ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

Border Sensitive Fuzzy Vector Quantization in Semi-Supervised Learning

Tina Geweniger, Marika Kästner, Thomas Villmann

Computational Intelligence Group, University of Applied Sciences Mittweida, Technikumplatz 17, 09648 Mittweida, Germany

Abstract. We propose a semi-supervised fuzzy vector quantization method for the classification of incompletely labeled data. Since information contained within the structure of the data set should not be neglected, our method considers the whole data set during the learning process. In difference to known methods our approach uses neighborhood cooperativeness for stable prototype learning known from Neural Gas. Further improvement of the classification accuracy is achieved by including class border sensitivity inspired by Support Vector Machines again improved by neighborhood learning.

1 Motivation

Supervised classification based on labeled data and unsupervised clustering based on unlabeled data are common tasks in the field of machine learning. There exist a variety of algorithms for either paradigm. Some famous clustering methods are c-Means, Self Organizing Maps (SOM), Neural Gas (NG), Affinity Propagation, and variants thereof to improve the performance, consider overlapping data (fuzziness), incorporate neighborhood relations or to attend to sparsity to name just a few. On the other hand there are classifiers like Learning Vector Quantizers (LVQ, GLVQ, RSLVQ etc.) or Nearest Prototype Classifiers (NPC, SNPC, FSNPC etc.) to solve classification problems.

Obtaining labeled data sets often is a difficult and costly procedure requiring expert knowledge and – especially if the labelling has to be done manually – a considerable amount of time. Therefore, sometimes only a fraction of a data set is labeled impeding complete classification learning. To utilize the above mentioned methods for this kind of data, either the labels are neglected to do unsupervised clustering or the data set itself is reduced considering only labeled data samples for the classification. In either case, information is lost. For this reason several semi-supervised classification methods namely FLSOM and FLNG [1] have been developed for crisp data. These methods closely follow the standard SOM and NG procedures yet also take labeled data into consideration. Pedrycz proposed an alternative semi-supervised vector quantization scheme based on Fuzzy c-Means (FCM) [2], but also pays attention to the labeled data samples. For this partially supervised clustering the original FCM cost function is extended by an additional term expressing a level of coincidence between the FCM membership degrees and the expert provided class information [3, 4]. This way the class information as well as the structure of the data inherent in the whole data set is taken into consideration.

Further, special interest frequently is given to the knowledge about decision borders between classes. This problem is explicitly addressed in Support Vector Machine (SVM) learning, which determines so-called support vectors approximating and indicating the borders between the classes [5, 6]. Recently, the idea

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

of emphasizing the class borders while training several instances of unsupervised (Fuzzy) c-Means algorithms (FCMs) based on the Euclidean distance is provided in [7]. In this model several FCM instances interact with each other while learning the classification task.

In this paper we propose partially supervised class border sensitive vector quantization by taking up the idea of partially supervised clustering as introduced in [3, 4] and extending this method with class border sensitive prototype positioning similar to [7]. For the latter aspect we extend the algorithm proposed in [8] by introducing neighborhood cooperativeness similar to FSOM [9] and FNG [10] on both levels, i. e. the prototype learning as well as the class border detection. Thus, both learning processes are stabilized, which drastically reduces the problem of getting stuck in local minima.

2 Partially supervised Vector Quantization

Commonly, the initial situation for performing any kind of clustering is an unlabeled data set. Yet in those cases, where (only partial) label information is available, this additional information should be considered adequately in the learning process. For crisp vector quantization schemes like SOM and NG the respective variants are proposed in [11, 12, 13]. In the context of fuzzy vector quantization two methods stand out: Cebron & Berthold combined unsupervised FCM with LVQ to handle incomplete labelling [14] and Pedrycz proposed Partially Supervised FCM (PS-FCM), where semi-supervised learning is integrated in the framework of FCM [3, 4]. Following we provide details of the latter method PS-FCM. The generalization to other fuzzy vector quantization methods is straight forward and follows afterwards.

In the first step each data point $\mathbf{v}_i \in V$ is equipped with a boolean variable $b_i = 1$ if a class label is available, else b_i is set to zero. The fuzzy data labels are stored in a matrix $\mathbf{F} \in \mathbb{R}^{N\times M}$ with $f_{i,k} \in [0, 1]$. The objective function is

$$J(\mathbf{U}, V, W, \mathbf{F}, \xi) = \sum_{k=1}^{M}\sum_{i=1}^{N}\left[(1-\xi)\cdot u_{i,k}^{m} + \xi\cdot(u_{i,k} - f_{i,k}b_i)^m\right](d_{i,k})^2 \tag{1}$$

with a convex weighting factor $\xi \in [0, 1)$ for the influence of the labeled data, the fuzzy assignments $u_{i,k}$ stored in $\mathbf{U}$ as known from FCM, and a distance $d_{i,k}$, usually the Euclidean distance. In this model it is assumed that each class C is represented by exactly one prototype $\mathbf{w}_k \in W$, i. e. $C = M$. [3]

Obviously, this assumption is not realistic. Therefore Bouchachia & Pedrycz suggested the following generalization to deal with this problem [4]: They consider a number of classes $C \leq M$ and a partition of the prototypes such that M_c prototypes are responsible for class $c \in \{1, \ldots, C\}$. Hence, $\sum_c M_c = M$ is valid. The cost function to be minimized is defined as

$$\tilde{J}\left(\mathbf{U}, \tilde{\mathbf{U}}, V, W, \tilde{\mathbf{F}}, \xi\right) = \sum_{k=1}^{M}\sum_{i=1}^{N}\left[(1-\xi)\cdot u_{i,k}^{m} - \xi\cdot(u_{i,k} - \tilde{u}_{i,k})^m\right](d_{i,k})^2 \tag{2}$$

whereby the new assignments $\tilde{u}_{i,k}$ are optimized according to stochastic gradient of the cost function

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

$$Q\left(\tilde{\mathbf{U}},\tilde{\mathbf{F}}\right)=\sum_{c=1}^{C}\sum_{i=1}^{N}b_i\left(\tilde{f}_{c,i}-\sum_{k=1}^{M}\delta_c\left(k\right)\tilde{u}_{i,k}\right)^2 \tag{3}$$

where the value $\delta_c\left(k\right)=1$ holds if the prototype $\mathbf{w}_k$ is defined to be responsible for class c and zero elsewhere. The matrix $\tilde{\mathbf{F}}\in\mathbb{R}^{C\times N}$ plays the role of given fuzzy assignments $\tilde{f}_{c,i}\in[0,1]$ for $\mathbf{v}_i$ to the class c. The value $\vartheta_{c,i}=\tilde{f}_{c,i}-\sum_{k=1}^{M}\delta_c\left(k\right)\tilde{u}_{i,k}$ triggers the minimization of the difference between (fuzzy) class assignments $\tilde{u}_{i,k}$ and the prototype assignments $u_{i,k}$ for the labeled data. In consequence, both quantities $\tilde{J}$ and Q have to be minimized in parallel [4].

The update rule for the class assignments $\tilde{u}_{i,k}$ necessary for calculating (2) is obtained by optimizing (3) and yields

$$\Delta\tilde{u}_{i,k}=2b_i\sum_{c=1}^{C}\left(\tilde{f}_{c,i}-\sum_{s=1}^{M}\delta_c(s)\tilde{u}_{i,s}\right). \tag{4}$$

The updates of the prototypes $\mathbf{w}_k$ and the fuzzy prototype assignments $u_{i,k}$ are derived as solutions of the Lagrange minimization problem

$$\tilde{J}_i\left(\mathbf{U},\tilde{\mathbf{U}},V,W,\tilde{\mathbf{F}},\xi\right)=\sum_{k=1}^{M}\left[(1-\xi)\cdot u_{i,k}^m+\xi\cdot\left(u_{i,k}-\tilde{u}_{i,k}\right)^m\right]\left(d_{i,k}\right)^2-L_i. \tag{5}$$

with the Lagrange term $L_i=\lambda_i\left(\sum_{k=1}^{M}u_{i,k}-1\right)$. For the special case of the quadratic Euclidean distance and the fuzzifier $m=2$ the update rules yield

$$\mathbf{w}_l=\frac{\sum_{i=1}^{N}\left[(1-\xi)\cdot u_{i,l}^2+\xi\cdot\left(u_{i,l}-\tilde{u}_{i,l}\right)^2\right]\mathbf{v}_i}{\sum_{i=1}^{N}\left[(1-\xi)\cdot u_{i,l}^2+\xi\cdot\left(u_{i,l}-\tilde{u}_{i,l}\right)^2\right]} \tag{6}$$

$$u_{i,s}=\frac{1-\frac{\xi}{2\xi-1}\sum_{j=1}^{M}\tilde{u}_{i,j}}{\sum_{j=1}^{M}\left(\frac{d_{i,s}}{d_{i,j}}\right)^2}+\frac{\xi\cdot\tilde{u}_{i,s}}{2\xi-1} \tag{7}$$

and have to be performed in alternating adaptation steps analogously to FCM. Note that this idea can also be transfered to other FCM variants proposed by Bezdek.

Neighborhood learning for prototypes easily can be incorporated simply by replacing the squared dissimilarities $\left(d_{i,k}\right)^2$ by local costs known from FSOM [9] and FNG [10]:

$$lc_\sigma^{SOM/NG}\left(i,k\right)=\sum_{l=1}^{M}h_\sigma^{SOM/NG}\left(k,l\right)\cdot\left(d_{i,l}\right)^2 \tag{8}$$

with the respective neighborhood functions

$$h_\sigma^{SOM}\left(k,l\right)=c_\sigma^{SOM}\cdot\exp\left(-\frac{\left(d_A\left(k,l\right)\right)^2}{2\sigma^2}\right) \tag{9}$$

$$h_\sigma^{NG}\left(k,l,W\right)=c_\sigma^{NG}\cdot\exp\left(-\frac{\left(rk_k\left(\mathbf{w}_l,W\right)\right)^2}{2\sigma^2}\right) \tag{10}$$

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

referring to the relation $d_A(k,l)$ of the prototypes to each other according to the external grid structure A (FSOM) or the distance ranks $rk_k(\mathbf{w}_l, W)$ of the prototypes to each other (FNG). Further details (also concerning the here left undefined variables) can be found in the mentioned articles.

3 Class Border Sensitive Vector Quantization

As mentioned before, border sensitive prototypes as known from SVMs (support vectors, [5, 6]) are an interesting feature exploring the class related data space. The in [8] proposed method positions prototypes in a LVQ-like manner, where close prototypes belonging to different classes are moved towards each other. It is assumed that the data are partitioned into subsets $V = V_1 \cup V_2 \cup \cdots \cup V_C$ based on known class assignments and in the first step a separate FCM network for each subset is utilized to place the prototypes as known from FCM. In the second step selected prototypes of different classes respectively FCM networks are drawn to each other, i.e. in direction of the class borders. An additional attraction force realizes the information transfer between the FCM networks, which means that neighborhood relations of the prototypes across the class borders are taken into account to move the prototypes in direction of the class borders. We combine this method with the PS-FCM as described in the last section and reduce the model to one FCM network, which takes the available (partial) class information into account implicitly.

The cost function of the model proposed in [8, 7] consists of a sum over the costs of the C seperate FCM networks and an additional force term:

$$E_{BS} = \sum_{l=1}^{C} \sum_{k=1}^{M_l} \sum_{i=1}^{N_l} u_{i,k}^m(l)(d_{i,k})^2 + F_C(W, V). \tag{11}$$

The force term is defined as $F_C(W,V) = \sum_{i=1}^{N} d\left(\mathbf{w}_{s^+(i)}, \mathbf{w}_{s^-(i)}\right)$ where $s^+(i)$ and $s^-(i)$ are determining the closest prototype of the correct class and the closest prototype of all incorrect classes for a given data vector $\mathbf{v}_i$.

We modify this method in several ways. First, we replace the first term of (11) by the PS-FCM cost function $\tilde{J}$ (2). This way a method utilizing just one network and being able to handle data sets with partial class information is obtained. And secondly, knowing the limits of FCM we introduce neighborhood cooperativeness to improve convergence. For the first term it is straightforward replacing the distance $(d_{i,k})^2$ with the local costs $lc_\sigma(i,k)$ (8). More attention has to be given to an appropriate redefinition of the attraction force $F_C(W,V)$ from (11). Let W_i^- be the set of all prototypes which are assigned to different classes than the class c_i for a given data point $\mathbf{v}_i$ and

$$h_{\sigma_-}^{NG}(k,l,W^-) = c_{\sigma_-}^{NG} \cdot \exp\left(-\frac{(rk_k(\mathbf{w}_l, W^-))^2}{2\sigma_-^2}\right) \tag{12}$$

be a NG-like neighborhood function according to (10) but restricted to W^- with neighborhood range σ_-. Then the new *neighborhood-attentive* attraction force

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

is defined as

$$F_{neigh}(W,V) = \sum_{i=1}^{N} \sum_{k=1 \wedge \mathbf{w}_k \in W^-}^{M} h_{\sigma_-}^{NG}\left(k, s^+(i), W^-\right) d\left(\mathbf{w}_{s^+(i)}, \mathbf{w}_k\right) \tag{13}$$

which reduces to $F_C(W,V)$ for $\sigma_- \to 0$. The term $F_{neigh}(W,V)$ again compels the prototypes to move to the class borders. However, the neighborhood cooperativeness speeds up and stabilizes this process scaled by neighborhood range σ. Thereby, the responsibilities of the prototypes for the different class borders of a certain class are not predetermined. Rather they are results of a self-organizing process, which provides a great robustness and stability. The new cost function yields

$$E_{BSneigh} = \sum_{k=1}^{M} \sum_{i=1}^{N} \left[(1-\xi) \cdot u_{i,k}^m - \xi \cdot \left(u_{i,k} - \tilde{u}_{i,k}\right)^m\right] lc_\sigma\left(i,k\right) + F_{neigh}\left(W,V\right). \tag{14}$$

Depending on the chosen neighborhood cooperativeness we refer to this new method as Border Sensitive FCM (BS-FCM), Border Sensitive FSOM (BS-FSOM), or Border Sensitive FNG (BS-FNG).

4 Illustrative Experiments

In this section we present two illustrative examples to demonstrate the improvement in classification accuracy achieved by considering a classification dataset with partially missing label information.

The first dataset consists of two two-dimensional Gaussian distributions, where we provided appropriate class labels for a part of each distribution, see Fig. 1 (left). Comparing the results obtained by performing BS-FNG according to (14) only for the labeled data and for the complete but partially labeled data set shows a significant improvement of the test classification accuracy from 0.64 to 0.98. Using the whole dataset in semi-supervised learning causes class specific prototypes to be positioned also in those regions of the data space, for which no class information is provided.

In the second experiment addressed to class border sensitive learning, two-dimensional data samples are randomly distributed as pictured in Fig. 1 (right). About 10% of the data points selected by chance are labeled referring to two classes. In the course of the learning process the prototypes move in direction of the class border and take class responsibility according to the surrounding labeled data samples. Classifying test data, which also cover the gap, yields an accuracy of 0.97 for semi-supervised learning as compared to 0.72 obtained by ignoring unlabeled data.

Further, we observe that the prototype density near the class border is higher than elsewhere as a consequence of border sensitive learning.

5 Conclusion

In this paper we propose neighborhood and class border sensitive learning in fuzzy vector quantization for partially labeled data. During the learning process the whole data set is considered, taking class information into account if

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

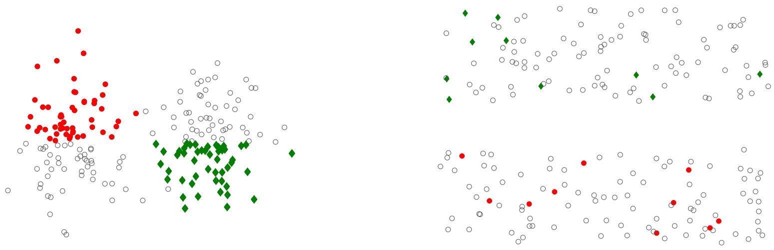

Fig. 1: Partially labeled Gaussian distributions (left) and randomly distributed data points (right). The class labels represent two different classes.

available. The work proposed in [3, 4] is based on FCM, yet was extended here to incorporate neighborhood cooperativeness to stabilize the learning process and ensure convergence. Further, inspired by [8] we added a force term to the cost function which makes the prototypes move in direction of the class borders again incorporating neighborhood cooperativeness. After all, we showed in illustrative examples that the classification accuracy of our method considering all data samples significantly improves compared to the classification based on labeled data only and that class border sensitive learning leads to an improved border detection.

References

[1] T. Villmann, B. Hammer, F.-M. Schleif, T. Geweniger, and W. Herrmann. Fuzzy classification by fuzzy labeled neural gas. *Neural Networks*, 19(6-7):772–779, 2006.

[2] J.C. Bezdek. A convergence theorem for the fuzzy isodata clustering algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(1):1–8, 1980.

[3] W. Pedrycz and J. Waletzky. Fuzzy clustering with partial supervision. *IEEE Transactions on System, Man, and Cybernetics - Part B*, 27(5):787–795, October 1997.

[4] A. Bouchachia and W. Pedrycz. Data clustering with partial supervision. *Data Mining and Knowledge Discovery*, 12:47–78, 2006.

[5] J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis and Discovery.* Cambridge University Press, 2004.

[6] B. Schölkopf and A. Smola. *Learning with Kernels.* MIT Press, 2002.

[7] C. Yin et al. Using cooperative clustering to solve multiclass problems. In Y. Wang and T Li, editors, *Foundation of Intelligent Systems - Proc. of the 6th Internat. Conf. on Int. Systems and Knowledge Eng. (ISKE 2011)*, volume 122, pages 327–334. Springer, 2012.

[8] Shengfeng et al. Tian. Cooperative clustering for training svms. In *Adv. in Neural Networks - ISNN 2006*, LNCS 3971, pages 962–967. Springer, 2006.

[9] M. Kästner, W. Hermann, and T. Villmann. Integration of structural expert knowledge about classes for classification using the fuzzy supervised neural gas. In *ESANN'2012*, pages 209–214, 2012.

[10] M. Kästner, M. Lange, and T. Villmann. Fuzzy supervised self-organizing map for semi-supervised vector quantization. In *AISC - Proc. the International Conference ICAISC, Zakopane*, volume 1 of *LNAI 7267*, pages 256–265. Springer, 2012.

[11] B. Hammer, M. Strickert, and Th. Villmann. Supervised neural gas with general similarity measure. *Neural Processing Letters*, 21:21–44, February 2005.

[12] M. Kästner and Th. Villmann. Fuzzy supervised neural gas for semi-supervised vector quantization – theoretical aspects. Machine learning report, University of Bielefeld, 2011.

[13] T. Villmann, E. Merényi, and W.H. Farrand. Unmixing hyperspectral images with fuzzy supervised self-organizing maps. In *ESANN'2012*, pages 185–190, 2012.

[14] N. Cebron and M.R. Berthold. Adaptive fuzzy clustering. In *Proc. of the Annual Meeting of the North American Fuzzy Inf. Proc. Soc. (NAFIPS) 2006*, pages 188 – 193. IEEE Society Press, 2006.