LEARNING THE RELEVANT IMAGE FEATURES WITH MULTIPLE KERNELS

Devis Tuia [1], *Giona Matasci* [1], *Gustavo Camps-Valls* [2], *Mikhail Kanevski* [1]

[1] Institute of Geomatics and Analysis of Risk, University of Lausanne, Switzerland
[2] Imaging & Processing Laboratory (IPL), Universitat de València, Spain

1. ABSTRACT

The increase in resolution of satellite sensors has provided new tools for describing and modelling the surface of the Earth. While hyperspectral sensor provide hundreds of bands giving an unprecedented insight of spectral signature of materials, multispectral sensors allow to capture spatial details at very fine scales. These theoretical advantages also pose some hard problems and new challenges in terms of remote sensing image processing: 1) the high number of redundant bands induce collinearity problems and the well-known overfitting phenomenon and, 2) since images are also spatially redundant, this knowledge must be included in the classifier through careful spatial processing techniques. Nevertheless, the evaluation of the relevance of the extracted (both spectral and spatial, contextual or textural) features is a difficult problem. Building models in such scenarios involve high-dimensional data processing and thus create the need for i) classifiers that are efficient and robust in high dimensional spaces and for ii) feature selection routines [1] capable to select features that are discriminative to solve the problem.

Regarding classifiers, kernel methods and support vector machines (SVM) have been shown to be robust methods capable of handling high dimensional input spaces. SVMs have been successfully applied to a wide range of remote sensing problems dealing with spectral [2], contextual [3] and multi-temporal and multi-source [4] information. Regarding feature selection, several strategies have been discussed in the remote sensing literature: great attention has been given to filters, i.e. feature selection methods working independently to the classification, and more complex strategies interacting with the classifier (*wrappers*) start drawing interest.

In this paper we propose a wrapper integrating feature selection and classification within the framework of multiple kernel learning (MKL, [5, 6]): we want to define an *ad-hoc* kernel defined by a convex combination of base kernels:

$$K(x,x') = \sum_{m=1}^{M} d_m K_m(x,x') \qquad \text{with} \quad d_m \geq 0, \sum_{m=1}^{M} d_m = 1 \tag{1}$$

In order to do that, the weights vector $\mathbf{d}$ must be optimized, simultaneously as the SVM problem using the sum kernel K. This can be achieved by wrapping the classical SVM solver with an optimization routine.

In this paper we apply SimpleMKL, a recently proposed method to find optimal linear combinations of kernels [7], on both hyperspectral (HyMap) and multispectral (QuickBird) images. This model performs a reduced gradient search over the primal of the SVM to optimize $\mathbf{d}$.

For the purpose of feature selection in multidimensional data, we may build as many kernels as we have initial variables. Nonetheless, in the case of remote sensing images and when confronted to datasets carrying more than one hundred bands, such a strategy would imply the creation (and storage in memory) of several hundreds of kernels. This problem precludes the use of the original formulation of SimpleMKL for problems including thousands of training pixels. To overcome this problem, we propose prameter estimation using kernel alignment [8]:

$$A = \frac{\langle K_m, yy^\top \rangle_F}{\sqrt{\langle K_m, K_m \rangle_F \langle yy^\top, yy^\top \rangle_F}} \tag{2}$$

where yy' is the ideal kernel giving the value 1 if two pixels belong to the same class and 0 otherwise. If two kernels are aligned with the labels vector and not aligned with eachother, their combination will be valuable to solve the problem, because

This work has been partly supported by the Swiss National Science Foundation (grant no.100012-113506) and by the Spanish Ministry of Education and Science under project CONSOLIDER/CSD2007-00018.

both kernels contain independent information. In our setting, we select the best candidates for SimpleMKL, by maximizing the alignment of each feature's kernels K_m with the output vector. Then, SimpleMKL selects the best combination to solve the problem.

Experiments are carried for several scenarios:

(a) One kernel per input variable is selected and kernel parameters are estimated using alignment;

(b) Each variable is associated to several kernels with different parameters (as proposed in [7]): this way, multiscale solutions are possible;

(c) Kernels encoding similarity between groups of features are used. These groups are created either by the type of feature (spectral-spatial in the case of QuickBird image) or by their physical meaning (in the case of HyMap image);

(d) Discriminative variables are selected using prior knowledge about their physical meaning and the best combination of kernels is optimized (HyMap image only).

Excellent classification results were observed in both multi and hyperspectral images. Significant improvements to the solution of the standard SVM are observed for all the experiments led. Additionally, the model returns automatically a rank of the most relevant features and hence the most important physical signal characteristics are discovered.

2. REFERENCES

[1] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, *Feature extraction: foundations and applications*, Springer, 2006.

[2] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. on Geosc. Rem. Sens.*, vol. 43, no. 6, pp. 1351– 1362, 2005.

[3] G. Camps-Valls, L. Gómez-Chova, J. Muñoz-Marí, J. Vila-Francés, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 1, pp. 93–97, Jan 2006.

[4] G. Camps-Valls, L. Gómez-Chova, M. Muñoz Marí, J.and Martínez-Ramón, and J. L. Rojo-Álvarez, "Kernel-based framework for multi-temporal and multi-source remote sensing data classification and change detection," *IEEE Trans. on Geosc. Rem. Sens.*, vol. 46, no. 6, pp. 1822–1835, Jun 2008.

[5] G. Lancricket, T. De Bie, N. Cristianini, M. Jordan, and W. Noble, "A statistical framework for genomic data fusion," *Bioinformatics*, vol. 20, pp. 2626–2635, 2004.

[6] A. Villa, M. Fauvel, J. Chanussot, P. Gamba, and J. A. Benediktsson, "Gradient optimization for multiple kernel parameters in support vector machines classification," in *Proc. of Geosc. Rem. Sens. Symp., IGARSS08*, 2008.

[7] A. Rakotomamonjy, F. R. Bach, S. Canu, and Y. Grandvalet, "Simple MKL," *Journal of Machine Learning Research*, vol. 9, pp. 2491–2521, 2008.

[8] N. Cristianini, J. Kandola, A. Elisseeff, and J. Shawe-Taylor, "On kernel target alignment," Tech. Rep. 2001-087, NeuroCOLT, 2001.