

INTEGRATION OF HYPERSPECTRAL IMAGE CLASSIFICATION AND UNMIXING FOR ACTIVE LEARNING

Jun Li^{1,2}, Antonio Plaza² and José M. Bioucas-Dias¹

¹Instituto de Telecomunicações, Instituto Superior Técnico, TULisbon, 1900-118, Lisboa, Portugal.

²Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, University of Extremadura, E-10071 Caceres, Spain.

ABSTRACT

Spectral unmixing is a growing area in remotely sensed hyperspectral image analysis. Many algorithms have been developed to retrieve pure spectral components and determine their sub-pixel abundance fractions in this kind of data. However, possible connections between spectral unmixing concepts and classification algorithms have been rarely investigated. In this work, we propose a new method to perform semi-supervised hyperspectral image classification exploiting the information retrieved with spectral unmixing. Our main contribution is the integration of a well-established discriminative classifier (the multinomial logistic regression) with a state-of-the-art technique for linear spectral unmixing (fully constrained least squares abundance estimation). Furthermore, we propose a new active sampling approach which integrates both the spatial and the spectral information. Our experimental results, conducted with a well-known hyperspectral image data set collected by the Airborne Visible Infra-Red Imaging Spectrometer (AVIRIS) over the Indian Pines region, NW Indiana, reveal that the proposed method can benefit from the newly developed integrated framework.

Index Terms— Semi-supervised learning, hyperspectral unmixing, unlabeled samples, classification

1. INTRODUCTION

Remotely sensed hyperspectral imaging [1] 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. 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]. Based on these observations, active learning, in which informative unlabeled samples are included in the analysis process by means of assigning labels to the newly selected query, is becoming a fast growing research topic since it can effectively reduce the need for labeled training samples available a priori [3].

Several techniques have been proposed in the recent literature in order to perform active query sampling based on different criteria, such as maximum entropy of the newly added samples, least confidence, or maximum disagreement between multiple learners [4–7]. These approaches lead to high performance in hyperspectral image classification and segmentation problems [8]. However, in general, there are many samples meeting the active sampling criteria, which brings difficulties from the viewpoint of devising mechanisms to effective selection of the most informative samples for active learning in practical applications.

In this work, we propose a new approach to select the most informative samples for active learning by exploiting the linear mixed model commonly used in spectral unmixing application [9]. Spectral unmixing aims at decomposing mixed pixels collected by remotely sensed hyperspectral imaging instruments into a set of pure spectral components (called *end-members*) and their fractional abundances at a sub-pixel level [10]. We exploit this concept because it is well-known that the pixels measured by remotely sensed imaging instruments for Earth observation are composed by different pure spectral substances that interact at sub-pixel levels. This means that most of the pixels in hyperspectral images are mixed in nature. A recent trend to address this issue (in the framework of supervised classification) consists of integrating classification and unmixing concepts to refine the initial classification [11]. This approach seems natural since many research efforts have been developed in both directions, but the synergies between the two approaches have not been widely explored as of yet.

In this work, we have developed a new two-step procedure to find informative samples for supervised classification of hyperspectral data by taking advantage of spectral unmixing concepts. First, we run active sampling to find a set of candidate samples. Then, unmixing is applied to analyze the obtained candidate set and derive which samples can better represent the classes. Here, we model the class posterior densities using a well-established discriminative classifier: the multinomial logistic regression (MLR) [12]. We have selected this classifier due to several reasons. First, it can learn the class posterior probabilities directly, thus effectively dealing with the high dimensionality of hyperspectral data. Second, it produces sparse solutions by adopting a Laplacian prior [13]. This is beneficial in order to control generalization

This work was supported by MRTN-CT-2006-035927 and AYA2008-05965-C04-02 projects.

properties. Finally, the considered classifier can be effectively combined with the proposed active sampling procedure based on spectral unmixing concepts.

The remainder of the paper is organized as follows. Section 2 introduces the proposed methodology and active sampling approach. Section 3 reports classification results based on the well-known hyperspectral image data set collected by the Airborne Visible Infra-Red Imaging Spectrometer (AVIRIS) [14] over the Indian Pines region, NW Indiana. Finally, section 4 concludes with some remarks and future lines.

2. PROPOSED APPROACH

First of all, we briefly define the notations used in this paper. Let $\mathcal{K} \equiv \{1, \dots, K\}$ denote a set of K class labels; let $\mathcal{S} \equiv \{1, \dots, n\}$ be a set of integers indexing the n pixels of an image, let $\mathbf{x} \equiv (\mathbf{x}_1, \dots, \mathbf{x}_n)$ be an image of d -dimensional feature vectors, let $\mathbf{y} \equiv (y_1, \dots, y_n)$ be an image of labels, let L be the number of labeled training samples. With these definitions in mind, in this section we first introduce the MLR model. Then, we present the newly proposed active sampling approach based on linear spectral unmixing concepts.

2.1. Multinomial logistic regression

The original multinomial logistic regression (MLR) [12] models the posterior class probabilities as follows:

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

where $\mathbf{h}(\mathbf{x}_i) \equiv [h_1(\mathbf{x}_i), \dots, h_l(\mathbf{x}_i)]^T$ is a vector of l fixed functions of the input, often termed features; $\boldsymbol{\omega}$ denotes the regressors and $\boldsymbol{\omega} \equiv [\boldsymbol{\omega}^{(1)T}, \dots, \boldsymbol{\omega}^{(K-1)T}]^T$. Since the density (1) does not depend on possible translations applied to the regressors $\boldsymbol{\omega}^{(k)}$, in this work we take $\boldsymbol{\omega}^{(K)} = \mathbf{0}$. It should be noted that the function \mathbf{h} may be linear (*i.e.*, $\mathbf{h}(\mathbf{x}_i) = [1, x_{i,1}, \dots, x_{i,d}]^T$, where $x_{i,j}$ is the j -th component of \mathbf{x}_i) or nonlinear. A kernel function is some symmetric function which offers a mechanism to deal with the nonlinear case, *i.e.*, $\mathbf{h}(\mathbf{x}_i) = [1, K_{\mathbf{x}_i, \mathbf{x}_1}, \dots, K_{\mathbf{x}_i, \mathbf{x}_l}]^T$, where $K_{\mathbf{x}_i, \mathbf{x}_j} = K(\mathbf{x}_i, \mathbf{x}_j)$ and $K(\cdot, \cdot)$. Kernels have been largely used in this context since they tend to improve data separability in the transformed space. In this work, we use the Gaussian Radial Basis Function (RBF) kernel: $K(\mathbf{x}, \mathbf{z}) \equiv -\exp(-\|\mathbf{x} - \mathbf{z}\|^2 / (2\sigma^2))$ kernel, which is widely used in hyperspectral image classification [15].

2.2. Linear spectral unmixing

Spectral unmixing [16] decomposes a (mixed) pixel observation into a collection of pure endmember spectra and their sub-pixel fractional abundances. Under the linear mixture assumption, for any $i \in \mathcal{S}$ we have:

$$\mathbf{x}_i = \mathbf{M}\mathbf{s}_i + \mathbf{n}_i, \quad (2)$$

where $\mathbf{M} \equiv [\mathbf{m}^{(1)}, \dots, \mathbf{m}^{(K)}]$ denotes a mixing matrix composed by the spectral endmembers, \mathbf{n}_i denotes the noise, and $\mathbf{s}_i = [s_i^{(1)}, \dots, s_i^{(K)}]^T$ denotes the fractional abundances of the endmembers at pixel $i \in \mathcal{S}$. We are assuming that each endmember corresponds to a class. Generally, nonnegative and sum-to-one constraints are imposed into this model [17]: $s_i^{(k)} \geq 0$, for $k \in \mathcal{K}$, and $\sum_{k=1}^K s_i^{(k)} = 1$. Let $\gamma_i = \|\mathbf{s}_i\|_\infty \equiv \max\{s_i^{(1)}, \dots, s_i^{(K)}\}$ denote the infinity norm. We note that that, $1/K \leq \gamma_i \leq 1$, $\gamma_i = 1/K$ iff $s_i^{(k)} = 1/K$ for $k \in \mathcal{K}$ (*i.e.*, the spectrum of the pixel is maximally mixed), and $\gamma_i = 1$, \mathbf{x}_i iff $s_i^{(k)} = 1$ for some $k \in \mathcal{K}$ (*i.e.*, the spectrum of the pixel is pure).

2.3. Active sampling

Active sampling, which aims at finding the most informative training set, has been widely studied in the recent literature. However, there are generally many candidate samples meeting the sampling criterion and this introduces difficulties from the viewpoint of how to select the most informative samples adequately. For instance, the breaking ties (BT) active sampling approach [5] minimizes the distance between the first two most probable classes. If the data is dominated by two classes only, it is very likely to have many samples in one single boundary. So when we want to label a limited number of new samples, a relevant question is which samples should be selected first for the learning process.

In this work, we address this issue through the proposal of a new active sampling approach which takes into account some special characteristic of hyperspectral data, such as the presence of mixed pixels. Specifically, we run a two-step procedure for active sampling: 1) first, we run a standard active sampling approach to find a set of candidate samples; 2) then we use a spectral unmixing algorithm to compute the abundances of the obtained samples, with the ultimate goal of identifying a limited number of new samples. The first step aims at finding samples close to the class boundaries. This is because those samples are generally more informative than those samples in the class centers. The second step aims at finding samples away from the boundaries. This is because we also want to include samples which can well represent the related classes. Such combination of intelligently selected samples provides good balance according to our experiments.

The basic strategy of our proposed active sampling approach can be simply summarized as follows. Let \mathcal{D}_U be the set of new sampling meeting the active sampling criterion, and U is the number of samples in \mathcal{D}_U . Let \mathcal{D}_u be the set we select in the active selection procedure and u be the number of samples in \mathcal{D}_u , and generally, $\mathcal{D}_u \subseteq \mathcal{D}_U$ and $u \leq U$. Therefore, a relevant question is ‘‘how to select these u samples?’’. In this work, we use a recently developed hyperspectral unmixing algorithm, termed simplex identification via split augmented Lagrangian (SISAL) [18], to compute the abundances of the samples in \mathcal{D}_U , thus computing the ℓ_∞ norms $\boldsymbol{\gamma}_U = \{\gamma_1, \dots, \gamma_U\}$. According to the above mentioned criterion, the proposed active sampling approach

selects samples have larger ℓ_∞ norms, which correspond to samples away from the boundary.

3. EXPERIMENTAL RESULTS

In this section, we evaluate the proposed active sampling approach on a real hyperspectral analysis scenario. The well-known AVIRIS Indian Pines scene was used in our experiments. These data were collected over Northwestern Indiana in June 1992 [14], and contains 145×145 pixels and 220 spectral bands. A total of 20 bands were removed prior to experiments due to noise and water absorption in those channels. The ground-truth data contains 16 mutually exclusive classes, and a total of 10366 labeled pixels. This image is a classic benchmark to validate the accuracy of hyperspectral image analysis algorithms and constitutes a challenging problem due to the significant presence of mixed pixels in all available classes, and also because of the unbalanced number of available labeled pixels per class.

For the first step of our newly developed approach, we run the BT active sampling approach [5] to select the initial candidate set. We select BT sampling for this purpose because it is widely used and well-established in active sampling applications with hyperspectral data [3, 19]. After we obtain the initial candidate set by BT, we run spectral unmixing based on SISAL in order to select the final set of samples. The initial set, 5 per class in our experiments, is randomly selected from the available labeled samples. The remaining samples are used for validation purposes. In order to increase statistic confidence, in our experiments each value is obtained as the mean of 10 Monte Carlo runs.

Figure 1 illustrates the obtained kappa statistic as a function of the number of training samples obtained by our proposed active sampling approach. It is noticeable that the proposed active sampling approach outperforms the original BT sampling. With the same size of training samples, the proposed approach obtained better accuracy. On the other hand, in order to obtain a similar accuracy the proposed method needs less training samples. This means that the samples actively selected by our proposed approach are more informative. For example, in order to obtain an accuracy around 80%, the BT sampling needs around 440 training samples. However, the proposed approach only needs around 390 samples. These preliminary results suggest that the proposed approach can lead to better sampling for active learning in hyperspectral image analysis applications, although further experiments with additional scenes and baseline comparison methods should be conducted in order to fully substantiate these remarks.

In order to better illustrate the performance achieved by our proposed method, Figure 2 shows the training samples obtained by the BT and the proposed active sampling approach for two classes. It is noticeable that the training samples obtained by these two methods are quite different, where the proposed approach have more diversity, which is very important to characterize the considered class.

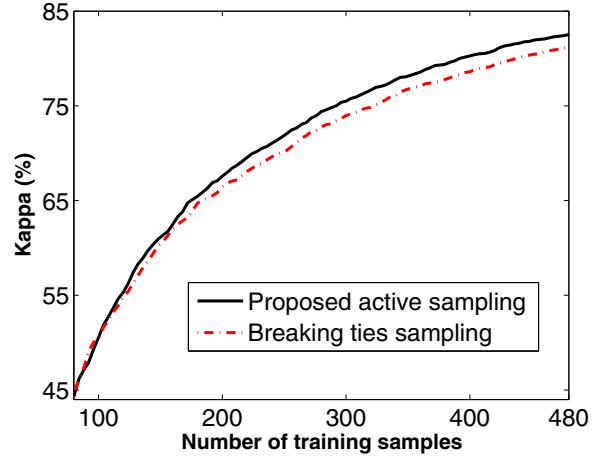


Fig. 1. Kappa statistic as a function of the number of training samples in the AVIRIS Indian Pines classification experiment.

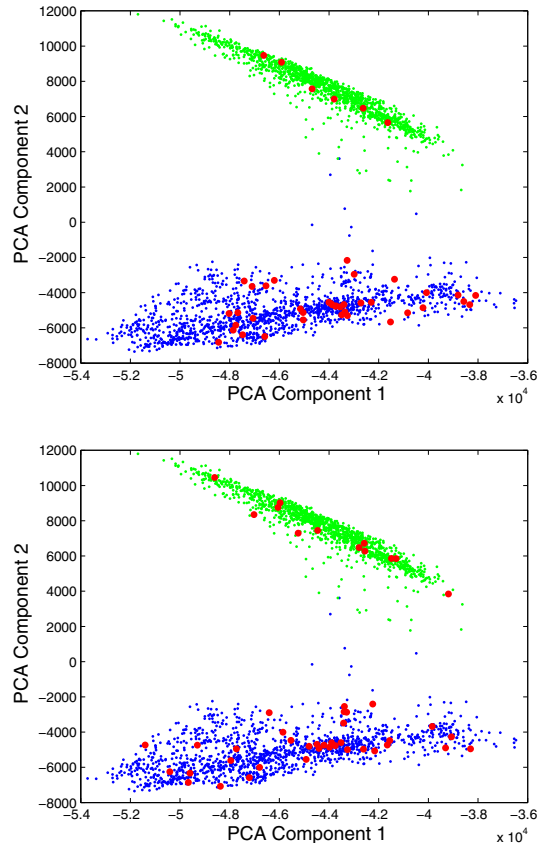


Fig. 2. Training samples (for two classes) obtained by (top) the BT sampling approach and (bottom) the proposed sampling approach, respectively

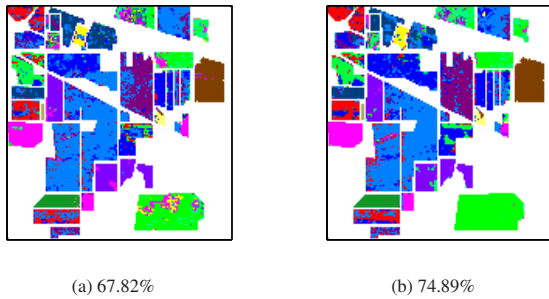


Fig. 3. Classification maps along with the kappa statistic obtained by (a) the BT sampling approach and (b) the proposed sampling approach.

For illustration purpose, Figure 3 shows the classification maps obtained by the BT and proposed active sampling method. Notice the good performance obtained by the proposed method, which outperforms the BT approach with a higher kappa of 7.07%. Furthermore, Figure 3 reveals the better classification map obtained through our proposed method.

4. CONCLUSION

In this work, we have proposed a new active sampling for hyperspectral image classification which integrates discriminative hyperspectral classification with linear spectral unmixing. Our preliminary results, conducted with a widely used hyperspectral data set, reveal that active sampling can benefit from the inclusion of a hyperspectral unmixing framework. Future work should analyze related parameters, conduct experiments with additional data sets as well as comparisons with other active sampling approaches.

5. ACKNOWLEDGMENTS

This research was supported by MRTN-CT-2006-035927 and AYA2008-05965-C04-02 projects. The authors gratefully thank D. Landgrebe for providing the hyperspectral data set.

6. 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] 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.
- [3] 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.
- [4] B. Krishnapuram, D. Williams, Y. Xue, A. Hartemink, L. Carin, and M. Figueiredo, "On semi-supervised classification," in *Proceeding of the 18th Annual Conference on Neural Information Processing Systems*, 2004.
- [5] T. Luo, K. Kramer, D. B. Goldgof, S. Samson, A. Rensen, T. Hopkins, and D. Cohn, "Active learning to recognize multiple types of plankton," *Journal of Machine Learning Research*, vol. 6, pp. 589–613, 2005.
- [6] P. Mitra, B. U. Shankar, and S. K. Pal, "Segmentation of multi-spectral remote sensing images using active support vector machines," *Pattern Recognition Letters*, vol. 25, no. 9, pp. 1067–1074, 2004.
- [7] S. Rajan, J. Ghosh, and M. M. Crawford, "An active learning approach to hyperspectral data classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, pp. 1231–1242, 2008.
- [8] J. Li, J. Bioucas-Dias, and A. Plaza, "Hyperspectral image segmentation using a new Bayesian approach with active learning," *IEEE Transactions on Geoscience and Remote Sensing*, accepted for publication, 2011.
- [9] N. Keshava and J. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, vol. 19, pp. 44–57, 2002.
- [10] A. Plaza, P. Martinez, R.M. Perez, and J. Plaza., "A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 3, pp. 650–663, 2004.
- [11] I. Dopido, M. Zortea, A. Villa, A. Plaza, and P. Gamba, "Unmixing prior to supervised classification of remotely sensed hyperspectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 4, pp. 760–764, 2011.
- [12] D. Böhning, "Multinomial logistic regression algorithm," *Annals of the Institute of Statistical Mathematics*, vol. 44, pp. 197–200, 1992.
- [13] 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.
- [14] R. Green, M. Eastwood, C. Sarture, T. Chrien, M. Aronsson, B. Chippendale, J. Faust, B. Pavri, C. Chovit, M. Solis, M. Olah, and O. Williams, "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sensing of Environment*, vol. 65, pp. 227–248, 1998.
- [15] 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.
- [16] C.-I Chang, *Hyperspectral imaging: techniques for spectral detection and classification*, New York: Kluwer, 2003.
- [17] C.-I Chang and D. Heinz, "Constrained subpixel spectral detection for remotely sensed images," *IEEE transactions on geoscience and remote sensing*, vol. 38, pp. 1144–1159, 2000.
- [18] J. M. Bioucas-Dias and M. Figueiredo, "Logistic regression via variable splitting and augmented Lagrangian tools," Tech. Rep., Instituto Superior Técnico, TULisbon, 2009.
- [19] 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, no. 11, pp. 4085–4098, 2010.