SPATIALLY ADAPTED MANIFOLD LEARNING FOR CLASSIFICATION OF HYPERSPECTRAL IMAGERY WITH INSUFFICIENT LABELED DATA

Wonkook Kim and Melba M. Crawford

Laboratory for Applications
of Remote Sensing,
Purdue University
West Lafayette, IN 47906

Joydeep Ghosh

Intelligent Data Exploration
and Analysis Laboratory,
University of Texas at Austin
Austin, TX 78712

ABSTRACT

A classifier derived from labeled samples acquired over an extended area may not perform well for a specific sub-region if the spectral signatures of classes vary across the image. However, characterizing the local effects are an ill-posed problem, particularly for hyperspectral data, since an adequate number of labeled samples is not typically available for every location. This problem is addressed using semi-supervised learning and manifold learning, which both exploit the information provided by unlabeled samples in the image. A spatially adaptive classification method that uses Laplacian regularization is proposed, with the updating scheme using a combination of labeled and unlabeled samples.

Index Terms— spatially adaptive, classification, hyperspectral, Laplacian regularization, SVM

1. INTRODUCTION

Remote sensing image data have spatial context within a 2-dimensional array, which is different from classical machine learning problems, where data typically have no spatial relationships and are assumed to be statistically independent. Local effects such as sun angle variation, bidirectional effects and changes in soil condition can cause the spectral distribution of samples of a given class to vary across the image. Unfortunately, attempting to accommodate these local effects in supervised classification is problematic as training samples are often limited in number and not well distributed across the image, and some classes may not occur in a given local region. The problem of limited sample size is critical for successful classification of hyperspectral data because of the well known "curse of dimensionality" [1], which is exacerbated if the probability distributions of the class signatures vary across the image.

Manifold learning can mitigate these problems by discovering the inherent lower dimensionality from the input space. Recent results also indicate that classification of hyperspectral data may be more robust on the manifold space, particularly when training data are limited and spectral signatures are nonstationary across the image. Under the assumption that spatial variation in signatures affects the spatially localized structure of the manifold, classification on a spatially localized manifold should yield improved classification results.

Even though the manifold coordinates can be used directly for classification, kernel based regression methods are widely used for the inductive formulation of the classifier. Recently, fully supervised classification was adapted to semi-supervised learning to cope with the problem of insufficient labeled samples in transductive SVM [2] and semi-supervised SVM [3], and investigated for remote sensing applications [4, 5]. However, many kernel based semi-supervised methods are computationally intensive and are known to have problems with convergence to local optima. Belkin et al. regularized the classifier based on the geometry of unlabeled samples in the spectral domain [6]. Smoothness of a function is computed, and an irregular regression function is penalized so that the smoothly connected samples can be taken as a cluster. In the remote sensing literature, Gomez-Chova et al. recently investigated the Laplacian regularization approach for cloud screening and image classification [7, 8].

In this paper, a novel methodology is proposed to handle the spatial adaptation problem using the Laplacian regularization method. A global classifier that is constructed from the labeled samples is adapted to a local area using the unlabeled samples in the area, as well as any locally available labeled samples. While the initial labeled samples are unchanged in most semi-supervised algorithms, here the labeled samples are incrementally updated with randomly selected unlabeled samples from the local region. Since the regularization method involves unlabeled samples in the training stage, importing appropriate unlabeled samples to the binary classification framework of the Laplacian regularization method is an issue for multi-class classification problems. The selection scheme is investigated here for both one-against-one and one-against-all strategies.

2. LAPLACIAN REGULARIZATION

The Laplace-Beltrami operator on a manifold provides a measure of the geometric smoothness of a function on the manifold. In the discrete case, the operator is approximated by a graph Laplacian. Once a kernel-based regression function for any two classes is assumed, the best regression function that fits the data is calculated by minimizing the functional

$$f^* = \underset{f \in H_K}{\operatorname{argmin}} \frac{1}{l} \sum_{i=1}^{l} V(x_i, y_i, f) + \gamma_A \|f\|_K^2 + \gamma_I \|f\|_I^2 ,$$

where V is a loss function such as squared loss or the hinge loss function as in Support Vector Machines (SVMs).

The second term is the norm of the function in the corresponding Reproducing Kernel Hilbert Space (RKHS) [9] for a Mercer kernel K, and the last term measures the geometric smoothness of the intrinsic manifold structure through a graph Laplacian. Penalizing the last two terms controls the smoothness of the regression function in ambient space and in intrinsic space, respectively. According to the Representer Theorem [10], the solution to this minimization is given as

$$f^*(x) = \sum_{i=1}^{l} \alpha_i K(x_i, x),$$

and the problem is reduced to finding optimal coefficients of the kernel function.

3. SPATIAL ADAPTATION

Adaptation of the classifier is accomplished by updating the existing labeled and unlabeled samples with local samples which better represent the local distribution of samples. Although the classifier is constrained by the smoothness factor and seeks to find the class boundaries which are not smooth, the boundaries between classes of remote sensing data are often difficult to differentiate only by clustering. Therefore, the adaptation procedure gradually modifies the group of labeled and unlabeled sets.

When data sets X_1 and X_2 are from two different distributions, the problem is to classify X_2 based on X_1 and its label set, Y_1. Each sample of the data, (X, Y), can be expressed as $\{x_j, y_j\}$ where $x_j \in \Re^N$ and $y_j \in \{0, 1, ..., N_{class}\}$. The classification is performed iteratively by updating the labeled set, which is set initially as $L^{(0)} (= Y_1)$. The current classifier, $C^{(i)}$ is applied to X_2 and generates the classification result, $Y_2^{(i)}$. The current labeled set, $L^{(i)}$ is then updated in such a way that $N_{exchange}$ samples are randomly selected from X_2 based on $Y_2^{(i)}$, and those samples replace the same number of samples in the labeled set, $L^{(i)}$, to produce the updated labeled set, $L^{(i+1)}$. The subsequent classifier, $C^{(i+1)}$, is developed based on both the labeled and unlabeled samples from this second set of data. Although some anomalous samples might be chosen for labeled samples, the smoothness term in the regularization formulation forces the labeled sample set to converge to the true values of the labels. The training phase is terminated when the stopping criterion is satisfied. The iteration stops when the number of labels being changed is less than threshold defined in terms of the ratio of the number being changed to the total number of pixels .

This binary classifier can be implemented for multi-class problems using one-against one (OAO) and one-against-all (OAA) strategies. The OAO strategy is known to maintain a more balanced and robust result compared to the OAA method and is more conducive to incorporation of the smoothness measure. However, in the semi-supervised setting, the unlabeled samples for two specific classes required by the OAO cannot be identified directly. In OAA, the metaclass which is comprised of multiple classes has more discontinuity than the single class branch of the classifier, which can cause semi-supervised learning to converge to local optima, particularly when the number of labeled samples near the boundary is small. Both the modified OAO scheme and the OAA method are investigated and compared in the current study.

With the proposed adaptation scheme, the global classifier is adapted to each local region which is a subset of the whole image. The adaptation can be applied recursively into smaller regions, as long as the number of pixels is adequate to construct the manifold structure. After updating, the locally classified image patches are combined to produce a global output.

4. EXPERIMENTS

Hyperion data acquired by the NASA EO-1 satellite over the Okavango Delta, Botswana in May, 2001 are used for this experiment. The image dimension for this experiment is 1476 x 256, and the land cover is classified into 9 classes with 145 bands.

The labeled samples are usually collected in spatially contiguous patches over sites, often resulting in highly correlated samples which can result in poor generalization of the classification results. For the experiments reported here, training and test samples are extracted randomly from spatially disjoint patches scattered throughout the scene. The training rate is varied from 25% to 75% of the candidate training samples to investigate the impact of training sample size.

First, the classifier is developed and applied to the whole image using all the labeled samples in the scene to determine the base performance of the global classifier. K-nearest neighbors (KNN), the Laplacian method with OAA and the Laplacian method with OAO are all investigated for the global classifier. Since the manifold is constructed with only labeled samples, each class is relatively well separated from other classes compared to the case when unlabeled samples are introduced to the classifier. However, because the labeled sam-

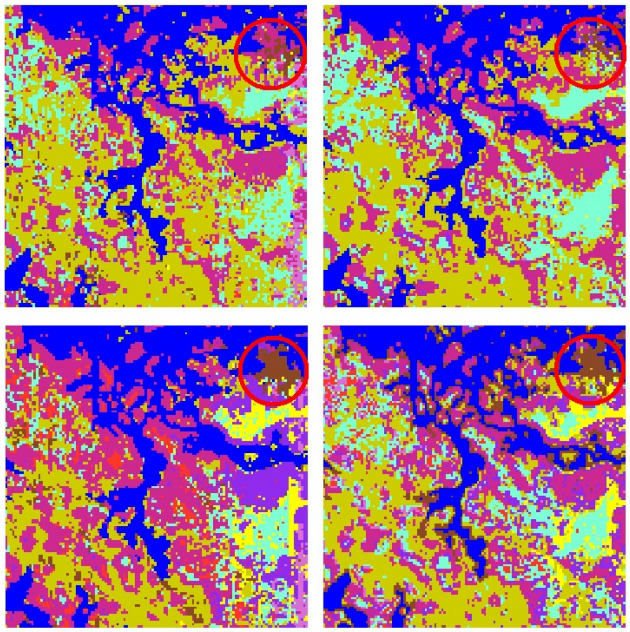

Fig. 1. Classification results on a subset image from (top left) global-OAA, (top right) global-OAO, (bottom left) adaptive-OAA and (bottom right) adaptive-OAO. The red circled area (originally fire scar) is classified into several other classes in the upper images. True labels are recovered by adapting the classifiers in the lower images.

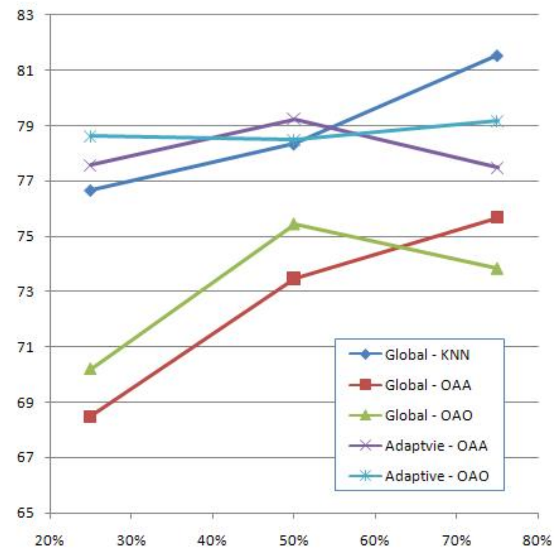

Fig. 2. Classification accuracies for Botswana data over varying sampling rate

ples are not well distributed across the scene, the performance of the global classifier is not guaranteed for local regions.

Through the spatial adaptation, the global classifier is adjusted to better reflect the distributions of the local samples. In this experiment, the entire region is divided into four subsections, each of dimension 369 x 256. Since the number of available unlabeled samples is so large compared to the number of labeled samples, only the randomly selected samples are used for the iterative training procedure. In case of the OAA strategy, any unlabeled samples can be used for training the N_{class} classifiers since the output space of the classifier exhausts all the classes. For the OAO scheme, only the unlabeled samples that are related to the specific two classes should be introduced to the classifier. However, it is impossible to identify such samples precisely, as this requires advanced knowledge of the labels of unlabeled samples. Here, samples that are classified during the initial global classification into the two respective classes are selected for the unlabeled samples to be imported to the binary classifier. The final label of a pixel is determined by voting after pairwise classification is completed.

The KNN global classifier yielded higher accuracies than either the OAA or OAO binary strategies implemented globally in the Laplacian framework, with performance of the Laplacian-based method deteriorating rapidly at lower sampling rates. For the localized regional implementation, the adaptive classification methods performed nearly as well as the global KNN. As the sampling rates decreased, the OAO strategy eventually yielded the highest accuracies, indicating that the semi-supervised learning compensated for the small number of training samples through exploitation of local information. On a class basis, the localized Laplacian method yielded the greatest improvement for the primary floodplain and fire scar classes, which the KNN method often confused. An example of improvement is illustrated in Fig.2. However, the proposed method yielded poor results on pairs of classes whose spectra are highly overlapped such that the natural boundary is hard to identify, e.g. riparian vs. woodlands which have similar vegetation components. The smoothness regularization may actually impact the results negatively for these cases.

5. CONCLUSIONS

A new spatially adaptive classification method which exploits unlabeled samples when proper labeled samples are not available was developed. Although spatial drift in spectral signatures is often difficult to determine in multi-class problems, it becomes critical to consider the bias when the number of labeled samples is limited and not well distributed over the image. The novelty of this paper is the ability to handle changes in the distribution of class spectra when labeled samples are regionally localized and limited in number.

Future work will include application of the proposed method to hyperspectral data from alternative locations and

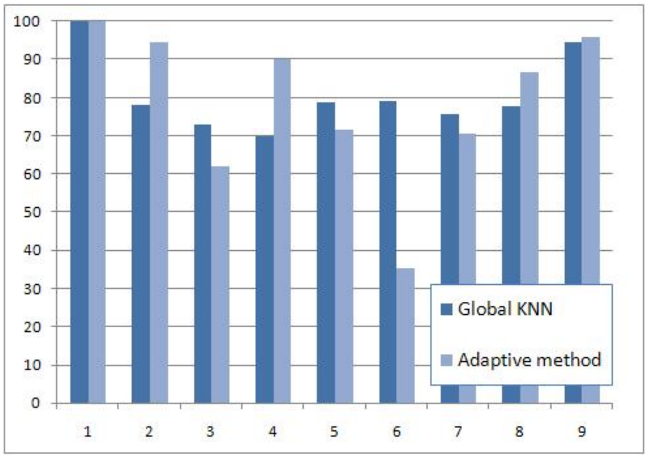

Fig. 3. Class dependent classification accuracies of the adaptive method are presented and compared to KNN results. The accuracies of adaptive methods are averaged from the two multi-class scheme, OAA and OAO.

	Class	Number of Samples
1	Water	158
2	Primary Floodplain	228
3	Riparian	237
4	Firescar	178
5	Island Interior	183
6	Woodlands	199
7	Savanna	162
8	Short Mopane	124
9	Exposed Soils	111

Table 1. Land cover types in the Botswana scene and corresponding numbers of samples per class.

higher resolution sensors. More promising results are expected for the high resolution images since Laplacian smoothness regularization is more beneficial for data with few mixed samples. Also the adaptation approach will be applied to lower levels of the quadtree hierarchy which represent smaller areas. The use of alternative binary decomposition methods, such as the Binary Hierarchical Classifier (BHC) [11], will be also investigated.

Acknowledgments

This research was supported by NSF (Grant IIS-0705836) and by Purdue University.

6. REFERENCES

[1] G. F. Hughes, "On the mean accuracy of statistical pattern recognisers," *IEEE Transactions on Information Theory*, vol. 1, pp. 55–63, 1968.

[2] V. N. Vapnik, *Statistical Learning Theory*, Wiley-Interscience, September 1998.

[3] K. Bennett and A. Demiriz, "Semi-supervised support vector machines," *Advances in Neural Information Processing Systems*, vol. 11, pp. 368–374, 1998.

[4] 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, November 2006.

[5] G. Camps-Valls, T.V. 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, October 2007.

[6] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, June 2003.

[7] L. L. Gomez-Chova, G. G. Camps-Valls, J. J. Munoz-Mari, and J. J. Calpe, "Semi-supervised cloud screening with Laplacian SVM," in *Proc. 2007 International Geoscience and Remote Sensing Symposium*, Barcelona, Spain, Jul. 23-27 2007, pp. 1521–1524.

[8] L. L. Gomez-Chova, G. G. Camps-Valls, J. J. Munoz-Mari, and J. J. Calpe, "Semi-supervised image classification with Laplacian support vector machines," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 3, pp. 336–340, 2008.

[9] N. Aronszajn, "Theory of reproducing kernels," *Transactions of the American Mathematical Society*, vol. 68, 1950.

[10] G. Wahba, *Spline models for observational data*, SIAM, Philadelphia, 1990.

[11] S. Rajan, J. Ghosh, and M. M. Crawford, "Exploiting class hierarchies for knowledge transfer in hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 11, pp. 3408 – 3417, 2006.