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.

A Sparse Kernelized Matrix Learning Vector Quantization Model for Human Activity Recognition

M. Kästner[1], M. Strickert[2], and T. Villmann[1]

1- University of Appl. Sciences Mittweida - Dept. of Mathematics
Mittweida, Saxonia - Germany

2- University Marburg - Fac. of Mathematics and Computer Sciences,
Knowledge Engineering & Bioinformatics Group, Germany

Abstract. The contribution describes our application to the ESANN'2013 Competition on Human Activity Recognition (HAR) using Android-OS smartphone sensor signals. We applied a kernel variant of learning vector quantization with metric adaptation using only one prototype vector per class. This sparse model obtains very good accuracies and additionally provides class correlation information. Further, the model allows an optimized class visualization.

1 Introduction

The present paper describes the general techniques applied to solve the ESANN'2013 competition to detect human activity and motion disorders based on Android-OS smartphone sensor signals[1]. For this purpose six different human activities were selected by the organizers to recognize: standing, sitting, laying, walking, walking upstairs and walking downstairs. The respective vectorial information were retrieved from smartphone inertial sensors such as accelerometers and gyroscope. The background of the competition lies in potential applications for assisted living technologies.

Classification of vectorial data is a challenging task for many real life problems. Although the goal is quite obvious, the realization frequently is not simple but requires sophisticated methods for satisfying solutions. One of the most prominent and successful methods for classification is the support vector machine (SVM) [6, 14, 16]. SVMs uses the implicit kernel mapping of the data into a high-dimensional data space such that the classes often become linearly separable in that space. In particular, SVMs optimize the separation margin. The implicit handling of this mapping is called the 'kernel trick'. One disadvantage of this approach is the complexity of the resulting model (numbers of support vectors), which can not be explicitly controlled. Further, the support vectors describe the class border, whereas in several applications (for example in medicine) class typical prototypes are preferred. An alternative to SVMs is the learning vector quantization approach based on Hebbian learning of more or less class typical prototypes [10]. Although heuristically motivated, there exist modifications such that a gradient learning according to a cost function takes place.

[1]The data set with its detailed description is available on the UCI data base http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones

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.

This method is known as Generalized Learning Vector Quantization (GLVQ, [12]), which optimizes implicitly the hypothesis margin [3, 13, 7]. Extensions thereof can deal with different parametrized metrics for optimal classification results [8, 9, 15]. Recent developments also include kernelized variants [9, 19]. In the proposed challenge we applied the *kernelized matrix GLVQ* (kGMLVQ). Thus, a combination of the ideas of SVMs (kernel mapping) with these of GLVQ (sparse models) is obtained. Additionally, we provide detailed simulation results from the given training and test data which offer additional insights to the data and their class correlations.

2 Classification of vectorial data by prototype based kernel models

We assume data vectors $\mathbf{v} \in V \subseteq \mathbb{R}^n$ with class labels $x_\mathbf{v} \in \mathcal{C} = \{1, \ldots, C\}$ for training. Many classification models require prototype vectors $\mathbf{w}_k \in \mathbb{R}^n$ to represent the classes. The prototypes $W = \{\mathbf{w}_k \in \mathbb{R}^n, k = 1 \ldots M\}$ are responsible for the different classes according to their prototype labels $y_{\mathbf{w}_k} \in \mathcal{C}$. For the calculation of the dissimilarities between prototypes and data vector frequently a parametrized positive semi-definite bilinear form

$$d_\Lambda(\mathbf{v}, \mathbf{w}) = (\mathbf{v} - \mathbf{w})^\top \Lambda (\mathbf{v} - \mathbf{w}) = (\Omega(\mathbf{v} - \mathbf{w}))^2 \tag{1}$$

is used with *classification correlation matrix* $\Lambda = \Omega^\top \Omega$ and the *mapping matrix* $\Omega \in \mathbb{R}^{m \times n}$ realizing a linear data mapping [4, 5]. Obviously, the Euclidean metric is obtained for Λ being the identity matrix.

For GMLVQ, the classification loss function is defined as

$$E_{\mathrm{GMLVQ}} = \frac{1}{2} \sum_{\mathbf{v} \in V} f(\mu(\mathbf{v})) \text{ with } \mu(\mathbf{v}) = \frac{d^+(\mathbf{v}) - d^-(\mathbf{v})}{d^+(\mathbf{v}) + d^-(\mathbf{v})} \tag{2}$$

being the classifier function and f frequently chosen the identity. $d^+(\mathbf{v}) = d_\Lambda(\mathbf{v}, \mathbf{w}^+)$ denotes the dissimilarity between the data vector $\mathbf{v}$ and the closest prototype $\mathbf{w}^+$ with the same class label $y_{\mathbf{w}^+} = x_v$, and $d^-(\mathbf{v}) = d_\Lambda(\mathbf{v}, \mathbf{w}^-)$ is the dissimilarity degree for the best matching prototype $\mathbf{w}^-$ with a class label $y_{\mathbf{w}^-}$. Prototype learning takes place as a stochastic gradient descent learning of the cost function E_{GMLVQ} taking into account the derivatives of the dissimilarity measure with respect to the prototype vectors. At the same time, the mapping matrix Ω is optimized also by gradient descent learning.

In SVMs the data are implicitly mapped into a reproducing kernel Hilbert space (RKHS) by a map Φ uniquely corresponding to an universal kernel κ_Φ in a canonical manner [2, 11]. The dissimilarity for images of vectors $\mathbf{v}$ and $\mathbf{w}$ is calculated as

$$d_\mathcal{H}(\Phi(\mathbf{v}), \Phi(\mathbf{w})) = \sqrt{\kappa_\Phi(\mathbf{v}, \mathbf{v}) - 2\kappa_\Phi(\mathbf{v}, \mathbf{w}) + \kappa_\Phi(\mathbf{w}, \mathbf{w})} \tag{3}$$

in the RKHS. Universal kernels imply that the image $\mathcal{I}_{\kappa_\Phi} = span(\Phi(V))$ is a subspace of $\mathcal{H}$ [17]. Generally, the map Φ is *non-linear* and the topological

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.

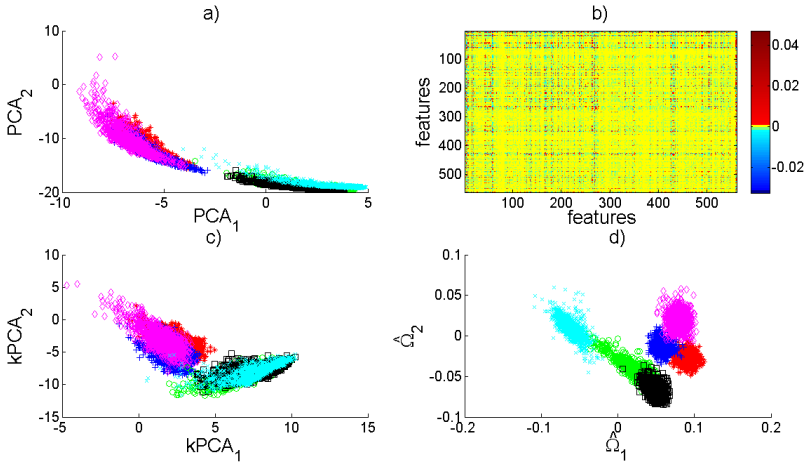

Figure 1: ***a)*** principal component projection of the training data ***b)*** kernel classification correlation matrix $\hat{\Lambda}$ ***c)*** kernel PCA projection of the data according to the eigenspace of the kernel classification correlation matrix $\hat{\Lambda}$ ***d)*** kernel projection of the data according to the limited rank mapping matrix $\hat{\Omega} \in \mathbb{R}^{2\times n}$. (Colored image can be obtained from the authors.)

richness of $\mathcal{I}_{\kappa_\Phi}$ frequently leads to good class separation properties of the mapped data.

Recently, a kernel variant of GLVQ (kGLVQ) was proposed, which uses *differentiable* continuous universal kernels [19, 18]. Hence, the resulting kernel distance (3) is also differentiable and prototype learning can be performed in the original data space but now equipped with the kernel distance (3). We refer to this space as the *kernel data space* V_κ. In contrast to SVMs, kGLVQ obviously allows a control of the model complexity simply by the chosen number of prototypes per class. In this way the advantages of kernel mapping can be combined with the model sparsity provided by GLVQ.

3 Application of the Model to the Competition Data Sets

The training data set consist of 7352 preprocessed vector data of smartphone inertial sensors such as accelerometers and gyroscope for six user motion activities: passive patterns (**g**-sitting, **k**-standing, **c**-laying) and active patterns (**r**-walking, **b**-walking upstairs, **m**-walking downstairs) [1]. The data dimension was $n = 561$. A principal component projection of the training data is given in Fig. 1. We observe a good visual class separation for the two subgroups whereas the overall class separability is not given in this projection.

We applied GMLVQ with *only one prototype per class* to learn the data using the metric (1) width full matrix Ω, i.e. $m = n$. The achieved accuracy for threefold cross validation test was 97.8%. Alternatively, we considered a kGLVQ

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.

	confusion cross validation						confusion test					
	r	**b**	**m**	**g**	**k**	**c**	**r**	**b**	**m**	**g**	**k**	**c**
r	1225	1	0	0	0	0	487	9	0	0	0	0
b	0	1071	2	0	0	0	19	452	0	0	0	0
m	0	5	981	0	0	0	3	12	405	0	0	0
g	0	1	0	1223	42	20	0	1	0	438	52	0
b	0	0	0	59	1315	0	0	0	0	15	517	0
c	0	0	0	0	0	1407	0	0	0	0	0	537

Table 1: Confusion matrices for the cross validation (left) and test data (**r**-walking, **b**-walking upstairs, **m**-walking downstairs, **g**-sitting, **k**-standing, **c**-laying)

with the parametrized RBF-kernel according to

$$k_\Lambda(\mathbf{v},\mathbf{w}) = \exp\left(-(\mathbf{v}-\mathbf{w})^\top \hat{\Lambda}(\mathbf{v}-\mathbf{w})\right) = \exp\left(-\left(\hat{\Omega}(\mathbf{v}-\mathbf{w})\right)^2\right) \quad (4)$$

and refer to this model as kGMLVQ. In this model the kernel width is implicitly adjusted by the *kernel mapping matrix* $\hat{\Omega}$. As for GMLVQ, kGMLVQ allows an automatic adaptation of the matrix $\hat{\Omega}$ by gradient descent learning. Hence, we obtain a kernel classification correlation matrix $\hat{\Lambda}$ providing information about classification correlations in the kernel data space V_κ. The kGMLVQ with full kernel mapping matrix $\hat{\Omega}$ and again only one prototype per class achieves an improved accuracy of 98.2%. The confusion matrix is depicted in Tab. 1 whereas the resulting kernel classification correlation matrix $\hat{\Lambda}$ is visualized in Fig. 1b). Additionally, the data projection according to the eigen decomposition of the kernel classification correlation matrix $\hat{\Lambda}$ is depicted in Fig. 1c).

The test data set for the competition consists of 2947 vectors. Our kGMLVQ approach yields 96.23% test accuracy whereas standard GMLVQ results only 95.83%. The confusion matrix of kGMLVQ is depicted in Tab. 1. We observe that the class probability distribution of the test data is slightly different from that of the training data, which courses the reduced accuracy values.

If we are interested in class visualization only, an explicit data projection is required which can be provided by a *limited rank* kernel mapping matrix $\hat{\Omega} \in \mathbb{R}^{2\times n}$ [4, 5, 9]. Learning the kGLMVQ system with such matrix generates a test accuracy of still 93.5%. Yet, the respective kernel projection of the data delivers well separated classes in the visualization space, see Fig. 1d.

4 Conclusion

In this contribution we present the results of the HAR competition. We applied kernel GMLVQ with only one prototype per class (sparse model), which process the data in a metric space isomorphic to a RKHS. The adaptive RBF-kernel is optimized during learning. At the end, the classification correlation matrix offers additional details regarding class dependencies. The model obtains excellent classification results and allows an optimized class visualization by limited rank kernel mapping.

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.

Acknowledgement

Marika Kästner is supported by a grand from European Science Foundation (ESF), Saxony, Germany.

References

[1] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. Reyes-Ortiz. Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. In *International Workshop of Ambient Assisted Living (IWAAL 2012)*, 2012.

[2] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68:337–404, 1950.

[3] M. Biehl, K. Bunte, F.-M. Schleif, P. Schneider, and T. Villmann. Large margin discriminative visualization by matrix relevance learning. In H. Abbass, D. Essam, and R. Sarker, editors, *Proc. of the International Joint Conference on Neural Networks (IJCNN), Brisbane*, pages 1873–1880, Los Alamitos, 2012. IEEE Computer Society Press.

[4] K. Bunte, B. Hammer, A. Wismüller, and M. Biehl. Adaptive local dissimilarity measures for discriminative dimension reduction of labeled data. *Neurocomputing*, 73:1074–1092, 2010.

[5] K. Bunte, P. Schneider, B. Hammer, F.-M. Schleif, T. Villmann, and M. Biehl. Limited rank matrix learning, discriminative dimension reduction and visualization. *Neural Networks*, 26(1):159–173, 2012.

[6] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines and other kernel-based learning methods.* Cambridge University Press, 2000.

[7] B. Hammer, M. Strickert, and T. Villmann. Relevance LVQ versus SVM. In L. Rutkowski, J. Siekmann, R. Tadeusiewicz, and L. Zadeh, editors, *Artificial Intelligence and Soft Computing (ICAISC 2004)*, Lecture Notes in Artificial Intelligence 3070, pages 592–597. Springer Verlag, Berlin-Heidelberg, 2004.

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

[9] M. Kästner, D. Nebel, M. Riedel, M. Biehl, and T. Villmann. Differentiable kernels in generalized matrix learning vector quantization. In *Proc. of the Internacional Conference of Machine Learning Applications (ICMLA'12)*, page in press. IEEE Computer Society Press, 2012.

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

[11] J. Mercer. Functions of positive and negative type and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society, London, A*, 209:415–446, 1909.

[12] A. Sato and K. Yamada. Generalized learning vector quantization. In D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, editors, *Advances in Neu-*

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.

ral Information Processing Systems 8. Proceedings of the 1995 Conference, pages 423–9. MIT Press, Cambridge, MA, USA, 1996.

[13] A. Sato and K. Yamada. An analysis of convergence in generalized LVQ. In L. Niklasson, M. Bodén, and T. Ziemke, editors, *Proceedings of ICANN98, the 8th International Conference on Artificial Neural Networks*, volume 1, pages 170–176. Springer, London, 1998.

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

[15] P. Schneider, B. Hammer, and M. Biehl. Adaptive relevance matrices in learning vector quantization. *Neural Computation*, 21:3532–3561, 2009.

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

[17] I. Steinwart. On the influence of the kernel on the consistency of support vector machines. *Journal of Machine Learning Research*, 2:67–93, 2001.

[18] T. Villmann and S. Haase. A note on gradient based learning in vector quantization using differentiable kernels for Hilbert and Banach spaces. *Machine Learning Reports*, 6(MLR-02-2012):1–29, 2012. ISSN:1865-3960, http://www.techfak.uni-bielefeld.de/~fschleif/mlr/mlr_02_2012.pdf.

[19] T. Villmann, S. Haase, and M. Kästner. Gradient based learning in vector quantization using differentiable kernels. In P. Estevez, J. Principe, and P. Zegers, editors, *Advances in Self-Organizing Maps: 9th International Workshop WSOM 2012 Santiage de Chile*, volume 198 of *Advances in Intelligent Systems and Computing*, pages 193–204, Berlin, 2012. Springer.