SEMI-SUPERVISED ACTIVE LEARNING FOR URBAN HYPERSPECTRAL IMAGE CLASSIFICATION

Inmaculada Dópido[1], *Jun Li*[1,2], *Antonio Plaza*[1] *and José M. Bioucas-Dias*[2]

[1]Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, University of Extremadura, E-10071 Caceres, Spain .
[2]Instituto de Telecomunicações, Instituto Superior Técnico, TULisbon, 1900-118, Lisbon, Portugal.

ABSTRACT

In this paper, we develop a new framework for semi-supervised learning which exploits active learning for unlabeled sample selection in hyperspectral data classification. Specifically, we use active learning to select the most informative unlabeled training samples with the ultimate goal of systematically achieving noticeable improvements in classification results with regard to those found by randomly selected training sets of the same size. Our experimental results, conducted with an urban hyperspectral scene collected by the Reflective Optics Spectrographic Imaging Instrument (ROSIS) of the Deutschen Zentrum for Luftund Raumfahrt (DLR, the German Aerospace Agency) over the city of Pavia, Italy, indicate that using active learning for unlabeled sample selection represents an effective and promising strategy in the context of urban hyperspectral data classification.

***Index Terms*—** Hyperspectral image classification, urban classification, semi-supervised learning, active learning.

1. INTRODUCTION

Remotely sensed hyperspectral imaging allows for the detailed analysis of the surface of the Earth using advanced imaging instruments which can produce high-dimensional images with hundreds of spectral bands [1]. Supervised hyperspectral image classification is a difficult task due to the unbalance between the high dimensionality of the data and the limited availability of labeled training samples in real analysis scenarios [2]. This is particularly the case in urban areas, which are dominated by complex regions and surface heterogeneity which often prevents the collection of reliable ground-truth samples [3]. While the collection of labeled samples is generally difficult, expensive and time-consuming, unlabeled samples can be generated in a much easier way. This observation has fostered the idea of adopting semi-supervised learning (SSL) techniques in hyperspectral image classification. The main assumption of such techniques is that the new (unlabeled) training samples can be obtained from a (limited) set of available labeled samples without significant effort/cost [4].

In this paper, we develop a new framework for SSL which exploits active learning (AL) for unlabeled sample selection. Specifically, we use AL to select the most informative unlabeled training samples with the ultimate goal of systematically achieving noticeable improvements in classification results with regard to those found by randomly selected training sets of the same size. It should be noted that, in our context, using AL for unlabeled sample selection is similar to using AL for labeled sample selection in supervised algorithms. However, in SSL the labels of the selected pixels are estimated by the classifier itself with the advantage that no extra cost is required for labeling the selected pixels when compared with the supervised AL. Our experimental results, conducted with an urban hyperspectral scene collected by the Reflective Optics Spectrographic Imaging Instrument (ROSIS) of the Deutschen Zentrum for Luftund Raumfahrt (DLR, the German Aerospace Agency) over the city of Pavia, Italy, indicate that AL for unlabeled sample selection represents an effective and promising strategy in the context of urban hyperspectral data classification.

2. PROPOSED APPROACH

Let $\mathcal{K} \equiv \{1, \ldots, K\}$ denote a set of K class labels, $\mathcal{S} \equiv \{1, \ldots, n\}$ a set of integers indexing the n pixels of an image, $\mathbf{x} \equiv (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathbb{R}^{d \times n}$ an image of d-dimensional feature vectors, $\mathbf{y} \equiv (y_1, \ldots, y_n)$ an image of labels, $\mathcal{D}_l \equiv \{(y_{l_1}, \mathbf{x}_{l_1}), \ldots, (y_{l_n}, \mathbf{x}_{l_n})\}$ a set of labeled samples, l_n the number of labeled training samples, $\mathcal{Y}_l \equiv \{y_{l_1}, \ldots, y_{l_n}\}$ the set of labels in $\mathcal{D}_l$, $\mathcal{X}_l \equiv \{\mathbf{x}_{l_1}, \ldots, \mathbf{x}_{l_n}\}$ the set of feature vectors in $\mathcal{D}_l$, $\mathcal{D}_u \equiv \{\mathcal{X}_u, \mathcal{Y}_u\}$ a set of unlabeled samples, $\mathcal{X}_u \equiv \{\mathbf{x}_{u_1}, \ldots, \mathbf{x}_{u_n}\}$ the set of unlabeled feature vectors in $\mathcal{D}_u$, $\mathcal{Y}_u \equiv \{y_{u_1}, \ldots, y_{u_n}\}$ the set of labels associated with

This work has been supported by the European Community's Marie Curie Research Training Networks Programme under contract MRTN-CT-2006-035927, Hyperspectral Imaging Network (HYPER-I-NET). Funding from the Portuguese Science and Technology Foundation, project PEst-OE/EEI/LA0008/2011, and from the Spanish Ministry of Science and Innovation (CEOS-SPAIN project, reference AYA2011-29334-C02-02) is also gratefully acknowledged.

$\mathcal{X}_u$, and u_n the number of unlabeled samples. With this notation in mind, the proposed strategy consists of two main ingredients: semi-supervised learning (SSL) and active learning (AL).

2.1. Semi-supervised Learning (SSL)

For the SSL part, we use multinomial logistic regression (MLR) to model the class posterior density, which is formally given by [5]:

$$p(y_i = k|\mathbf{x}_i, \boldsymbol{\omega}) = \frac{\exp(\boldsymbol{\omega}^{(k)^T}\mathbf{h}(\mathbf{x}_i))}{\sum_{k=1}^{K}\exp(\boldsymbol{\omega}^{(k)^T}\mathbf{h}(\mathbf{x}_i))}, \quad (1)$$

where $\mathbf{h}(\mathbf{x}) = [h_1(\mathbf{x}), ..., h_l(\mathbf{x})]^T$ is a vector of l fixed functions of the input, often termed features; $\boldsymbol{\omega}$ are the regressors and $\boldsymbol{\omega} = [\boldsymbol{\omega}^{(1)^T}, ..., \boldsymbol{\omega}^{(K)^T}]^T$. Notice that, the function $\mathbf{h}$ may be linear, *i.e.*, $\mathbf{h}(\mathbf{x}_i) = [1, x_{i,1}, ..., x_{i,d}]^T$, where $x_{i,j}$ is the j-th component of $\mathbf{x}_i$; or nonlinear, *i.e.*, $\mathbf{h}(\mathbf{x}_i) = [1, K_{\mathbf{x}_i,\mathbf{x}_1}, ..., K_{\mathbf{x}_i,\mathbf{x}_l}]^T$, where $K_{\mathbf{x}_i,\mathbf{x}_j} = K(\mathbf{x}_i, \mathbf{x}_j)$ and $K(\cdot,\cdot)$ is some symmetric kernel function. Kernels have been largely used because they tend to improve the data separability in the transformed space. In this paper, we use a Gaussian Radial Basis Function (RBF) $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma^2)$ kernel, which is widely used in hyperspectral image classification [6]. From now on, d denotes the dimension of $\mathbf{h}(\mathbf{x})$. Under the present setup, learning the class densities amounts to estimating the logistic regressors. Following the work in [7,8], we can compute $\boldsymbol{\omega}$ by computing the maximum a posteriori (MAP) estimate:

$$\widehat{\boldsymbol{\omega}} = \arg\max_{\boldsymbol{\omega}} \quad \ell(\boldsymbol{\omega}) + \log p(\boldsymbol{\omega}), \quad (2)$$

where $p(\boldsymbol{\omega}) \propto \exp(-\lambda\|\boldsymbol{\omega}\|_1)$ is a Laplacian prior to promote the sparsity and λ is a regularization parameter controlling the degree of sparseness of $\widehat{\boldsymbol{\omega}}$ in [7,8]; $\ell(\boldsymbol{\omega})$ is the log-likelihood function over the training samples $\mathcal{D}_{l+u} \equiv \mathcal{D}_l + \mathcal{D}_u$, given by:

$$\ell(\boldsymbol{\omega}) \equiv \sum_{i=1}^{l_n+u_n} \log p(y_i = k|\mathbf{x}_i, \boldsymbol{\omega}) \quad (3)$$

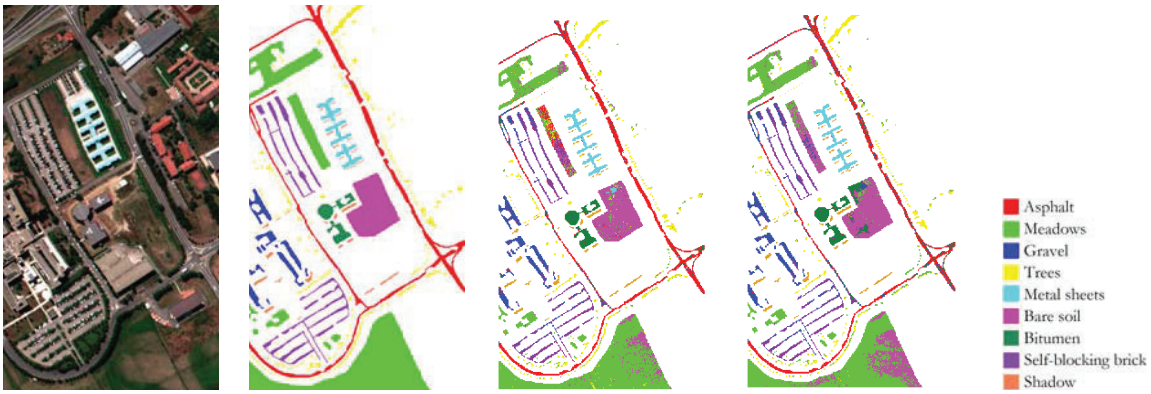
As shown by Eq. (3), labeled and unlabeled information is integrated to learn the regressors $\boldsymbol{\omega}$. The considered SSL approach belongs to the family of self-learning approaches, where the training set $\mathcal{D}_{l+u}$ is incremented under the following criterion. Let $\mathcal{D}_{\mathcal{N}(i)} \equiv \{(\widehat{y}_{i_1}, \mathbf{x}_{i_1}), \ldots, (\widehat{y}_{i_n}, \mathbf{x}_{i_n})\}$ be the set of neighboring set of samples of $(y_i, \mathbf{x}_i)$ for $i \in \{l_1, \ldots, l_n, u_1, \ldots, u_n\}$, where i_n is the number of samples in $\mathcal{D}_{\mathcal{N}(i)}$ and $\widehat{y}_{i_j}$ is the maximum a posteriori probability (MAP) estimate from the MLR classifier, with $i_j \in \{i_1, \ldots, i_n\}$. If $\widehat{y}_{i_j} = y_i$, we increment the unlabeled training set by adding $(\widehat{y}_{i_j}, \mathbf{x}_{i_j})$, *i.e.*, $\mathcal{D}_u = \{\mathcal{D}_u, (\widehat{y}_{i_j}, \mathbf{x}_{i_j})\}$. This increment is reasonable due to the following considerations. First, from a global viewpoint, samples which have the same spectral structure likely belong to the same class. Second, from a local viewpoint, it is very likely that two neighboring pixels also belong to the same class. Therefore, the newly included samples are reliable for learning the classifier. In this work, we run an iterative scheme to increment the training set as this strategy can refine the estimates and enlarge the neighborhood set such that the set of potential unlabeled training samples is increased.

2.2. Active Learning (AL)

In this work, we adopt the AL concept from supervised learning [9–11] and combine it with SSL. In this way, we can find the most informative samples without the need for human supervision. In this case, the labels are predicted by the considered SSL algorithm as mentioned in section 2.1. Let $\mathcal{D}_c$ be the newly generated unlabeled training set at each iteration, which meets the criteria of the considered SSL algorithm. Now we can run AL algorithms over $\mathcal{D}_c$ to find the most informative set $\mathcal{D}_u$, such that $\mathcal{D}_u \subseteq \mathcal{D}_c$. It should be noted that we use $\mathcal{D}_c$ as the candidate set for the AL process instead of the whole image. This is because, as compared with the user-oriented strategy in supervised learning in which the labels are given by the end-users, here we use machine-machine interaction so that the new labels are predicted by the learning algorithm itself. Therefore, in order to have a good control of the newly generated samples, high-confidence estimates are preferred. Furthermore, due to the fact that we use a discriminative classifier and a self-learning strategy for the SSL algorithm, AL algorithms which focus on the boundaries are preferred. In our study, we use the well-known breaking ties (BT) sampling approach [10] to illustrate our newly proposed framework.

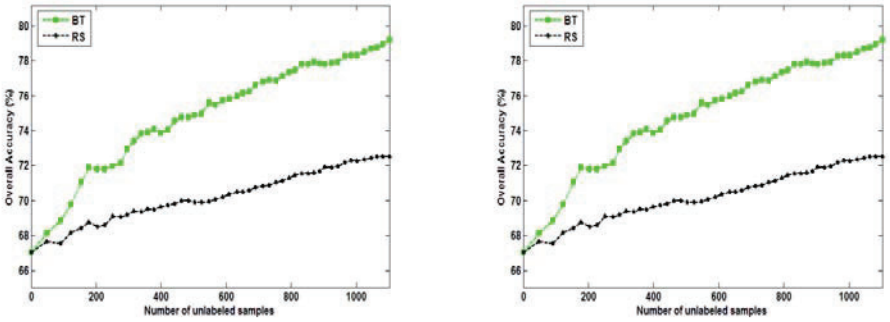
3. EXPERIMENTAL RESULTS

The hyperspectral data set used in our study is an urban data set collected by the ROSIS sensor over the urban area of the University of Pavia, Italy. The flight was operated by DLR in the framework of the HySens project, managed and sponsored by the European Union. The image size in pixels is 610×340, with very high spatial resolution of 1.3 meters per pixel. The number of data channels in the acquired image is 103 (with spectral range from 0.43 to 0.86 μm). Fig. 1(a) shows a false color composite of the image, while Fig. 1(b) shows nine ground-truth classes of interest, which comprise urban features, as well as soil and vegetation features, with 43923 samples. Fig. 2 reports the overall accuracies (OAs) obtained in the classification of the ROSIS hyperspectral image using 10 and 20 labeled training samples per class. First and foremost, these plots shows clear advantages of using unlabeled samples for the SSL algorithm in comparison with the supervised algorithm. Furthermore, it can be observed



(a) Original image (b) Ground truth (c) BT [OA=79.20%] (d) RS [OA=72.55%]

Fig. 1. Classification results obtained by the proposed semi-supervised approach with breaking ties (BT) and random sampling (RS) for active learning, using the ROSIS Pavia University scene.



$l = 90$ $l = 180$

Fig. 2. Overall accuracy (OA) results as a function of the number of unlabeled samples for the proposed semi-supervised approach with active learning, using the ROSIS Pavia University scene.

that the BT sampling approach for unlabeled training sample selection greatly improved the accuracies in comparison with random selection (RS). For illustrative purposes, Fig. 1(c-d) shows the obtained classification maps obtained after applying BT and RS, respectively, for the AL part of our implementation. Overall, our results in this paper indicate that our proposed semi-supervised strategy can greatly assist in improving the classification of urban areas using hyperspectral imagery.

4. REFERENCES

[1] A. F. H. Goetz, G. Vane, J. E. Solomon, and B. N. Rock, "Imaging spectrometry for Earth remote sensing," *Science*, vol. 228, pp. 1147–1153, 1985.

[2] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*, John Wiley & Sons: New York, 2003.

[3] L. Bruzzone, M. Chi, and M. Marconcini, "Classification of remote sensing images from urban areas using a fuzzy probabilistic model," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, pp. 40–44, 2006.

[4] L. Bruzzone, M. Chi, and M. Marconcini, "A novel transductive svm for the semisupervised classification of remote-sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 11, pp. 3363–3373, 2006.

[5] D. Böhning, "Multinomial logistic regression algorithm," *Annals of the Institute of Statistical Mathematics*, vol. 44, pp. 197–200, 1992.

[6] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, pp. 1351–1362, 2005.

[7] B. Krishnapuram, L. Carin, M. Figueiredo, and A. Hartemink, "Sparse multinomial logistic regression: Fast algorithms and generalization bounds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 957–968, 2005.

[8] J. Li, J. Bioucas-Dias, and Antonio Plaza, "Hyperspectral image segmentation using a new bayesian approach with active learning," *IEEE Transactions on Geoscience and Remote Sensing (accepted)*, 2010.

[9] Suju Rajan, Joydeep Ghosh, and Melba M. Crawford, "An active learning approach to hyperspectral data classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, pp. 1231–1242, 2008.

[10] D. Tuia, F. Ratle, F. Pacifici, M. F. Kanevski, and W. J. Emery, "Active learning methods for remote sensing image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 7, pp. 2218–2232, 2009.

[11] J. Li, J. Bioucas-Dias, and A. Plaza, "Semi-supervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, pp. 4085–4098, 2010.