SEMI-SUPERVISED HYPERSPECTRAL IMAGE CLASSIFICATION USING A NEW (SOFT) SPARSE MULTINOMIAL LOGISTIC REGRESSION MODEL

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

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

ABSTRACT

In this work, we propose a new semi-supervised classification algorithm for remotely sensed hyperspectral images. The main contribution of this work is the development of new soft sparse multinomial logistic regression (S^2MLR) model which exploits both *hard* and *soft* labels. In our terminology, these labels respectively correspond to labeled and unlabeled training samples. In order to obtain the soft labels, we use a recently proposed subspace-based MLR algorithm (MLRsub). The proposed semi-supervised algorithm represents an innovative contribution with regards to conventional semi-supervised learning algorithms that only assign hard labels to unlabeled samples. The effectiveness of our proposed method is evaluated via experiments with a widely used hyperspectral image collected by the Airborne Visible Infra-Red Imaging Spectrometer (AVIRIS) over the Indian Pines region in Indiana. Our results indicate that the proposed method provides state-of-the-art performance when compared to other methods.

Index Terms— Hyperspectral image classification, semi-supervised learning, soft labels, unlabeled training samples, sparse multinomial logistic regression.

1. INTRODUCTION

Remotely sensed hyperspectral image classification is an active area of research [1]. It amounts at taking advantage of the detailed information contained in hyperspectral pixel vectors (spectral signatures) to generate thematic maps. A relevant challenge for supervised classification techniques (which assume prior knowledge in the form of class labels for some spectral signatures) is the limited size of labeled training sets, since their collection generally involves expensive ground campaigns. 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 fostered the idea of semi-supervised learning techniques, in which the main assumption is that new (unlabeled) training samples can be obtained from the (limited) set of available labeled samples without significant effort/cost, and without the need to design a ground campaign.

Semi-supervised learning has evolved into more complex generative models [2–5], self-learning models [6, 7], multi-view learning models [8,9], transductive support vector machines (SVMs) [10,11], and graph-based methods [12, 13]. A survey of semi-supervised learning algorithms is available in [14]. Most semi-supervised learning algorithms use some type of regularization which encourages that "similar" features belong to the same class. The effect of this regularization is to push the boundaries between classes towards regions of low data density [15], where a rather usual way of building such regularizer is to associate the vertices of a graph to the complete set of samples and then build the regularizer depending on variables defined on the vertices. This trend has been successfully adopted in several remote sensing image classification studies [16–21].

In general, these algorithms usually add *hard* labels to the set of unlabeled samples, then use jointly the labeled and unlabeled training set to perform classification [18, 22]. Although these methods exhibit very good performance, difficulties may arise from the viewpoint of the complexity of the model and its high computational cost. Furthermore, when a hard label is inappropriately estimated (which may often happen when limited training samples are available) the learning process is driven by unlabeled samples. It is well-known that hyperspectral images are dominated by mixed pixels, and hence assigning a *hard* label to a pixel as a whole may be a potential source of errors if several spectral constituents participate in the spectral signature associated to the pixel [20]. This calls for new developments in the area of semi-supervised learning, which could exploit unlabeled information (e.g. by means of *soft* labels) in a more effective way. The use of soft classification labels has been very rarely studied in semi-supervised learning. In [23] soft labels are combined with harmonic energy minimization into an external classifier based on the graph strategy. In [24], posterior marginals are considered as *soft labels* which lead to good performance. In [25], a novel fuzzy-input fuzzy-output SVM classifier (F^2SVM) was designed to address sub-pixel classification problems. The proposed F^2SVM algorithm uses an input fuzzy membership function to model the sub-pixel abundances of unknown patterns in the learning process.

In this work, we develop a soft sparse multinomial logistic regression (S^2MLR) model which belongs to the family of self-learning algorithms. Compared with the original model in which it is inspired [26,27], the S^2MLR makes use of both hard and soft classification labels. As opposed to other self-learning methods which expand the training set by using unlabeled samples with hard labels, the proposed semi-supervised algorithm exploits the concept of soft labels generating unlabeled training samples, where the soft labels are obtained by using a recently proposed subspace-based MLR algorithm (MLRsub) [28]. This algorithm is specifically designed to address the problem of mixed pixels and provides high confidence for the posterior probabilities estimates for such pixels.

The remainder of the paper is organized as follows. Section 2 introduces the S^2MLR model and presents the proposed semi-supervised learning strategy. Section 3 evaluates the effectiveness of the proposed method via experiments with a widely used hyperspectral image, collected by the Airborne Visible Infra-Red Imag-

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

978-1-4577-2201-1/11/$26.00 ©2011 IEEE

ing Spectrometer (AVIRIS), over the Indian Pines region in Indiana. Section 4 concludes the paper with some remarks and hints at plausible future research lines.

2. SOFT SPARSE MULTINOMIAL LOGISTIC REGRESSION (S^{2}MLR) MODEL

First, we briefly define the notations used in this paper. 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; $\mathbf{x} \equiv (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ an image of d-dimensional feature vectors, $\mathbf{y} \equiv (y_1, \ldots, y_n)$ an image of labels, and $\mathbf{y}_i = [y_i^{(1)}, ..., y_i^{(K)}]^T$ denote a "1-of-K" encoding of the K classes, where $\sum_{k=1}^{K} y_i^{(k)} = 1$, $y_i^{(k)} \in \{0, 1\}$ for hard labels and $y_i^{(k)} \in [0, 1]$ for soft labels; L be the number of labeled training samples; U be the number of unlabeled training samples. In the following, we describe the classic sparse multinomial logistic regression (SMLR) model and our proposed variation (S^{2}MLR).

2.1. The original SMLR

The original SMLR [26] models the posterior class probabilities as

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

where $\mathbf{h}(\mathbf{x}_i) \equiv [h_1(\mathbf{x}_i), ..., h_l(\mathbf{x}_i)]^T$ is a vector of l fixed functions of the input, often termed features; $\boldsymbol{\omega}$ denotes the regressors and $\boldsymbol{\omega} \equiv [{\boldsymbol{\omega}^{(1)}}^T, ..., {\boldsymbol{\omega}^{(K-1)}}^T]^T$. Since the density (1) does not depend on translations on 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.*, $\mathbf{h}(\mathbf{x}_i) = [1, x_{i,1}, ..., x_{i,d}]^T$, 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.*, $\mathbf{h}(\mathbf{x}_i) = [1, K_{\mathbf{x}_i,\mathbf{x}_1}, ..., K_{\mathbf{x}_i,\mathbf{x}_l}]^T$, 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 [29].

In order to control the machine complexity and, thus, its generalization capacity, following the SMLR algorithm introduced in [27], we model $\boldsymbol{\omega}$ as a random vector with Laplacian density $p(\boldsymbol{\omega}) \propto \exp(-\lambda\|\boldsymbol{\omega}\|_1)$, where λ is the regularization parameter controlling the degree of sparsity of $\boldsymbol{\omega}$. Under the present setting, learning the class densities amounts to estimating the logistic regressors $\boldsymbol{\omega}$ [21, 27, 30]. Adopting the maximum a posteriori (MAP) estimation criterium, we have

$$\widehat{\boldsymbol{\omega}}_{\text{MAP}} = \arg\max_{\boldsymbol{\omega}} \quad \ell_L(\boldsymbol{\omega}) + \log p(\boldsymbol{\omega}), \qquad (2)$$

where $\ell_L(\boldsymbol{\omega})$ is the log-likelihood function on the labeled information and

$$\ell_L(\boldsymbol{\omega}) = \sum_{i=1}^{L}\left(\sum_{k=1}^{K} y_i^{(k)}\boldsymbol{\omega}^{(k)}\mathbf{x}_i - \log\sum_{k=1}^{K}\exp(\boldsymbol{\omega}^{(k)}\mathbf{x}_i)\right). \qquad (3)$$

Notice that, $y_i^{(k)}$ are *hard labels*, *i.e.*, $y_i^{(k)} \in \{0,\ 1\}$.

2.2. Soft Sparse Multinomial Logistic Regression (S^{2}MLR)

Let $\{(y_i, \mathbf{x}_i) : i = 1, \ldots, L\}$ and $\{(y_i, \mathbf{x}_i) : i = L+1, \ldots, L+U\}$, respectively, the labeled the unlabeled data sets. Since our approach is semi-supervised, we will learn the classifier from the labeled data, usually a set of small size, and from the unlabeled data sets, usually a larger set. The usual way to do inference with unobserved data is the expectation maximization (EM) algorithm. In the present context, the E-step and M-step this algorithm are given by

E-step: $Q(\boldsymbol{\omega}|\widehat{\boldsymbol{\omega}}_t) \equiv \ell_L(\boldsymbol{\omega}) + \ell_U(\boldsymbol{\omega}) + \log p(\boldsymbol{\omega})$,
M-step: $\widehat{\boldsymbol{\omega}}_{t+1} \equiv \arg\max Q(\boldsymbol{\omega}|\widehat{\boldsymbol{\omega}}_t)$,

where $\ell_L(\boldsymbol{\omega})$ is given by (3) and

$$\ell_U(\boldsymbol{\omega}) = \sum_{i=L+1}^{L+U}\left(\sum_{k=1}^{K}\widehat{y}_i^{(k)}\boldsymbol{\omega}^{(k)}\mathbf{x}_i - \log\sum_{k=1}^{K}\exp(\boldsymbol{\omega}^{(k)}\mathbf{x}_i)\right), \quad (4)$$

where $\widehat{y}_i^{(k)} \equiv E[y_i^{(k)} \,|\, \mathcal{X}_U, \widehat{\boldsymbol{\omega}}_t] = p(y_i = k|\mathcal{X}_U, \widehat{\boldsymbol{\omega}}_t)$; *i.e.*, $\ell_U(\boldsymbol{\omega})$ has the same structure of (3) but now $\widehat{y}_i^{(k)}$ denotes probabilities, often called *soft labels* to contrast with the hard labels supplied with the training set.

In this work, we heuristically replace the soft labels $\widehat{y}_i^{(k)} \equiv E[y_i^{(k)} \,|\, \mathcal{X}_U, \widehat{\boldsymbol{\omega}}_t] = p(y_i = k|\mathcal{X}_U, \widehat{\boldsymbol{\omega}}_t)$ with the probabilities given by the MLR*sub* algorithm [28]. This is because the MLR*sub* algorithm [28] is very well suited to mixed pixels, *i.e.*, pixels with materials appearing in more than one class and thus difficult to classify. More details about the MLR*sub* algorithm can be found in [28].

It should be noted that, if the features are given by kernels, then the original SMLR is limited to data sets with products $L \times K$ not larger than, say, 1000. Therefore, most hyperspectral data sets are beyond the reach of this algorithm. This difficulty was recently removed by the introduction of the logistic regression via variable splitting and augmented Lagrangian (LORSAL) algorithm [31], which is able to deal with training sets with a few thousands training samples, regardless of the number of classes. LORSAL plays a central role, for example, in [21, 30]. In this paper, following our previous work, we resort to the LORSAL algorithm to estimate the regressors.

3. EXPERIMENTAL RESULTS

The well-known AVIRIS Indian Pines scene was used in our experiments. The data were collected over Northwestern Indiana in June of 1992 [1], 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. In our experiments, labeled training samples are randomly selected from the ground-truth data, whereas the remaining samples are used as the validation set. In order to increase the statistical significance of the results, each value of overall accuracy (OA), kappa statistic (κ) and average accuracy (AA) reported in this work is obtained as the average of 10 Monte Carlo (MC) runs.

Fig. 1 illustrates the obtained OA, κ and AA results as a function of the number of unlabeled training samples with 160 labeled samples (10 samples per class). This constitutes a very low number of samples, thus leading to a very difficult classification problem

regardless of whether a supervised or a semi-supervised technique is adopted. In the former case, poor generalization capability is expected. In the latter case, the trade-off between a large number of unlabeled samples versus a small number of labeled samples could bias the learning process. To address these problems, we resort to the following iterative scheme:

1. The posterior probabilities of unlabeled samples are estimated by the MLR*sub* algorithm [28].
2. A set of U unlabeled samples (with $U \leq L$) are selected and incorporated to the set of L labeled samples for joint consideration by the proposed S^{2}MLR-based classifier.
3. Steps 1-2 are repeated in iterative fashion until some *stopping criterion* is met. The criterion considered in this work is a maximum number of unlabeled samples.

The proposed iterative scheme offers two main advantages. First and foremost, the MLR*sub* classifier can be trained jointly with labeled and unlabeled samples by means of the S^{2}MLR model, thus leading to better posterior probability estimates which will can be used as soft labels in the proposed framework. Second, with the setting of $U \leq L$, the influence of unlabeled training samples can be more balanced in the case of poor generalization capability by simply expanding the set of unlabeled samples iteratively.

Fig. 1 reports the classification accuracies obtained by the proposed semi-supervised method as the number of unlabeled samples is increased. It can be observed that the proposed algorithm leads to very good performance in terms of OA, κ and AA scores. For instance, with 160 labeled samples and no unlabeled samples, the proposed algorithm obtained OA=61.70%, $\kappa = 57.22$ and AA=73.04%. However, by including 1577 unlabeled samples (which come at very low cost) the proposed algorithm increased the classification accuracy to OA=69.37%, $\kappa = 65.45\%$ and AA=79.13%. Note that all the aforementioned metrics were significantly improved with 7.67%, 8.23% and 6.09% improvement, respectively. Another important observation is that the classification results increase as the number of unlabeled samples increases.

For illustrative purposes, Table 1 shows the classification results obtained for each of the 10 conducted MC runs. The table reveals that the classification results are improved in all cases by using unlabeled samples which are very easy to generate in our proposed framework. From these results we can also conclude that the performance of the proposed semi-supervised algorithm is relevant to the size of the unlabeled samples. Furthermore, in the case of poor generalization (*see* MC4 in Table 1), the proposed semi-supervised algorithm still highly improved the accuracies. This is because the S^{2}MLR model allows using unlabeled samples in conjunction with labeled samples to better optimize the class boundaries and associate together all of the samples sharing common properties in the same region, while preserving the relevance of labeled samples in the process thus allowing a better balance between labeled and unlabeled samples in the classification process.

For illustrative purposes, Fig. 2 shows the classification maps obtained in one of the conducted Monte Carlo runs (MC10 in Table 1) along with the respective ground-truth map. The improvements in classification obtained by means of the proposed strategy can be graphically observed in the results reported in this figure.

4. CONCLUSION

In this work, we have developed a new soft sparse multinomial logistic regression model (S^{2}MLR) which uses both *hard* and *soft* classification labels as opposed to other semi-supervised learning algorithm which are usually based on hard labels when deriving unlabeled training samples. Our proposed strategy allows us to model better the phenomenon of mixed pixels present in hyperspectral images by accepting soft labels from many different sources. In this work, we use the posterior probabilities obtained by a recently proposed subspace-based multinomial logistic regression algorithm (MLR*sub*) as soft labels, mainly because these probabilities exhibit high confidence in the estimates provided for mixed pixels. The effectiveness of the proposed method has been evaluated via experiments with a widely used hyperspectral image, collected by the Airborne Visible Infra-Red Imaging Spectrometer (AVIRIS) over the Indian Pines region in Indiana. Our results indicate that the proposed semi-supervised algorithm provides state-of-the-art performance with very limited training samples. In the future, additional sources for the generation of soft labels (e.g., the abundances obtained after classic spectral unmixing) as well as additional hyperspectral data sets will be used to fully substantiate the contributions of our newly introduced framework. Additional studies will also be conducted on the influence of the size of set of unlabeled samples available at each algorithm iteration. In future work, we will also target additional mechanisms for exploiting the unlabeled information in a more efficiently way, *e.g.*, by means of active learning.

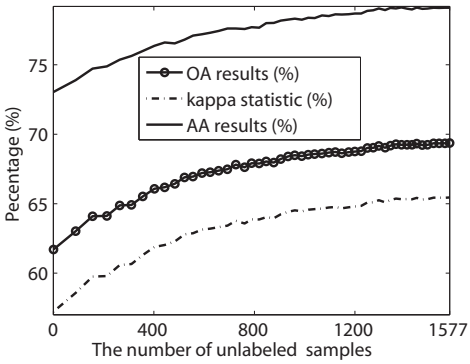


Fig. 1. OA [%], κ [%] and AA [%] results as a function of the number of unlabeled samples obtained after 10 Monte Carlo (MC) runs using the proposed semi-supervised algorithm with 160 labeled samples (10 per class).

5. REFERENCES

[1] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*, John Wiley, Hoboken, NJ, 2003.

[2] Shumeet Baluja, "Probabilistic modeling for face orientation discrimination: Learning from labeled and unlabeled data," in *Neural Information Procesing systems (NIPS '98)*, 1998.

[3] Avrim Blum and Tom Mitchell, "Combining labeled and unlabeled data with co-training," 1998, pp. 92–100, Morgan Kaufmann Publishers.

[4] Tom M. Mitchell, "The role of unlabeled data in supervised learning," in *In Proceedings of the Sixth International Colloquium on Cognitive Science*, 1999.

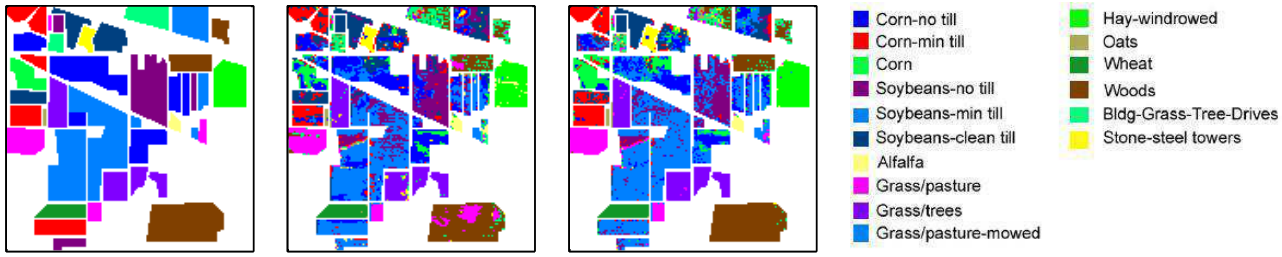
[5] Akinori Fujino, Naonori Ueda, and Kazumi Saito, "A hybrid generative/discriminative approach to semi-supervised classifer degin," in *AAAI'05 Proceedings of the 20th national coference on Artificial intelligence*, 2005, vol. 2.

[6] David Yarowsky, "Unsupervised word sense disambiguation rivaling supervised methods," in *Proceedings of the 33rd annual meeting on Association for Computational Linguistics*, 1995, ACL'95, pp. 189–196.

[7] Charles Rosenberg, Martial Hebert, and Henry Schneiderman, "Semi-supervised self-training of object detection models," in *Seventh IEEE Workshop on Applications of Computer Vision*, January 2005.

Table 1. OA [%] OA [%], κ [%] and AA [%] results for 10 Monte Carlo (MC) runs for the proposed semi-supervised algorithm with 0 and 1577 unlabeled samples, respectively.

			MC runs									
			MC1	MC2	MC3	MC4	MC5	MC6	MC7	MC8	MC9	MC10
Unlabeled Samples	0	OA	63.87	59.54	64.8	56.92	61.44	62.14	61.01	64.17	60.88	62.26
		κ	59.57	55.09	60.50	52.59	56.7	57.42	56.67	59.86	56.32	57.48
		AA	73.71	74.1	74.89	72.37	71.24	73.33	73.57	73.89	71.72	71.52
	1577	OA	69.17	69.4	67.06	68.92	73.21	65.49	69.21	69.45	71.51	70.23
		κ	65.17	65.71	62.53	65.35	69.59	60.89	65.4	65.92	67.79	66.15
		AA	78.94	79.76	78.63	80.36	78.81	77.88	78.81	79.83	79.41	78.64



(a) (b) (c)

Fig. 2. Classification maps obtained by the proposed semi-supervised algorithm for the AVIRIS Indian Pines image in one of the conducted Monte Carlo runs (MC10 in Table 1) using 160 labeled samples and 1577 unlabeled samples. (a) Ground-truth image. (b) S^2MLR-based classifier (no unlabeled samples). (c) S^2MLR-based classifier (1577 unlabeled samples).

[8] In Cowan, Tesauro G, and Virginia R. De Sa, "Learning classification with unlabeled data," 1994.

[9] Ulf Brefeld, Thomas Gärtner, Tobias Scheffer, and Stefan Wrobel, "Efficient coregularised least squares regression," in *Proceedings of the 23rd international conference on Machine learning*, 2006, ICML'06, pp. 137–144.

[10] V. Vapnik, *Statistical Learning Theory*, John Wiley, New York, 1998.

[11] Thorsten Joachims, "Transductive inference for text classification using support vector machines," in *Proceedings of the Sixteenth International Conference on Machine Learning*, 1999, ICML'99, pp. 200–209.

[12] Avrim Blum and Shuchi Chawla, "Learning from labeled and unlabeled data using graph mincuts," in *Proceedings of the Eighteenth International Conference on Machine Learning*, 2001, ICML'01, pp. 19–26.

[13] Xiaojin Zhu, Zoubin Ghahramani, Tommi Jaakkola, and Mit Ii Abstract, "Semi-supervised learning with graphs (technical report cmu-cald-02-106)," Tech. Rep., Carnegie Mellon University, 2005.

[14] Xiaojin Zhu, "Semi-supervised learning literature survey," Tech. Rep. 1530, Computer Sciences, University of Wisconsin-Madison, 2005.

[15] O. Chapelle, M. Chi, and A. Zien, "A continuation method for semi-supervised svms," in *Proceedings of the 23rd International Conference on Machine Learning*. 2006, pp. 185–192, ACM Press.

[16] Y. Zhong, L. Zhang, B. Huang, and L. Pingxiang, "An unsupervised artificial immune classifier for multi/hyperspectral remote sensing imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 2, pp. 420–431, 2006.

[17] 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. 11, pp. 3363–3373, 2006.

[18] G. Camps-Valls, T. Bandos, and D. Zhou, "Semi-supervised graph-based hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, pp. 3044–3054, Oct 2007.

[19] Devis Tuia and Gustavo Camps-Valls, "Semi-supervised hyperspectral image classification with cluster kernels," *IEEE Geoscience and Remote Sensing Letters. Accept for publication*, 2009.

[20] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sensing of Environment*, vol. 113, pp. 110–122, September 2009.

[21] 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.

[22] J. Li, J. Bioucas-Dias, and A. Plaza, "Semi-supervised hyperspectral image classification based on a markov random field and sparse multinomial logistic regression," in *Proc. IEEE International Geoscience and Remote sensing Symposium*, 2009.

[23] Xiaojin Zhu, Zoubin Ghahramani, and John Lafferty, "Semi-supervised learning using gaussian fields and harmonic functions," in *The 20th International Conference on Machine Learning (ICML)*, 2003, pp. 912–919.

[24] J. Li, J. Bioucas-Dias, and Antonio Plaza, "Exploiting spatial information in semi-supervised hyperspectral image segmentation," in *IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS'10)*, 2010.

[25] F. Bovolo, L. Bruzzone, and L. Carline, "A novel technique for subpixel image classification based on support vection machine," *IEEE Transactions on Image Processing*, vol. 19, pp. 2983–2999, 2010.

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

[27] 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.

[28] J. Li, J. Bioucas-Dias, and Antonio Plaza, "pectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and markov random fields," *IEEE Transactions on Geoscience and Remote Sensing (submitted),http://www.lx.it.pt/∼jun/*, 2011.

[29] 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.

[30] 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 (submitted)*, 2010.

[31] J.M. Bioucas-Dias and M. Figueiredo, "Logistic regression via variable splitting and augmented lagrangian tools," Tech. Rep., Instituto Superior Técnico, TULisbon, 2009.