2}} =: \frac{\mathcal{B}}{\sqrt{\mathcal{C} \cdot \mathcal{D}}}$$

	GRLVQ	FLSOM
training data set ($N \approx 10000$)	89.5%	77.5%
whole data set ($N \approx 1500000$)	95.8%	73.9%

Table 1: Classification accuracies for the different algorithms using the crisp labeled data. For FLSOM, majority voting is applied.

between entries of the source distances and the reconstructed distances. While, for correlation maximization, powers of $(1-r)^K$ were minimized in the old formulation (HiT-MDS) [11], the new formulation (HiT-MDS-2) makes use of Fisher's Z′-transform for improved convergence of the stress function S:

$$S = -\mathrm{Z}' \circ r \circ \mathbf{D} \circ \mathbf{T} = \min \text{ with } \mathrm{Z}' = \frac{1}{2}\log\left(\frac{a+r}{a-r}\right) \stackrel{!}{=} \max,\ a = 1+\epsilon. \quad (4)$$

Locations of points $\mathbf{t}_i$ in target space are obtained by iterative gradient descent $\triangle t_i^k = -\varepsilon \frac{\partial S}{\partial t_i^k}$ of step size ε on the stress function S using the chain rule:

$$\frac{\partial \mathrm{Z}'}{\partial r} = -\frac{a}{r^2-a^2} \quad (5)$$

$$\frac{\partial r}{\partial d_{i,j}} = \frac{(\delta_{i,j}-\mu_\delta)\cdot\mathcal{D} - (d_{i,j}-\mu_d)\cdot\mathcal{B}}{\mathcal{D}\cdot\sqrt{\mathcal{C}\cdot\mathcal{D}}} \quad (6)$$

$$\frac{\partial d_{i,j}}{\partial t_i^k} = \frac{2\left(t_i^k - t_j^k\right)}{d_{i,j}} \quad \text{(for Euclidean target space)} \quad (7)$$

In addition to better convergence than the old HiT-MDS with its critical choice of the combination of exponent K and learning rate, HiT-MDS-2 makes use of two non-critical parameters, the learning rate $\varepsilon = 0.1$ and $\epsilon = 0.001$ in Z′ to prevent infinity that could occur in Fisher's original formula with $a = 1$.

In the applications here, we identify the original data vectors of MDS with the prototype labels $\mathbf{y_r}$.

4 Application

We applied the above mentioned algorithm to the available exemplary data set. We used for GRLVQ 55 prototypes (5 for each class). We used a 7×7 FLSOM of approximately the same complexity and to meet the topographic mapping requirements for SOMs assessed by the topographic product [12]. For FLSOM we applied both the crisp labeled data and the fuzzified version. For comparison, we determined the classification accuracy for crisp data by matching for GRLVQ and by majority voting for FLSOM. The results are depicted in Tab. 1. For result visualization we embedded both prototype vectors and prototype label vectors into the RGB-color space separately using the similarity preserving HiT-MDS for each configuration. The three-dimensional prototype embeddings for GRLVQ and FLSOM are visualized in Fig. 1. One can clearly observe the improved color coding by the utilization of FLSOM class fuzzy labels, which inherently contain

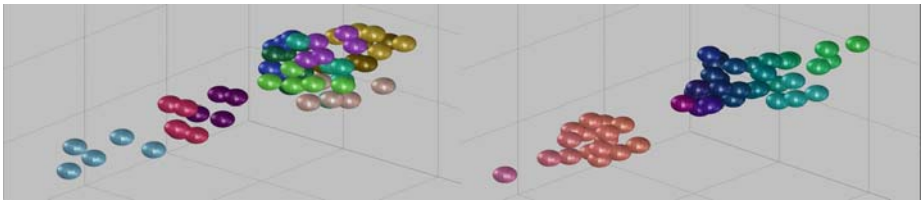

Fig. 1: Color representation of labels for GRLVQ (left) and FLSOM (right) prototypes obtained by HiT-MDS. The learned class similarities represented by similar colors for FLSOM labels are obvious, whereas for GRLVQ all class similarities are assumed as equidistant.

the class similarity information learned by FLSOM, whereas class similarities for GRLVQ assumed to be equidistant. Further, the visualization of the full image according to the obtained (fuzzy) classification and applied color space embedding is depicted in Fig. 2 [1]. We remark the improved visualization by the usage of fuzzy classification. In particular, the usage of class similarity information detected by FLSOM yields a more adequate color representation.

5 Conclusion

In this paper we demonstrate the usage of (advanced) MDS for visualization of fuzzy classification results. For an exemplary image segmentation application of transversal barley grain sections we obtain good visualization results.

This work was supported by the German Ministry of Education and Research (grants no. 0313115 and 0312706A).

References

[1] 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.

[2] Petri Vuorimaa. Fuzzy self-organizing map. *Fuzzy Sets and Systems*, 66(2):223–231, Sept 1994.

[3] 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.

[4] Teuvo Kohonen. *Self-Organizing Maps*, volume 30 of *Springer Series in Information Sciences*. Springer, Berlin, Heidelberg, 1995. (Second Extended Edition 1997).

[5] 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.J. 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.

[6] T. Heskes. Energy functions for self-organizing maps. In E. Oja and S. Kaski, editors, *Kohonen Maps*, pages 303–316. Elsevier, Amsterdam, 1999.

[1] A color version of the figures can be obtained from the authors on request.

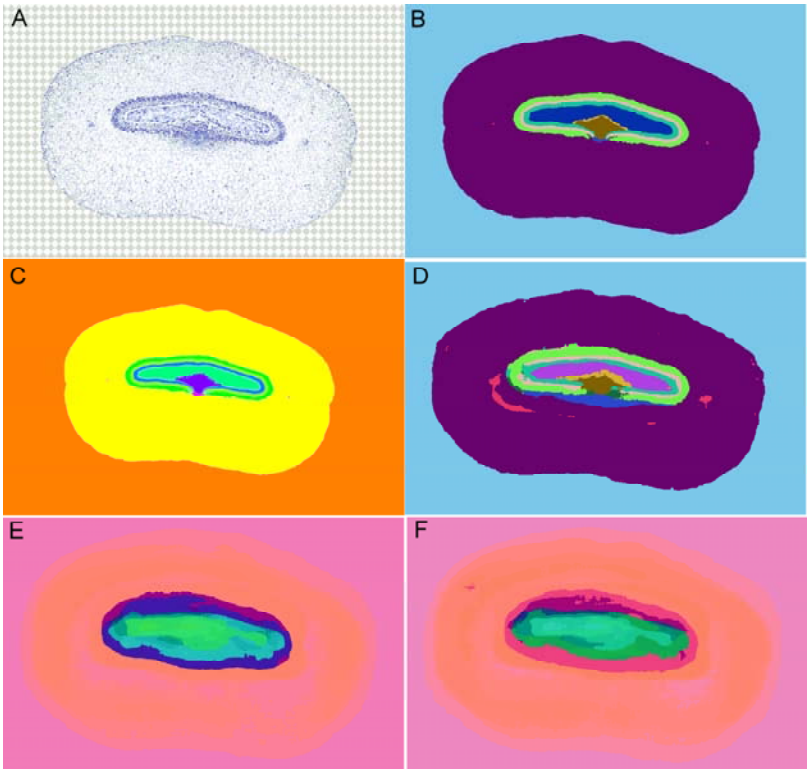

Fig. 2: Image segmentation results for FLSOM and GRLVQ - color coding is obtained by HiT-MDS-2. **A**: original image, **B**: manually crisp labeled data, **C**: manually fuzzy labeled data, **D:** GRLVQ for crisp data, **E**: FLSOM for crisp data, **F:** FLSOM for fuzzy data.

[7] B. Hammer and Th. Villmann. Generalized relevance learning vector quantization. *Neural Networks*, 15(8-9):1059–1068, 2002.

[8] C. Brüß, F. Bollenbeck, F.-M. Schleif, W. Weschke, Th. Villmann, and U. Seiffert. Fuzzy image segmentation with fuzzy labeled neural gas. In M. Verleysen, editor, *Proc. of European Symposium on Artificial Neural Networks (ESANN'2006)*, pages 563–568, Brussels, Belgium, 2006. d-side publications.

[9] C.M. Bishop. *Pattern Recognition and Machine Learning*. Springer Science+Business Media, LLC, New York, NY, 2006.

[10] A. Buja, D. Swayne, M. Littman, N. Dean, and H. Hofmann. Interactive Data Visualization with Multidimensional Scaling. Report, University of Pennsylvania, 2004. http://www.ggobi.org/.

[11] M. Strickert, S. Teichmann, N. Sreenivasulu, and U. Seiffert. High-Throughput Multi-Dimensional Scaling (HiT-MDS) for cDNA-array expression data. In W. Duch et al., editor, *Artificial Neural Networks: Biological Inspirations, Part I, LNCS 3696*, pages 625–634. Springer, 2005. http://hitmds.webhop.net/.

[12] 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.