ADVANCED ACTIVE SAMPLING FOR REMOTE SENSING IMAGE CLASSIFICATION

Michele Volpi, Devis Tuia, Mikhail Kanevski

Institute of Geomatics and Analysis of Risk, University of Lausanne (Switzerland)
michele.volpi@unil.ch

1. INTRODUCTION

During the last decades, remote sensing image classification has shown its relevancy for modern Earth Observation applications. The increase of both spatial and spectral resolution made standard algorithms difficult to apply, in order to obtain relevant solution with classical methods. Models such as Support Vector Machines (SVM) [1] have shown their reliability in order to find non-linear solutions to classification problems [2, 3] and are now becoming standard tools for accurate image classification.

The bottleneck of the classification of new generation imagery (VHR) remains the definition of an accurate and meaningful training set. The difficulty of manually selecting the set lies on covering intraclass variance of the image as possible by keeping the number of labeled training patterns $\{\mathbf{x}_i, y_i\}_{i=1}^n$ small. The high cost of labeling a sufficient (a priori unknown) number of pixels with terrain campaigns often forces the user to the visual definition of regions of examples. By this approach, the patterns retained are highly redundant (e.g. neighboring pixels) and the explanation of the complete variability is not ensured. By choosing a correct set of pixels, an accurate classification can be obtained both with less complex models, and improved generalization ability (i.e. avoiding overfitting).

Active learning is a machine learning framework that aims at solving the problem of finding a reliable training set starting with few and sub-optimal examples. With this new information, the classifier can create a model that outperforms the previous one, converging to the optimal classification accuracy faster than when selecting the training samples randomly. From the practical point of view, the model returns to the user the patterns selected from a pool of unlabeled pixels (candidates) $\{\mathbf{x}_j\}_{j=1}^m$ for which the classifier is less confident. By adding these samples to the training set, the classifier is forced to solve the regions of uncertainty. When using SVM, a natural approach is to query the user for labels of the candidates nearest to the classification margin. This method is known as Margin Sampling (MS) [4].

Despite the advantages that active learning can provide, they can rarely be found in the remote sensing community. Mitra *et al.* [5] adapted an active learning method similar to MS for object-oriented segmentation. Rajan *et al.* [6] proposed a probabilistic approach based on a bounded loss of the learner, and thus choosing pixels that minimizes this loss. Jun and Ghosh [7] propose a development of this method by introducing an adjustment for distributions of labeled data prior and posterior to the sampling. Liu *et al.* [8] proposed a combination of the active learning approach and semi-supervised classifier, based on a mutual information measure in order to detect targets. Recently, Tuia *et al.* [9] proposed a new approach based on classification disagreement by a committee of learners. These preliminary studies raised several questions and opened research directions for active sampling strategies better considering class overlapping and high redundancy. This is particular true when considering VHR image classification [9]: when sampling several candidates at each iteration, the MS heuristic can return highly redundant points without exhaustively exploring the whole pool of candidates. Here, the problem of exploring the whole margin and enforcing diversity in the sampling by querying for more appropriate points is addressed by clustering the feature space induced by the SVM kernel to weight the MS heuristic by a measure of the distribution of the points in such space. This new advanced MS-based method is proven to outperform classical MS and other active learning methods by overcoming the problem of redundancy. In order to cluster the points in the same space where SVM shatter data, the kernel k-means algorithm is proposed [10, 11].

This work is supported by the Swiss National Science Foundation project N. 200021-126505 and PBLAP2-127713.

2. EXPLORING THE SVM MARGIN

The solution of a SVM model for classification can be summarized as the maximization of a margin that separates patterns of different classes under some constraints. The solution is then given by the decision function of form:

$$f(\mathbf{x}) = \sum_i \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}) + b. \tag{1}$$

Where α_i are the coefficients ($\alpha_i \neq 0$ for support vectors), y_i the associated labels and $k(\mathbf{x}_i, \mathbf{x})$ the kernel between points $\mathbf{x}$ and $\mathbf{x}_i$. The points lying within the hyperplanes (margins) have a decision function value $0 < |f(\mathbf{x})| < 1$. These patterns can carry important information which can be used to update the model for better classification boundary. Adopting the MS heuristic, the points $\mathbf{x}$ showing minimal distance to the hyperplane are added to the set and are removed from the pool. By using this MS approach, despite its versatility, the added information can be sub-optimal in terms of redundancy of information or of ignored areas. In order to solve these problems, the proposed approach aims at finding points that both minimize the distance to the margin (filling the uncertainty of the classifier in this region of the feature space) and maximize the distance from the clusters centers, i.e. avoiding redundancy of a partition of the margin. The latter can be view as applying some declustering algorithm in the feature space. Thus, prior to choose points from the candidate pool, all the points lying within the margin are clustered in the same space of the SVM (by using the same kernel function and hyperparameters). In this new active learning heuristic, a point per cluster is added by evaluating the score given by including both the real valued SVM function $f(\mathbf{x})$ and the distance to the center of the cluster in the reproducing kernel Hilbert space.

Kernel k-means algorithm is defined as the iterative minimizer of the distance from patterns $\phi(\mathbf{x})$ to the cluster centers $\mathbf{m}_k$ in the feature space, with $k = 1, ..., K$ (cluster k) and C_k the elements of the cluster:

$$d(\phi(\mathbf{x}), \mathbf{m}_k) = ||\phi(\mathbf{x}) - \mathbf{m}_k||^2 = \sum_{k=1}^{K} \sum_{i,j=1}^{|C_k|} (\phi(\mathbf{x}) - \mathbf{m}_k)(\phi(\mathbf{x}) - \mathbf{m}_k)^\top = k(\mathbf{x}, \mathbf{x}) - \frac{2}{|C_k|} k(\mathbf{x}, \mathbf{x}_i) + \frac{1}{|C_k|^2} k(\mathbf{x}_i, \mathbf{x}_j). \tag{2}$$

The point to be selected is the one that minimizes the score $s(\mathbf{x}_i) = (1 - \lambda) f(\mathbf{x}_i) + \lambda\, d(\phi(\mathbf{x}_i), \mathbf{m}_k)$, where λ is a user defined trade off parameter $0 \leq \lambda \leq 1$.

3. EXPERIMENTAL RESULTS

The proposed heuristic has been tested and validated on a QuickBird pansharpened image of the city of Zurich. The experiments were performed by comparing the proposed method with random sampling (RS) and with margin sampling (MS). Learning curves (Figure 1) illustrate the performance of the proposed method for different values of λ.

The set up of the experiments was the following. For each experiment, 10 runs were carried out starting from different sub-optimal training sets composed by 10 points per class (for a total of 5 classes). In Figure 1, two sets of curves are presented: on the left side, experiments by varying the value of λ are shown, while on the right the best KKM-SVM is compared with RS and with MS. Curves are the means of 10 runs. Experiments are truncated at 850 training points for visualization purposes.

The proposed approach is faster in terms of convergence and slightly better in terms of accuracy. Full SVM was trained with 3000 points, showing an accuracy of 85.41% ($\mathcal{K} = 0.803$). The proposed heuristic performs at the same accuracy with a training set size of 700 pixels. MS and RS are both outperformed by the proposed Kernel k-Means SVM (KKM-SVM) with an appropriate λ parameter. The curves obtained for different $\lambda \in \{0; 0.125; 0.25; 0.375; 0.5; 0.625\}$ show different behaviors but a similar trend: by approaching the active learning problem weighting the functional with cluster membership, the convergence is more uniform comparing to MS. The λ showing a better learning rate is $\lambda = 0.25$.

4. CONCLUSIONS

In this paper we proposed a new approach to active learning with SVM, useful when dealing with redundant information. The proposed sampling exploits the natural clustering of data in the feature space and maximizes the informativeness of the queried patterns. It is worth noting that the worst KKM-SVM run is comparable to the MS heuristic and the lower bound on the accuracy given by the RS is respected on the average. In terms of required number of pixels the convergence of this new methodology is faster than for other tested margin-based active learning approaches.

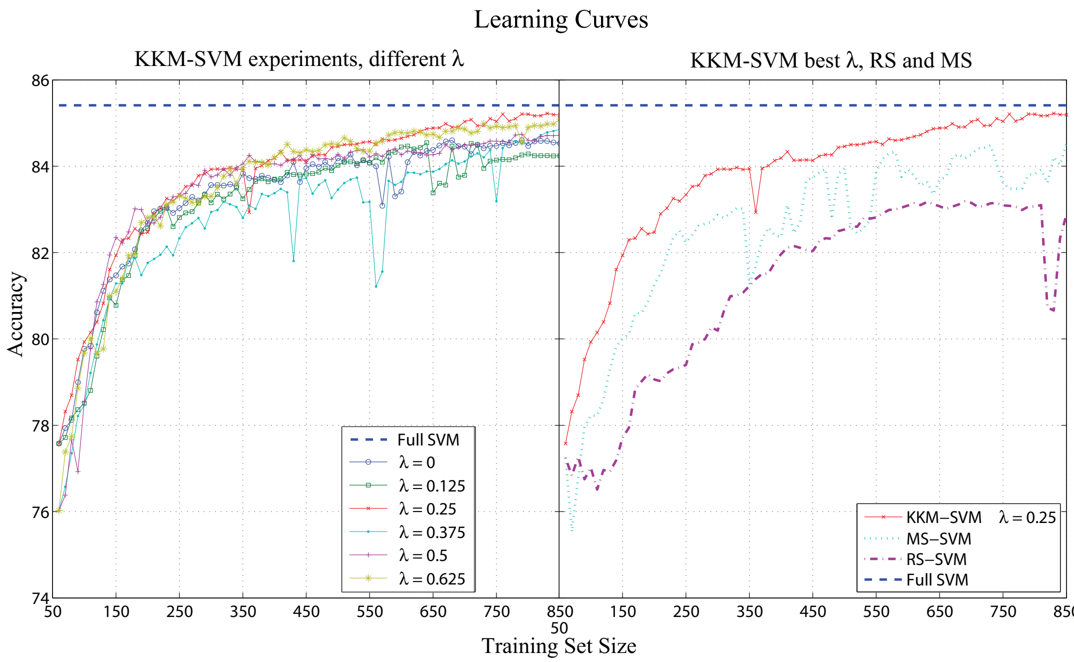

Fig. 1. *Learning Curves as a function of the training set size. At each iteration 10 points were added, starting from a sub-optimal training subset composed by 10 points per class (a total of 5 classes).*

5. REFERENCES

[1] B. Boser, I. Guyon, and V. Vapnik, "A training algorithm for optimal margin classifiers," in *5th ACM Workshop on Comput. Learn. Theory*, Pittsburgh, USA, 1992, pp. 144–152.

[2] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, 2004.

[3] G. Camps-Valls and L. Bruzzone, *Kernel methods in Remote Sensing Image Processing*, 2009.

[4] G. Schohn and D. Cohn, "Less is more: active learning with support vectors machines," in *Proc. ICML*, Stanford, USA, 2000, pp. 839–846, Morgan Kaufmann.

[5] P. Mitra, B.U. Shankar, and S.K. Pal, "Segmentation of multispectral remote sensing images using active support vector machines," *Pattern Recogn. Lett.*, vol. 25, no. 9, pp. 1067–1074, 2004.

[6] S. Rajan, J. Ghosh, and M.M. Crawford, "An active learning approach to hyperspectral data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46(4), pp. 1231–1242, 2008.

[7] G. Jun and J. Ghosh, "An efficient active learning algorithm with knowledge transfer for hyperspectral data analysis," in *Proc. IEEE IGARSS*, 2008.

[8] Q. Liu, X. Liao, and L. Carin, "Detection of unexploded ordnance via efficient semisupervised and active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 9, pp. 2558–2567, 2008.

[9] D. Tuia, F. Ratle, F. Pacifici, M. Kanevski, and W.J. Emery, "Active learning methods for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 47(7), no. 7, pp. 2218–2232, 2009.

[10] R. Zhang and A.I. Rudnicky, "A large clustering scheme for kernel k-means," in *Proc. IEEE ICPR*, Quebec City, Canada, 2002, vol. 4, pp. 289–292.

[11] I.S. Dhillon, Y. Guan, and B. Kulis, "Kernel k-means, spectral clustering and normal cuts," in *Int. Conf. Knowl. Disc. Data Mining KDD*, Seattle, USA, 2004, pp. 551–556.