

A NEW SEMI-SUPERVISED APPROACH FOR HYPERSPECTRAL IMAGE CLASSIFICATION WITH DIFFERENT ACTIVE LEARNING STRATEGIES

Inmaculada Dórido¹, Jun Li^{1,2}, Antonio Plaza¹ and José M. Bioucas-Dias²

¹Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, University of Extremadura, E-10071 Caceres, Spain .

²Instituto de Telecomunicações, Instituto Superior Técnico, TULisbon, Lisbon, Portugal.

ABSTRACT

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. 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. 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 and further evaluate two different strategies for active sample selection. In this work, the proposed approach is illustrated with the sparse multinomial logistic regression (SMLR) classifier learned with the MLR via variable splitting and augmented Lagrangian (LORSAL) algorithm. Our experimental results with a real hyperspectral image collected by the NASA Jet Propulsion Laboratory's Airborne Visible Infra-Red Imaging Spectrometer (AVIRIS) indicate that the use of AL for unlabeled sample selection represents an effective and promising strategy in the context of semi-supervised hyperspectral data classification.

Index Terms— Hyperspectral image classification, semi-supervised learning, active learning, sparse multinomial logistic regression.

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]. A relevant challenge for supervised classification techniques (which assume prior knowledge in the form of class labels for different spectral signatures) is the limited availability of labeled training sets, since their collection generally involves expensive ground campaigns [2]. While the collection of labeled samples

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 Spanish Ministry of Science and Innovation (HYPERCOMP/EODIX project, reference AYA2008-05965-C04-02) and from Portuguese Science and Technology Foundation, project PEst-OE/EEI/LA0008/2011 are also gratefully acknowledged.

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 new (unlabeled) training samples can be obtained from a (limited) set of available labeled samples without significant effort/cost [3].

In contrast to supervised classification, the aforementioned SSL algorithms generally assume that a limited number of labeled samples are available *a priori* and then enlarge the training set using unlabeled samples, thus allowing these approaches to address ill-posed problems. However, in order for this strategy to work several requirements need to be faced. First and foremost, the new (unlabeled) samples should be generated without significant cost/effort. Second, unlabeled samples should be representative enough in order for the SSL classifier to model the available classes without the need for a large number of unlabeled samples. In other words, if the unlabeled samples are not properly selected these may confuse the classifier, thus introducing divergence or even reducing the classification accuracy obtained with the initial set of labeled samples. In order to address these issues, it is very important that the most highly informative unlabeled samples are identified, so that significant improvements in classification performance can be observed without the need to increase a very high number of unlabeled samples.

In this work, we evaluate the feasibility of using active learning (AL) techniques for automatically selecting unlabeled samples. In the literature, AL techniques have been mainly exploited in a supervised context, *i.e.* a given supervised classifier is trained with the most representative training samples selected after a machine-user interaction process in which the training samples are actively selected according to some criteria based on the considered classifier, and then the labels of these samples are assigned by a trained expert in fully supervised fashion [4]. However, in our proposed approach we exploit AL in semi-supervised fashion by allowing the classifier to actively select and label new training samples itself. The idea is to first use a subset of the available (labeled) training set as the candidate pool for the AL process. However, different strategies can be considered for generating such candidate pool. In this work, we consider two different strategies for this purpose and illustrate them with the sparse multinomial logistic regression (SMLR) classifier, which is shown to achieve significant improvements in classification accuracy resulting from its combination with the proposed AL-based strategies. 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.

The remainder of the paper is organized as follows. Section 2

describes proposed framework for semi-supervised AL. Section 3 reports classification results using a real hyperspectral image collected by the Airborne Visible Infra-Red Imaging Spectrometer (AVIRIS) [5] over the Indian Pines region in NW Indiana. Finally, section 4 concludes the paper with some remarks and hints at plausible future research lines.

2. PROPOSED APPROACH

Let $\mathcal{K} \equiv \{1, \dots, K\}$ denote a set of K class labels, $\mathcal{S} \equiv \{1, \dots, n\}$ a set of integers indexing the n pixels of an image, $\mathbf{x} \equiv (\mathbf{x}_1, \dots, \mathbf{x}_n) \in \mathbb{R}^{d \times n}$ an image of d -dimensional feature vectors, $\mathbf{y} \equiv (y_1, \dots, y_n)$ an image of labels, $\mathcal{D}_l \equiv \{(y_{l_1}, \mathbf{x}_{l_1}), \dots, (y_{l_n}, \mathbf{x}_{l_n})\}$ a set of labeled samples, l_n the number of labeled training samples, $\mathcal{Y}_l \equiv \{y_{l_1}, \dots, y_{l_n}\}$ the set of labels in \mathcal{D}_l , $\mathcal{X}_l \equiv \{\mathbf{x}_{l_1}, \dots, \mathbf{x}_{l_n}\}$ the set of feature vectors in \mathcal{D}_l , $\mathcal{D}_u \equiv \{\mathcal{X}_u, \mathcal{Y}_u\}$ a set of unlabeled samples, $\mathcal{X}_u \equiv \{\mathbf{x}_{u_1}, \dots, \mathbf{x}_{u_n}\}$ the set of unlabeled feature vectors in \mathcal{D}_u , $\mathcal{Y}_u \equiv \{y_{u_1}, \dots, y_{u_n}\}$ the set of labels associated with \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 [6]:

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

where $\mathbf{h}(\mathbf{x}) = [h_1(\mathbf{x}), \dots, h_l(\mathbf{x})]^T$ is a vector of l fixed functions of the input, often termed features; $\boldsymbol{\omega}$ are the regressors and $\boldsymbol{\omega} = [\boldsymbol{\omega}^{(1)T}, \dots, \boldsymbol{\omega}^{(K)T}]^T$. Notice that the function \mathbf{h} may be linear, *i.e.*, $\mathbf{h}(\mathbf{x}_i) = [1, x_{i,1}, \dots, x_{i,d}]^T$, where $x_{i,j}$ is the j -th component of \mathbf{x}_i ; or nonlinear, *i.e.*, $\mathbf{h}(\mathbf{x}_i) = [1, K_{\mathbf{x}_i, \mathbf{x}_1}, \dots, K_{\mathbf{x}_i, \mathbf{x}_{l_n}}]^T$, 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 [7]. 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 [8, 9], we infer $\boldsymbol{\omega}$ by computing the maximum a posteriori (MAP) estimate:

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

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 $\hat{\boldsymbol{\omega}}$, set empirically in this work to $\lambda = 0.001$. The optimization problem (2) is solved by the LORSAL algorithm [10] (see also the appendix of [9]). $\ell(\boldsymbol{\omega})$ 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 in this work using AL techniques described in the following subsection.

2.2. Active Learning (AL)

In this work, we adopt the AL concept from supervised learning [4, 11, 12] and combine it with the SSL strategies described in the previous subsection. 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 using two different strategies:

1. **Strategy 1.** Let $\mathcal{D}_{\mathcal{N}(i)} \equiv \{(\hat{y}_{i_1}, \mathbf{x}_{i_1}), \dots, (\hat{y}_{i_n}, \mathbf{x}_{i_n})\}$ be the set of neighboring set of samples of (y_i, \mathbf{x}_i) for $i \in \{1, \dots, l_n, u_1, \dots, u_n\}$, where i_n is the number of samples in $\mathcal{D}_{\mathcal{N}(i)}$ and \hat{y}_{i_j} is the maximum a posteriori probability (MAP) estimate from the MLR classifier, with $i_j \in \{1, \dots, i_n\}$. If $\hat{y}_{i_j} = y_i$, we increment the unlabeled training set by adding $(\hat{y}_{i_j}, \mathbf{x}_{i_j})$, *i.e.*, $\mathcal{D}_u = \{\mathcal{D}_u, (\hat{y}_{i_j}, \mathbf{x}_{i_j})\}$. 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.
2. **Strategy 2.** A second strategy is to increment the unlabeled training set by adding $(\hat{y}_{i_j}, \mathbf{x}_{i_j})$, *i.e.*, $\mathcal{D}_u = \{(\hat{y}_{i_j}, \mathbf{x}_{i_j})\}$ in each iteration, *i.e.*, the previously selected labeled and unlabeled training samples are removed from the pool of candidates at each iteration, so that at each iteration a completely new set of unlabeled samples is selected. We emphasize that, 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.

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 (instead of user-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 between the classes are preferred. In our study, we use three different AL techniques [13] to evaluate the proposed framework: 1) margin sampling (MS), 2) breaking ties (BT), and 3) modified breaking ties (MBT) [9] in addition to random selection (RS).

3. EXPERIMENTAL RESULTS

The hyperspectral image used in experiments was collected by the AVIRIS sensor over the Indian Pines region in Northwestern Indiana in 1992. This scene, with a size of 145 lines by 145 samples, was acquired over a mixed agricultural/forest area, early in the growing season. The scene comprises 202 spectral channels in the wavelength range from 0.4 to 2.5 μm , nominal spectral resolution of 10 nm, moderate spatial resolution of 20 meters by pixel, and 16-bit radiometric resolution. After an initial screening, several spectral bands were removed from the data set due to noise and water absorption phenomena, leaving a total of 164 radiance channels to be used

in the experiments. These data, including ground-truth information, are available online¹, a fact which has made this scene a widely used benchmark for testing the accuracy of hyperspectral data classification algorithms. This scene constitutes a challenging classification problem due to the presence of mixed pixels in all available classes, and because of the unbalanced number of available labeled pixels per class.

Table 1 shows the overall, average, individual classification accuracies (in percentage) and the κ statistic obtained by the supervised MLR (trained using only $l = 5$, $l = 10$ and $l = 15$ labeled samples per class), and by the proposed SSL approach (based on the same classifier) using the four considered AL algorithms (executed using 50 iterations) and the strategy 1 for candidate selection. Similarly, Table 2 performs the same experiments but using the strategy 2 for candidate selection. Several conclusions can be obtained from Tables 1 and 2. First of all, we can notice that the inclusion of unlabeled samples significantly improves the classification results in all cases. This is expected, since the uncertainty of the classifier boundaries decreases as the training set size increases. It is also remarkable that, in the case of the strategy 2 for candidate selection, the improvements are more significant than the results obtained with the strategy 1. This is because the strategy 2 selects a set of new candidates at each iteration and the AL algorithms can perform a better sample selection from the candidate pool, which contains new samples at each iteration and allows exploring the spatial-contextual information to a wider extent.

In a second experiment we evaluated the impact of the number of unlabeled samples used to increase the classification performance achieved by the considered classifier with the two considered sample selection strategies. Fig. 1 shows the classification accuracies obtained by the proposed SSL approach (using only $l = 5$, $l = 10$ and $l = 15$ labeled samples per class and an increasing number of unlabeled samples) with the two considered strategies for active learning. Again, significant improvements can be seen when the AL strategy 2 is used when compared to the AL strategy 1. Finally, Fig. 2 shows the classification maps obtained with the strategy 2 using $l = 10$ labeled samples per class. Effective classification results can be appreciated in these maps.

4. CONCLUSIONS AND FUTURE WORK

In this paper, we have developed a new framework for semi-supervised classification of hyperspectral images in which unlabeled samples are actively selected using two different strategies. Specifically, we use active learning to select the most informative unlabeled training samples with the ultimate goal of improving classification results obtained using randomly selected training samples. In our semi-supervised context, the labels of the selected training samples are estimated by the classifier itself, with the advantage that no extra cost is required for labeling the selected samples when compared to supervised active learning. Our experimental results, conducted using the sparse multinomial logistic regression (MLR) classifier, indicate that the proposed approach can increase the classification accuracies obtained in the supervised case through the incorporation of unlabeled samples which can be obtained with very little cost and effort. In future work, we are planning on combining the proposed approach with other classifiers in order to confirm the advantages that can be gained by including actively selected unlabeled samples in the context of semi-supervised classification of remotely sensed hyperspectral image data sets.

5. 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] F. Bovolo, L. Bruzzone, and L. Carline, "A novel technique for subpixel image classification based on support vector machine," *IEEE Transactions on Image Processing*, vol. 19, pp. 2983–2999, 2010.
- [3] 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.
- [4] 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.
- [5] R. O. Green, M. L. Eastwood, C. M. Sarture, T. G. Chrien, M. Aronsson, B. J. Chippendale, J. A. Faust, B. E. Pavri, C. J. Chovit, M. Solis *et al.*, "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sensing of Environment*, vol. 65, no. 3, pp. 227–248, 1998.
- [6] D. Böhning, "Multinomial logistic regression algorithm," *Annals of the Institute of Statistical Mathematics*, vol. 44, pp. 197–200, 1992.
- [7] 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.
- [8] 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.
- [9] 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*, vol. 49, no. 10, pp. 3947–3960, 2011.
- [10] J. Bioucas-Dias and M. Figueiredo, "Logistic regression via variable splitting and augmented lagrangian tools," Instituto Superior Técnico, TULisbon, Tech. Rep., 2009.
- [11] 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.
- [12] 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.
- [13] D. Tuia and G. Camps-Valls, "Urban image classification with semisupervised multiscale cluster kernels," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 4, no. 1, pp. 65–74, march 2011.

¹Available online: <http://dynamo.ecn.purdue.edu/biehl/MultiSpec>

Table 1. Overall, average, individual classification accuracies [%], and κ statistic (in the parentheses) obtained using the MLR classifier when applied to the AVIRIS Indian Pines hyperspectral data set. Strategy 1 is used for candidate selection at each AL iteration.

	Number of labeled samples per class														
	$l = 5$					$l = 10$					$l = 15$				
	Supervised	MS	BT	MBT	RS	Supervised	MS	BT	MBT	RS	Supervised	MS	BT	MBT	RS
OA	51.78	75.07	73.90	73.32	63.33	60.12	77.50	79.09	76.28	67.36	66.20	78.82	79.50	77.40	72.54
AA	63.82	79.40	78.86	80.79	70.97	71.74	81.87	83.74	84.22	78.88	77.39	84.58	84.92	85.46	79.97
κ	46.26	71.63	70.39	69.92	58.53	55.43	74.41	76.30	73.32	63.29	62.09	75.90	76.71	74.57	68.87

Table 2. Overall, average, individual classification accuracies [%], and κ statistic (in the parentheses) obtained using the MLR classifier when applied to the AVIRIS Indian Pines hyperspectral data set. Strategy 2 is used for candidate selection at each AL iteration.

	Number of labeled samples per class														
	$l = 5$					$l = 10$					$l = 15$				
	Supervised	MS	BT	MBT	RS	Supervised	MS	BT	MBT	RS	Supervised	MS	BT	MBT	RS
Alfalfa (54)	83.88	84.29	84.49	86.33	81.63	83.64	84.32	85.68	88.86	77.50	85.38	84.87	84.62	86.67	81.79
Corn-Notill (1434)	39.10	71.20	72.32	70.22	48.12	48.38	71.83	76.08	75.18	63.01	51.40	74.50	74.27	74.93	58.44
Corn-Min (834)	34.26	52.64	54.62	60.33	38.70	47.65	67.69	65.52	65.95	55.01	59.61	72.20	73.28	71.95	57.80
Corn (234)	53.62	78.38	76.86	82.27	61.09	70.63	87.01	89.06	93.13	66.79	78.54	87.72	87.72	92.56	77.81
Grass-Pasture (497)	60.33	68.78	69.23	69.88	68.64	75.42	86.39	86.26	87.19	82.87	81.47	87.28	87.95	90.02	84.32
Grass-Trees (747)	78.68	96.47	96.67	97.30	94.61	86.01	96.70	96.45	96.81	94.75	93.44	96.89	96.89	97.45	96.09
Grass-Pasture-Mowed (26)	90.48	93.81	91.43	93.81	90.48	88.12	91.87	93.13	91.87	85.00	94.55	95.45	96.36	93.64	92.73
Hay-Windrowed (489)	74.42	99.07	99.30	99.42	96.26	88.89	98.18	98.79	99.16	97.72	90.70	98.35	98.71	98.95	97.19
Oats (20)	95.33	97.33	98.67	100.00	92.00	98.00	92.00	92.00	99.00	98.00	100.00	96.00	100.00	100.00	100.00
Soybeans-Notill (968)	52.72	74.55	76.53	76.50	63.71	58.68	82.67	84.67	82.22	68.38	61.72	77.28	78.94	77.99	65.67
Soybeans-Min (2468)	42.59	75.59	74.96	65.41	57.32	44.85	78.50	76.00	65.45	53.46	51.56	76.22	77.82	66.09	63.17
Soybeans-Clean (614)	40.21	80.66	79.98	82.25	57.45	52.50	82.52	83.56	86.90	64.39	62.25	89.65	87.41	89.72	72.95
Wheat (212)	95.31	98.99	99.03	99.23	99.08	98.76	99.50	99.36	99.60	99.55	99.24	99.39	99.19	99.44	99.54
Woods (1294)	65.03	95.77	94.77	95.02	93.20	75.63	95.02	94.89	95.13	90.79	82.24	96.51	96.46	96.80	93.74
Bldg-Grass-Tree-Drives (380)	36.85	54.27	62.24	61.79	39.36	50.84	67.41	69.59	70.92	56.97	59.15	66.41	69.89	70.74	61.10
Stone-Steel-Towers (95)	78.33	88.33	91.78	93.78	79.44	79.88	84.24	81.06	91.41	77.41	87.00	89.87	88.50	92.25	84.75
OA	51.78	78.13	78.64	76.91	65.40	60.12	82.33	82.48	80.18	69.85	66.20	82.68	83.31	80.87	72.97
AA	63.82	81.88	82.68	83.35	72.57	71.74	85.37	85.76	86.80	76.98	77.39	86.79	87.38	87.45	80.44
κ	46.26	75.08	75.69	73.93	60.84	55.43	79.88	80.09	77.67	66.02	62.09	80.29	80.99	78.37	69.38

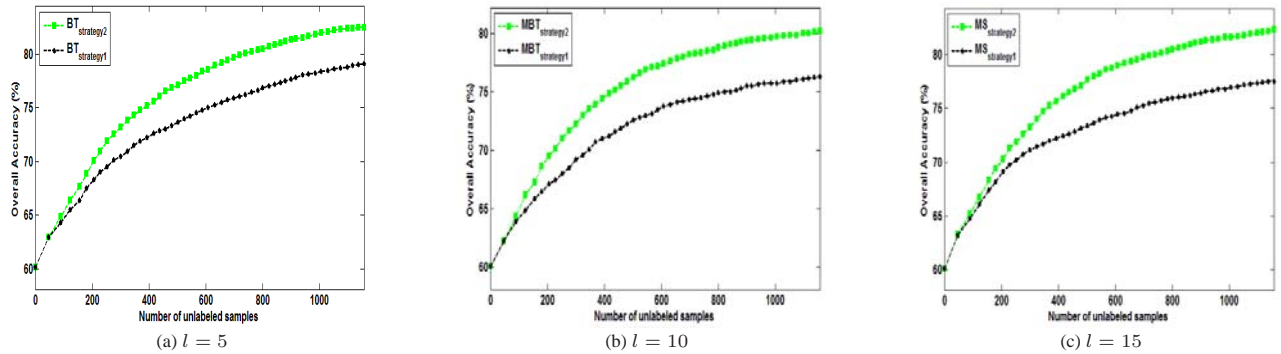


Fig. 1. Overall classification accuracies (as a function of the number of unlabeled samples) obtained for the AVIRIS Indian Pines data set using the proposed SSL approach with the two considered AL strategies.

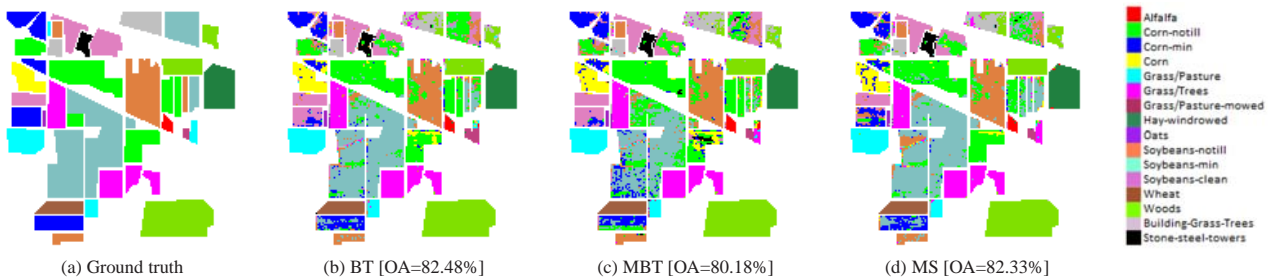


Fig. 2. Classification maps and overall classification accuracies (in the parentheses) obtained after applying the MLR classifier to the AVIRIS Indian Pines data set using the proposed AL framework based on strategy 2 (in all cases, $l = 10$).