SEMI-SUPERVISED CO-TRAINING AND ACTIVE LEARNING FRAMEWORK FOR HYPERSPECTRAL IMAGE CLASSIFICATION

Sathishkumar Samiappan and Robert J. Moorhead II

Geosystems Research Institute, Mississippi State University, Mississippi State, MS 39762

ABSTRACT

Hyperspectral imaging enables detailed ground cover classification with hundreds of spectral bands at each pixel. Rich spectral information can be a drawback since supervised classification of hyperspectral image requires a balance between number of training samples and its dimension. Achieving this balance requires large number of training or ground truth samples, which is generally difficult, expensive and time-consuming. This led researchers to explore the use of semi-supervised learning techniques where new training samples (unlabeled) are obtained from small set of available labeled samples without significant effort. In this paper, we propose a semi-supervised approach which adapts active learning to a co-training framework in which the algorithm automatically selects new training samples from abundant unlabeled pixels. Efficacy of the proposed approach is validated using probabilistic support vector machine classifier. Our experimental results with Indian pines hyperspectral image collected by the National Aeronautics and Space Administration Jet Propulsion Laboratory's Airborne Visible-Infrared Imaging Spectrometer indicate that the use of this co-training based approach represents promising strategy in the context of hyperspectral image classification.

***Index Terms*—** Semi-supervised learning, Active Learning, Co-training, Support vector machines

1. INTRODUCTION

Hyperspectral imagery (HSI) contains detailed spectral information in each pixel. The narrow bands in HSI over a wide range of wavelengths can provide excellent discrimination capability for subtly different classes compared to other remote sensing images. In a supervised classification scenario, a major challenge with HSI is the availability of limited ground truth (labeled sample) information. Collection of ground truth usually involves field trips which is time consuming and expensive. To alleviate this problem semi-supervised learning (SSL) techniques are adopted for HSI classification [1][2][3]. The SSL purports to use the available limited labeled data to enlarge the training set by using unlabeled samples. A SSL system trains a classifier on both the labeled and unlabeled data in such a way that it is better than supervised classifier trained on the labeled data alone [4]. For SSL to be effective, the unlabeled samples should be generated with significantly lower cost and the number of unlabeled samples should not be overwhelming to an extent that it adversely increase the computational complexity.

The quality of unlabeled samples determines the overall performance of SSL system. These samples need to be non-redundant with initial set of labeled samples and if not chosen carefully these unlabeled samples could deteriorate the performance of the classifier. To avoid the aforementioned issues, highly informative unlabeled samples need to be identified ensuring the performance of the classifier improves with enlarged training set. In supervised classification framework, active learning (AL) techniques are popularly used to select the highly informative samples for labeling [5]–[7]. In contrast with the random selection or random sampling (RS) of training sets, AL algorithm intelligently selects most useful unlabeled samples with the goal of achieving better classification performance with a smaller training set. The selected samples are usually labeled by a human expert, in [1] and the proposed approach this step is completely eliminated with SSL algorithm. In [1], the human expert is replaced with a self-learning SSL classifier and in the proposed approach, a Co-training classifier is used. In another approach [8], a transductive Support Vector Machine (TSVM) is used for semi-supervised classification of remote sensing images.

2. PROPOSED METHODOLOGY

In this work, we explored the combination of Co-training CT and AL for classification of HSI. CT is a complex generative SSL model originally proposed in [9] for classification of web pages. CT assumes the existence of two separate views of the input feature which are conditionally independent. In this paper, a) we define the concept of two different views for HSI by using a band grouping strategy and show that two views are conditionally independent, b) implement a CT and AL based SSL system for HSI classification and c) Study the performance of proposed framework for agricultural land cover classification problem by using support vector machine classifiers.

Let $x \equiv (x_1, x_2, \ldots x_n)$ be a HSI with n pixels such that $x \in \mathbb{R}^{d \times n}$ where d number of spectral bands in each pixel vector, $\mathcal{D}_l \equiv \{x_l, y_l\}$ be set of pixels with l known class labels, $\mathcal{D}_u \equiv \{x_u\}$ be u unlabeled pixels of HSI ($l \ll n$).

2.1 Co-training

CT has a resemblance to the one proposed in [1] where a classifier uses its confident predictions on unlabeled pixels to train itself. The critical difference with CT is usage of two classifiers and they teach each other. This is achieved by making three following underlying assumptions a) the training data has two views, $x = [x^1, x^2]$, b) each view alone is good enough to predict the class labels and c) the two views are conditionally independent. To satisfy the first assumption, HSI spectrum can be partitioned into two by using an intelligent band grouping strategy. To achieve the second assumption, the proposed implementation uses a spectral band grouping proposed in [10] to perform the partitioning. HSI with narrow spectral spacing exhibits a high degree of correlation between successive bands. Band grouping splits the spectrum at a point where there is minimum correlation between successive contiguous bands. In [10], the classifiers built on independent spectral band groups to decide the final class label is proven to be very efficient. So it is reasonable to assume that each band group is in itself enough to make a good classification with adequate labeled pixels. The third assumption calls for conditional independence of two spectral band groups. This ensures that the newly inducted samples of one classifier are dissimilar to the ones inducted by the second classifier making it more informative for the other classifier. With the two independent classifiers ($\mathcal{C}_1$ and $\mathcal{C}_2$) operating with different views of spectral bands they are conditionally independent. Since CT is a wrapper approach any learning algorithm could be employed for two classifiers. A probabilistic SVM classifier is used in this work. CT algorithm implemented in this work is elaborated in Figure 1 (a). Although the entire training set is presented to both the classifiers, $\mathcal{C}_1$ only pays attention to x^1 and completely ignores x^2. The classifier $\mathcal{C}_2$ is other way around.

2.2 Active Learning

The samples selected from the unlabeled set $\mathcal{D}_u$ for SSL need to correctly represent the class boundaries. AL achieves this by sampling discriminative pixels from $\mathcal{D}_u$. Suitability of AL to accurately select the pixels for supervised remote sensing image classification is demonstrated in [7]. The focus of this research is to study the interaction between CT proposed in section 2.1 and the AL discriminative sampling schemes. General AL algorithm implemented with CT is shown in Figure 1(b). The proposed implementation uses Margin Sampling (MS) heuristic [7] to select discriminative samples. The MS algorithm samples [11] the pixels that are lying within the margin of one-against all SVM hyper plane. It takes the advantage of geometrical properties of SVM.

Algorithm 1: Co-training

Inputs

- Labeled data $\mathcal{D}_l$
- Unlabeled data set from AL algorithm $\mathcal{D}_{uAL}$

1: **initialize**
    Call band grouping algorithm
    Split $\mathcal{D}_l$ into two views based on band grouping
    Training data for two classifiers $\mathcal{C}_1$& $\mathcal{C}_2$, $\mathcal{L}_1$=$\mathcal{L}_2$=$\mathcal{D}_l$

2: **repeat**
3: Train $\mathcal{C}_1$ from $\mathcal{L}_1$ and $\mathcal{C}_2$ from $\mathcal{L}_2$
4: Classify $\mathcal{D}_{uAL}$ with $\mathcal{C}_1$ and $\mathcal{C}_2$ separately
5: Add $\mathcal{C}_1$'s predictions to $\mathcal{L}_1$.
6: Add $\mathcal{C}_2$'s predictions to $\mathcal{L}_2$.
7: Remove $\mathcal{D}_{uAL}$ from $\mathcal{D}_u$

Fig. 1. (a): Co-training algorithm

Algorithm 2: Active Learning

Inputs

- Labeled data $\mathcal{D}_l$
- Unlabeled data $\mathcal{D}_u$
- Learning speed k

1: **repeat**
    Train model with current $\mathcal{D}_l$
2: **for** each sample in $\mathcal{D}_u$ do
        Evaluate AL heuristic MS or RS
    end for
3: rank the samples in $\mathcal{D}_u$ by its heuristic score
4: select k top ranked pixels into $\mathcal{D}_{uAL}$
5: Call Co-training - CT algorithm assigns the label
6: add the assigned labels to current training set $\mathcal{D}_l$
7: **until** adding sufficient number of training samples

Fig. 1. (b): Active Learning algorithm

3. EXPERIMENTAL STUDY

The HSI dataset used in this study was acquired using National Aeronautics and Space Administration Jet Propulsion Laboratory's Airborne Visible-Infrared Imaging Spectrometer collected over the Northwest Indiana's Indian Pine test site in June 1992. The image represents a vegetation-classification scenario with 145x145 pixels and 220 bands in the 400 to 2450nm region of the visible and infrared spectrum. This dataset has 16 classes, spectral resolution of 10nm and spatial resolution of 20m per pixel. Fig. 2(a) shows the pseudo colored RGB image of the Indian pines scene, Fig. 2(b) shows the ground-truth map and Fig. 2(c) shows the 16 mutually exclusive classes. This data provides a challenging classification problem due to the presence of unbalanced number of ground-truth pixels for each class.

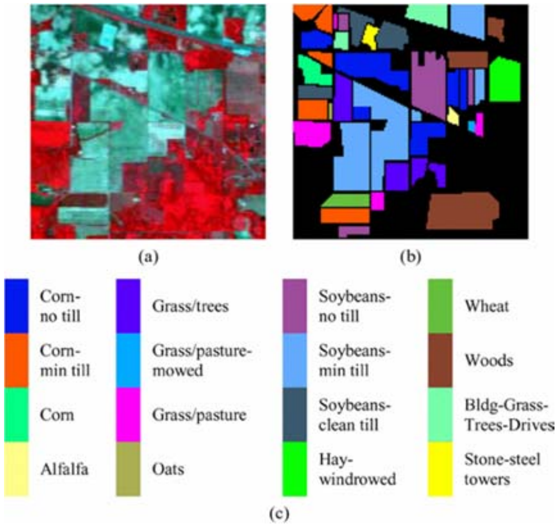

Fig. 2. (a) Indian pines pseudo colored RGB image (b) Indian pines ground truth (c) Indian pines class labels

4. RESULS

In the first experiment, the efficacy of proposed approach is evaluated with 5% training data (labeled) from each of 16 classes. Particularly, the performance of MS, a large-margin based AL heuristic is compared with RS. Since there is an imbalance in number of available training samples between classes, the overall accuracy (OA) of supervised classification with this training set is about 40%. For comparison, the OA of 7% training data is 62%. In our experiments, labeled training data is randomly selected from the original Indian pines ground truth. Fig.3 shows the accuracy as a function of number of iterations it takes to achieve this 62% with learning parameter k = 20. These samples are automatically selected by the algorithm without human supervision which is a key aspect. This plot reveals the advantages of using an active learning heuristic instead of random sampling.

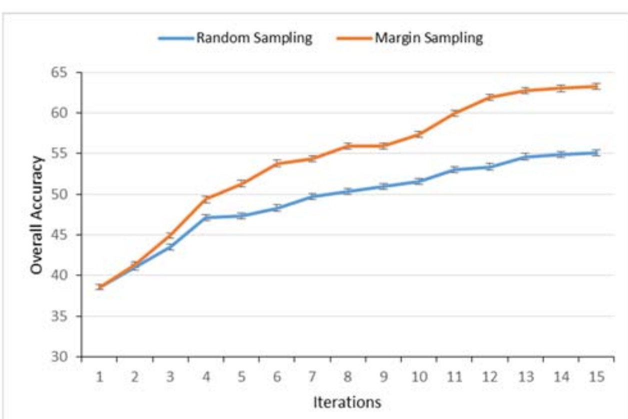

Fig 3 Overall classification accuracies as a function of iteration with 5% training samples & k = 20

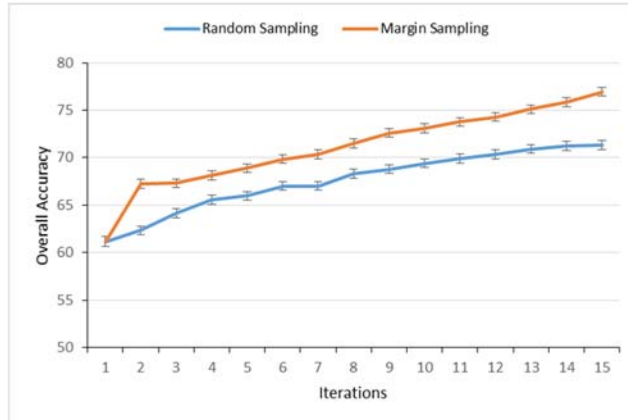

Fig 4 Overall classification accuracies as a function of iteration with 7% training samples & k = 20

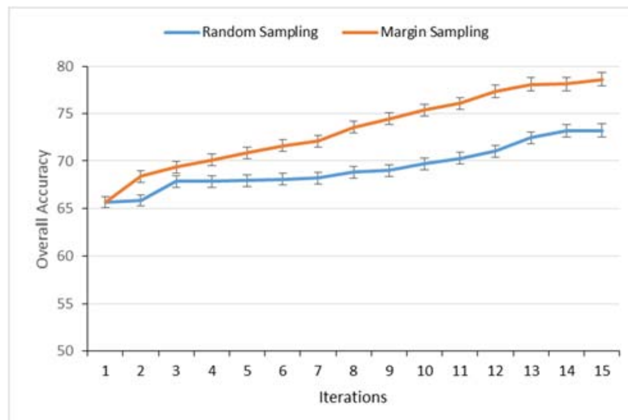

Fig 5 Overall classification accuracies as a function of iteration with 10% training samples & k = 20

MS achieves faster convergence to the upper bound represented by 7% training accuracy where MS performs poorly after adding same number of samples to the original training set.

Fig. 4 and Fig. 5 shows the performance when evaluated with 7% and 10% labeled data respectively. With 10% training, supervised SVM classifier yields OA of 67% and 75% with 15% training. It can be observed from the plots that these bounds are achieved with fewer iterations. It can be seen that MS performs consistently better than RS and the convergence takes fewer iterations to achieve the upper bound.

5. CONCLUSION AND FUTURE WORK

In this paper, we demonstrated a new approach for SSL for HSI classification using CT. Specifically, we studied the performance of using MS heuristic and compared it with RS. The results with respect to OA, kappa statistic and convergence to the upper bound is promising. Our experimental results indicate that this approach can greatly improve the accuracy of supervised classification by inducting unlabeled samples without human effort. In future work, we are planning to study more AL heuristics along with different classifiers. The current experiments focused on studying the performance with respect to OAs, studying class accuracies and their improvements will provide a better insight of the approach. The initial labeled data plays a major role in convergence and performance of the algorithm so, incorporating some form of spatial information to select these initial labels will be an interesting study. Results showed in this work uses the original hyperspectral bands without any feature selection and feature generation. Other recent studies with HSI - SVM demonstrated improvement in class accuracies and OAs with some form of feature selection before classification. So we are considering the use of feature selection methods under the proposed framework. The data views required by CT algorithm is arrived by using a band grouping strategy which considers HSI bands in contiguous manner. However, other non-contiguous band techniques such as spectral band clustering is worth exploring. Finally, another important research deserving future attention is usage of a multi-classifier system instead of using a single classifier. This may allow more accurate predictions and that can lead to improved overall performance of the proposed approach.

6. ACKNOWLEDGEMENT

The authors would like to thank D. Landgrebe and L. Biehl for making AVIRIS Indian pines data available to the community and D. Tuia, M. Volpi and Munoz-Mari for making AL Matlab toolbox openly available.

7. REFERENCES

[1] I. Dopido, J. Li, P. R. Marpu, A. Plaza, J. M. Bioucas Dias, and J. A. Benediktsson, "Semisupervised Self-Learning for Hyperspectral Image Classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 51, no. 7. pp. 4032–4044, 2013.

[2] G. Camps-Valls, T. Bandos Marsheva, and D. Zhou, "Semi-Supervised Graph-Based Hyperspectral Image Classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 45, no. 10. pp. 3044–3054, 2007.

[3] D. Tuia and G. Camps-Valls, "Semisupervised Remote Sensing Image Classification with Cluster Kernels," IEEE Geoscience and Remote Sensing Letters, vol. 6, no. 2. pp. 224–228, 2009.

[4] X. Zhu, "Semi-supervised learning literature survey," Tech Report University of Wisconsin-Madison, WI, 2008.

[5] 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, no. 4. pp. 1231–1242, 2008.

[6] W. Di and M. M. Crawford, "Active Learning via Multi-View and Local Proximity Co-Regularization for Hyperspectral Image Classification," IEEE Journal of Selected Topics in Signal Processing, vol. 5, no. 3. pp. 618–628, 2011.

[7] D. Tuia, M. Volpi, L. Copa, M. Kanevski, and J. Munoz-Mari, "A Survey of Active Learning Algorithms for Supervised Remote Sensing Image Classification," IEEE Journal of Selected Topics in Signal Processing, vol. 5, no. 3. pp. 606–617, 2011.

[8] L. Bruzzone, M. Chi, and M. Marconcini, "A Novel Transductive SVM for Semisupervised Classification of Remote-Sensing Images," IEEE Transactions on Geoscience and Remote Sensing, vol. 44, no. 11. pp. 3363–3373, 2006.

[9] A. Blum and T. Mitchell, "Combining Labeled and Unlabeled Data with Co-training," in Proceedings of the Eleventh Annual Conference on Computational Learning Theory, 1998, pp. 92–100.

[10] S. Prasad and L. M. Bruce, "Decision Fusion with Confidence-Based Weight Assignment for Hyperspectral Target Recognition," IEEE Transactions on Geoscience and Remote Sensing, vol. 46, no. 5, pp. 1448–1456, 2008.

[11] M. Balcan, A.Broder and T.Zhang, "Margin based active learning" in Proceedings of the 20th Annual Conference on Learning Theory, 2007, pp35-50.