ESANN'2009 proceedings, European Symposium on Artificial Neural Networks - Advances in Computational Intelligence and Learning. Bruges (Belgium), 22-24 April 2009, d-side publi., ISBN 2-930307-09-9.

Neural Maps and Learning Vector Quantization - Theory and Applications

Frank-Michael Schleif[1] **and Thomas Villmann**[2]

1- University Leipzig, Dept. of Medicine, Semmelweisstrasse 10, 04103 Leipzig, Germany
2- University of Applied Sciences, Technikumplatz 17, 09648 Mittweida, Germany

Abstract. Neural maps and Learning Vector Quantizer are fundamental paradigms in neural vector quantization based on Hebbian learning. The beginning of this field dates back over twenty years with strong progress in theory and outstanding applications. Their success lies in its robustness and simplicity in application whereas the mathematics beyond is rather difficult. We provide an overview on recent achievements and current trends of ongoing research.

Keywords: Self organizing maps, vector quantization, unsupervised and supervised learning

1 Introduction

Supervised and unsupervised vector quantization is a basic task in many machine learning applications. The main players in unsupervised prototype based learning can be identified by the family of *c-mean* algorithms and its neural counterparts *Self Organizing Map* (SOM) [24] and *Neural Gas* (NG) [27]. In (semi-)supervised learning the primal prototype based learning models are the family of *learning vector quantizers* (LVQ) and supervised extensions of SOM and NG. Subsequently we tackle each of these paradigms to highlight current trends and recent developments.

2 Unsupervised Learning

Unsupervised vector quantization can be seen as a projection of possibly high-dimensional data of an input space onto a set of prototypes, which may have an external ordering defined by outer structures. Example for such outer structures may be regular grids, graphs, trees and so on. In case of such structures we result a mapping from the data space to this output structure. The generic principle behind vector quantization is the representation of data by its most similar prototype of the vector quantization model or its equivalent in the outer structure. The most prominent representant of unsupervised vector quantizers without outer-structure is the famous c-means algorithm [19].

Neural maps are biologically based vector quantizers. Here the adaptation schemes for the prototype vectors are usually based on Hebbian learning. Moreover the external structure can also be motivated by biological models like the SOM. However, the basic principle of representation by most similar prototypes is kept.

2.1 Convergence, energy functions and topographic mapping

For the famous c-means the convergence has been shown in [37], but it has been found to be rather instable with respect to its initialization. This motivates extensions such as fuzzy-c-means [1] or SOM/NG. Related to c-means are approaches based on the principle of statistical physics like deterministic annealed learning vector quantization. Subsequently, the initial work of Kohonen given in [23, 22, 24] has provided a new neural paradigm of prototype based vector quantization. The SOM is the most applied neural vector quantizer [24], having a regular low dimensional grid as an external topological structure between prototypes. Yet, for its original algorithm, the convergence analysis is very difficult and only solved for special (simple) cases. A redefinition of being the most similar prototype leads to the Heskes variant of the SOM [20]. The respective learning follows a stochastic gradient descent on a cost function which ensures the convergence. The NG circumvents the convergence problem of the original SOM by the introduction of a dynamical external neighborhood structure of prototypes, which is determined by the shape of the data space to be partitioned. Both approaches, SOM and NG, have in common that an annealing scheme for the range of the neighborhood cooperativeness is used. However, it is not a unique correspondence to the deterministic annealing according to the principles of statistical physics [52, 20].

The statistical properties of the data are implicitly preserved by the magnification property [51]. Whereas for NG the structure preserving mapping (topology preservation) is inherently given, this property may be violated in SOMs due to the predefined external grid structure. Growing grids as well as more complicated grid structures like hyperbolic lattices or graphs and trees allow a broader range of SOM applications without loosing topography [29, 28]

2.2 Variants for specific data structures

Originally most of the vector quantizers were designed for the analysis of Euclidean data. In this case the derivatives of the distance are directly feeded in the update equations for prototypes. In case of gradient descent learning this follows from the derivatives of the cost function. In the last years the appropriate choice of other (differentiable), problem specific, metrics become popular, such as: Lee metric [26] as a generalization of the Minkowski metric, Sobolev norms [50] or kernels there of. In this line also kernelized SOMs come apart where direct distance calculations are replaced by scalar products in a kernel space [49, 48]. This also motivates kernelized vector quantization as shown by Geist [7] and others [47]. Initials steps on unsupervised prototype base feature selection by relevance learning have been done in [44]

A subsequent analysis of prototypes can be applied for robust cluster analysis, which is less sensitive to noise or initialization, compared to traditional cluster analysis techniques, applied to original data. Also prototype based fuzzy clustering provides a substantial key to advanced cluster analysis as shown in [2, 21].

If only similarities between data points are available, traditional approaches are not applicable. For those data structures batch variants are used. In particular, the median or relational variants of the vector quantizers given above are of interest [17, 16, 6]. A median based fuzzy c-means for similarity data is presented by Geweniger et al. [8].

Other extensions of SOM may deal with even more complicated data structures like hyperbolic spaces or, as shown here by Rossi et al. [29] for graph structures. Another special class of data structures occurs by means of time-series and time-related data. Here recursive variants of the SOM and NG take into account the special data structure by context learning [41, 25, 11, 42]. Yet, theoretical properties of these algorithms are not completely understood so far. Special variants of vector quantizers for mathematical and geometrical objects, like manifolds and subspaces, are also in the focus of interest [45].

3 Supervised and Semi-Supervised Learning

Supervised learning requires additional label information during learning. This labeling may be provided in form of class labels, potential continuous output attributes (regression), graduated class assignments (fuzzy label) or auxiliary data relations [40].

In prototype based classification this additional information is utilized for improved class separation compared to unsupervised learning with simple post-labeling, but keeping the Hebbian learning paradigm.

3.1 Convergence, energy functions and relation to margin optimization

One of the most popular learning vector quantizer is standard LVQ as introduced by Kohonen (in its variants) [24]. These approaches are heuristically motivated. Yet, they offer a rich variability in behavior and a mathematical analysis of this is quite difficult [61]. Recent approaches for stability and convergence analysis are based on methods from statistical physics [3, 9]. Newer extensions of LVQ try to overcome the stability problem by different strategies as improved windowing for standard LVQ or probabilistic modeling like Soft Nearest Prototype Classification (SNPC) or Robust Soft-Learning Vector Quantization [39, 38] and neighborhood cooperativeness as shown in Supervised Neural Gas (SNG) [13, 58]. The latter ones have in common that learning follows a (stochastic) gradient descent on a cost function. The generalized LVQ also belongs to this cost function based methods as one of the earliest cost function based extensions of standard LVQ [30].

It is straight forward that prototype based classification is strongly related to empirical risk minimization (ERM) and margin optimization. While in traditional ERM the sample margin is optimized, prototype based classification focus on distance based margins [34, 12, 3]. These distance based optimization is applicable also for faster learning strategies e.g. active learning [34, 18]. Thus prototype based classification can be seen as an alternative approach to Support Vector Machines (SVM) which focus on structural risk minimization (SRM) [14].

A new methodology for development of new classification algorithms, also applicable for semi-supervised learning, is to equip cost function based unsupervised learning approaches with an additional term judging the classification accuracy. Both optimization subjects can be weighted gradually. Thus statistical data information like densities and shape are merged with class label information for classification decision. For example the Heskes variant of SOM and NG can serve as generic starter for such algorithms.

These approaches can also be used for class visualization according to their underlying topographic properties [4, 55] which also contributes to stability in learning. Another approach for visualization is the Exploratory Observation Machine (XOM) for parallel structure preserving dimensionality reduction and data clustering [60].

3.2 Variants for specific data structures

As in unsupervised learning, the underlying data embedding assumption is also in the case of supervised prototype based learning an Euclidean one. However, the incorporation of label information offers extended possibilities to improve the modeling. The approach with the most impact following this line is a data specific metric adaptation taking the label information into account. Prominent candidates of this type are the scaled Euclidean metric, which weights each data dimension for improvement of the classification performance. Thereby the scaling parameter are determined taking the labeling into account and improving e.g. class separation. Multiple prototype based methods have been extended in this way [35, 59, 15]. The straight forward generalization is to use the Mahalanobis distance whereby the positive semi definite matrix of the respective bilinear form is optimized subject to the classification performance [32, 31, 36]. Although the number of adaptation parameters is drastically increased, the convergence is self-stabilizing in terms of eigen-analysis [31]. This behavior is also observed in pure matrix learning of bilinear forms [43]. From this methodology a projection technique can be derived, providing a supervised projection of the data on a lower dimensional space as shown in [5]. This technique can be used for optimal visualization of class separation.

Generally, these matrix based approaches can be seen as classifiers taking linear relations into account. Higher order correlations can be incorporated in feature selection for classification performance improvement, by utilization of information theoretic measures for example entropies, mutual information or divergence measures [46, 53]. Obviously all kernel based vector quantizers take non-linear information into account, too.

If only incomplete label information is available semi-supervised methods can combine unsupervised and supervised learning such that as much as possible information is extracted from the data. For this purpose the above mentioned supervised extensions of the standard unsupervised prototype based algorithms offer an appropriate learning scheme [57, 33, 56, 55]. These approaches have the additional feature that a gradually classification (fuzzy) is provided. Another fuzzy classifier is the FSNPC as the fuzzy extension of SNPC, however it is not semi-supervised[54].

As for unsupervised learning, many of the prototype based classifier can be extended such that they are able to deal only with dissimilarity data. These batch/median or relational variants include the Supervised Relational NG and Relational SOM [10] to name just a few. They offer a greater variability according to the underlying similarity structure which is not necessarily assumed to be a metric in case of median based algorithms. Another strategy would be to incorporate the knowledge about the structural nature of the data by means of special norms like Lee or Sobolev norms for functional data as outlined for unsupervised learning above. Other specific norms could be general minkowski norm or hyperbolic distances.

4 Conclusion

This introductory paper spotlights recent developments, achievements and research foci of prototype based supervised and unsupervised vector quantization. It can be stated that also while quite established, their are still many interesting open problems. Moreover the field offers continuously new challenging questions both, in theory and driven by outstanding applications. Some of them can be found in this volume as contributions to the special session dedicated to this topic at this European Symposium on Neural Networks 2009.

References

[1] J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York, 1981.

[2] J. C. Bezdek, E. C. K. Tsao, and N. R. Pal. Fuzzy Kohonen clustering networks. In *Proc. IEEE International Conference on Fuzzy Systems*, pages 1035–1043, Piscataway, NJ, 1992. IEEE Service Center.

[3] M. Biehl, A. Gosh, and B. Hammer. Dynamics and generalization ability of lvq algorithms. *Journal of Machine Learning Research*, 8:323–360, 2007.

[4] C. Brüß, F. Bollenbeck, F.-M. Schleif, W. Weschke, T. Villmann, and U. Seiffert. Fuzzy image segmentation with fuzzy labelled neural gas. In *Proc. of ESANN 2006*, pages 563–569, 2006.

[5] K. Bunte, P. Scheider, B. Hammer, F.-M. Schleif, T. Villmann, and M. Biehl. Discriminative visualization by limited rank matrix learning. *Machine Learning Reports*, 2(MLR-03-2008), 2008. ISSN:1865-3960 http://www.uni-leipzig.de/compint/mlr/mlr_03_2008.pdf.

[6] M. Cottrell, B. Hammer, A. Hasenfuss, and T. Villmann. Batch and median neural gas. *Neural Networks*, 19:762–771, 2006.

[7] M. Geist, O. Pietquin, and G. Fricout. Kernelizing vector quantization algorithms. In M. Verleysen, editor, *Proceedings of the European Symposium on Artificial Neural Networks (ESANN'2009)*, page in press, Brussels, 2009. Editions D'Facto.

[8] T. Geweniger, D. Zuehlke, B. Hammer, and T. Villmann. Median variant of fuzzy c-means. In M. Verleysen, editor, *Proceedings of the European Symposium on Artificial Neural Networks (ESANN'2009)*, page in press, Brussels, 2009. Editions D'Facto.

[9] A. Gosh, M. Biehl, and B. Hammer. Performance analysis of lvq algorithms: a statistical physics approach. *Neural Networks*, 19:817–829, 2006.

[10] B. Hammer, A. Hasenfuss, F. Rossi, and M. Strickert. Topographic processing of relational data. In H. Ritter, editor, *Proceedings of WSOM'2007)*, pages 1–6, Bielefeld, 2007.

[11] B. Hammer, A. Micheli, N. Neubauer, A. Sperduti, and M. Strickert. Self organizing maps for time series. In M. Cottrell, editor, *Proceedings of WSOM'2005)*, pages 115–122, Paris, 2005.

[12] B. Hammer, F.-M. Schleif, and Th. Villmann. On the generalization ability of prototype-based classifiers with local relevance determination. *Technical Reports University of Clausthal IfI-05-14*, page 18 pp, 2005.

[13] B. Hammer, M. Strickert, and T. Villmann. Supervised neural gas with general similarity measure. *Neural Proc. Letters*, 21(1):21–44, 2005.

[14] B. Hammer, M. Strickert, and Th. Villmann. Relevance lvq vs. svm. In *Artificial-Intelligence and Soft-Computing (ICAISC 2004)*, pages 592–597. Springer LNAI 3070, 2004.

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

[16] A. Hasenfuss and B. Hammer. Relational topographic maps. In *Proceedings of IDA'2007)*, pages 93–105, 2007.

[17] A. Hasenfuss, B. Hammer, and F. Rossi. Patch relational neural gas. In *Proceedings of ANNPR'2008)*, pages 1–12, Paris, 2008.

[18] M. Hasenjäger, H. Ritter, and K. Obermayer. Active data selection for fuzzy topographic mapping of proximities. In G. Brewka, R. Der, S. Gottwald, and A. Schierwagen, editors, *Fuzzy-Neuro-Systems'99 (FNS'99), Proc. Of the FNS-Workshop*, pages 93–103, Leipzig, 1999. Leipziger Universitätsverlag.

[19] T. Hastie. *The Elements of Statistical Learning*. Springer, New York, 2001.

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

[21] N. B. Karayiannis and J. C. Bezdek. An integrated approach to fuzzy learning vector quantization and fuzzy c-means clustering. *IEEE Transactions on Fuzzy Systems*, 5(4):622–8, 1997.

[22] T. Kohonen. The self-organizing map. In *Proceedings of the IEEE*, volume 78, pages 1464–1480, 1990.

[23] Teuvo Kohonen. Learning Vector Quantization. *Neural Networks*, 1(Supplement 1):303, 1988.

[24] Teuvo Kohonen. *Self-Organizing Maps*, volume 30 of *Springer Series in Information Sciences*. Springer, Berlin, Heidelberg, 1995. (2nd Ext. Ed. 1997).

[25] T. Koskela, M. Varsta, J. Heikkonen, and K. Kaski. Recurrent som with local linear models in time series prediction. In *6th European Symposium on Artificial Neural Networks. ESANN'98. Proceedings*, pages 167–72, Brussels, Belgium, 1998. D-Facto.

[26] J. Lee and M. Verleysen. Generalization of the l_p norm for time series and its application to self-organizing maps. In M. Cottrell, editor, *Proc. of Workshop on Self-Organizing Maps (WSOM) 2005*, pages 733–740, Paris, Sorbonne, 2005.

[27] Thomas M. Martinetz, Stanislav G. Berkovich, and Klaus J. Schulten. 'Neural-gas' network for vector quantization and its application to time-series prediction. *IEEE Trans. on Neural Networks*, 4(4):558–569, 1993.

[28] H. Ritter. Self-organizing maps on non-euclidean spaces. In E. Oja and S. Kaski, editors, *Kohonen Maps*, pages 97–110. Elsevier, Amsterdam, 1999.

[29] F. Rossi and N. Villa. Topologically ordered graph clustering via deterministic annealing. In M. Verleysen, editor, *Proceedings of the European Symposium on Artificial Neural Networks (ESANN'2009)*, page in press, Brussels, 2009. Editions D'Facto.

[30] A. Sato and K. Yamada. Generalized learning vector quantization. In D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, editors, *Advances in Neural Information Processing Systems 8. Proceedings of the 1995 Conference*, pages 423–9. MIT Press, Cambridge, MA, USA, 1996.

[31] P. Scheider, K. Bunte, H. Stickemer, B. Hammer, T. Villmann, and M. Biehl. Regularization in matrix relevance learning. *Machine Learning Reports*, 2(MLR-02-2008), 2008. ISSN:1865-3960 http://www.uni-leipzig.de/compint/mlr/mlr_02_2008.pdf.

[32] F.-M. Schleif. *Prototype based Machine Learning for Clinical Proteomics*. PhD thesis, Technical University Clausthal, Technical University Clausthal, Clausthal-Zellerfeld, Germany, 2006.

[33] F.-M. Schleif, T. Elssner, M. Kostrzewa, Th. Villmann, and B. Hammer. Analysis and visualization of proteomic data by fuzzy labeled self organizing maps. In *Proc. of CBMS 2006*, pages 919–924, 2006.

[34] F.-M. Schleif, B. Hammer, and Th. Villmann. Margin based active learning for lvq networks. *Neurocomputing*, 70(7-9):1215–1224, 2007.

[35] F.-M. Schleif, Th. Villmann, and B. Hammer. Fuzzy labeled soft nearest neighbor classification with relevance learning. In M. Arif Wani, Krzysztof J. Cios, and Khalid Hafeez, editors, *In Proceedings of the 4th International Conference on Machine Learning and Applications (ICMLA) 2005*, pages 11–15, Los Alamitos, CA, USA, 2005. IEEE Press.

[36] P. Schneider, F.-M. Schleif, T. Villmann, and M. Biehl. Generalized matrix learning vector quantizer for the analysis of spectral data. In *Proceedings of ESANN'2008)*, pages 451–456, 2007.

[37] S. Z. Selim and M. A. Ismail. K-means type algorithms: a generalized convergence theorem and characterization of local optimality. *IEEE T-PAMI*, 6:81–87, 1984.

[38] S. Seo, M. Bode, and K. Obermayer. Soft nearest prototype classification. *IEEE TNN*, 14(2):390–398, 2003.

[39] S. Seo and K. Obermayer. Soft learning vector quantization. *Neural Computation*, 15:1589–1604, 2003.

[40] Janne Sinkkonen and Samuel Kaski. Clustering based on conditional distributions in an auxiliary space. *Neural Computation*, 14:217–239, 2002.

[41] M. Strickert, T. Bojer, and B. Hammer. Generalized relevance LVQ for time series. In *ARTIFICAL NEURAL NETWORKS-ICANN 2001, PROCEEDINGS*, pages 677–683, 2001.

[42] M. Strickert and B. Hammer. Merge som for temporal data. *Neurocomputing*, 64:39–72, 2005.

[43] M. Strickert, P. Schneider, J. Keilwagen, T. Villmann, M. Biehl, and B. Hammer. Discriminatory data mapping by matrix-based supervised learning metrics. In *Proceedings of ANNPR'2008)*, pages 78–89, Paris, 2008.

[44] M. Strickert, N. Sreenivasulu, W. Weschke, H.-P. Mock, and U. Seiffert. Unsupervised feature selection for biomarker identification in chromatography and gene expression data. In *Proceedings of ANNPR'2006)*, pages 274–285, Ulm, 2006.

[45] P. Tino and I. Nabney. Hierarchical GTM: constructing localized nonlinear projection manifolds in a principled way. *IEEE-Transactions-on-Pattern-Analysis-and-Machine-Intelligence*, 24:639–56, 2002.

[46] K. Torkkola. Mutual information in learning feature transformations. *Journal of Machine Learning Research*, 3:1415–1438, 2003.

[47] M. M. Van Hulle. Towards an information-theoretic approach to kernel-based topographic map formation. In *Advances in Self-Organising Maps*.

[48] V Vapnik. *Statistical Learning Theory*. Wiley, New York, 1998.

[49] V Vapnik. *The Nature of Statistical Learning Theory*. Springer, New York, 2000.

[50] T. Villmann. Sobolev metrics for learning of functional data - mathematical and theoretical aspects. *Machine Learning Reports*, 1(MLR-03-2007), 2007. ISSN:1865-3960 http://www.uni-leipzig.de/compint/mlr/mlr_03_2007.pdf.

[51] T. Villmann and J. C. Claussen. Magnification control in self organizing maps and neural gas. *Neural Computation*, 18(2):446–469, 2006.

[52] T. Villmann, B. Hammer, and M. Biehl. Some theoretical aspects of neural gas vector quantizers. In M. Biehl, B. Hammer, M. Verleysen, and T. Villmann, editors, *Similarity based clustering*, page in press. Springer, Berlin, 2009.

[53] T. Villmann, B. Hammer, F.-M. Schleif, W. Herrmann, and M. Cottrell. Fuzzy classification using information theoretic learning vector quantization. *NeuroComputing*, 71(16-18):3070–3076, 2008.

[54] T. Villmann, F.-M. Schleif, and B. Hammer. Prototype-based fuzzy classification with local relevance for proteomics. *NeuroComputing*, 69:2425–2428, 2006.

[55] T. Villmann, F.-M. Schleif, B. Hammer, and M. Kostrzewa. Exploration of mass-spectrometric data in clinical proteomics using learning vector quantization methods. *Briefings in Bioinformatics*, 9(2):129–143, 2008.

[56] T. Villmann, F.-M. Schleif, E. Merenyi, and B. Hammer. Fuzzy labeled self organizing map for classification of spectra. In *In Proceedings of the 9th International Work-Conference on Artificial Neural Networks (IWANN) 2007*, pages 556–563, Berlin, Heidelberg, Germany, 2007. Springer.

[57] T. Villmann, F.-M. Schleif, M. v.d.Werff, A. Deelder, and R. Tollenaar. Associative learning in soms for fuzzy-classification. In *Proc. of ICMLA 2007*, pages 581–586, 2007.

[58] Th. Villmann, F.-M. Schleif, and B. Hammer. Supervised neural gas and relevance learning in learning vector quantisation. In Takeshi Yamakawa, editor, *In Proceedings of the 4th Workshop on Self Organizing Maps (WSOM) 2003*, pages 47–52, Hibikino, Kitakyushu, Japan, 2003. Kyushu Institute of Technology on CD-ROM (C) 2003 WSOM'03 Organizing Committee.

[59] Th. Villmann, F.-M. Schleif, and B. Hammer. Comparison of relevance learning vector quantization with other metric adaptive classification methods. *Neural Networks*, 19(15):610–622, 2005.

[60] A. Wismueller. A computational framework for exploratory data analysis. In M. Verleysen, editor, *Proceedings of the European Symposium on Artificial Neural Networks (ESANN'2009)*, page in press, Brussels, 2009. Editions D'Facto.

[61] A. Witoelar, M. Biehl, and B. Hammer. Equilibrium properties of off-line lvq. In M. Verleysen, editor, *Proceedings of the European Symposium on Artificial Neural Networks (ESANN'2009)*, page in press, Brussels, 2009. Editions D'Facto.