SPATIALLY ADAPTIVE CLASSIFICATION OF HYPERSPECTRAL DATA WITH GAUSSIAN PROCESSES

Goo Jun *Joydeep Ghosh*

Department of Electrical and Computer Engineering
The University of Texas at Austin, Austin TX 78712, USA
{gjun, ghosh}@ece.utexas.edu

1. INTRODUCTION

Automated classification of land cover types based on hyperspectral imagery often involves a large geographical area, but class labels are available for only small portions of the entire area. Moreover, the spectral signature of the same land cover class may vary substantially over different locations. When a classifier is trained on a specific geographical location and applied to other areas, it often performs poorly because of such spatial variation of spectral signatures. In this paper, we propose a novel framework for classification of hyperspectral data: a Gaussian-Process Maximum-Likelihood (GP-ML) model where the mean of each spectral band is spatially modeled using a Gaussian process. Our framework provides a practical and effective way to model spatial variations of high dimensional data such as hyperspectral images for classification problems.

2. RELATED WORKS

There have been algorithms developed for hyperspectral data that are intended for spatially distant datasets, or that use spatial information. For example, Rajan *et al* [1] provided a framework to transfer knowledge between different spatial and temporal locations, but this approach does not utilize spatial relations between locations. Another approach is to add information from homogeneous neighborhoods as in [2], but it does not model varying spectral signatures of hyperspectral data directly. The closest approach to this paper is by Goovaerts[3], where the prior probability of the i-th class, $P_i(\mathbf{s})$, is modeled by indicator kriging. Gaussian process has been long known in spatial statistics as *kriging* [4], but kriging has been considered to be only suitable for modeling of single or small number of target variables. In this paper, we directly model spatially adaptive class-conditional distributions for each each band..

3. METHODS

Let $\mathbf{x} = (x_1, x_2, ..., x_d)^T$ be a d-dimensional vector representing spectral bands of a pixel in a hyperspectral image, and $y \in \{y_1, y_2, ..., y_c\}$ be a class label that indicates land cover type. The class-conditional probability distribution $p(\mathbf{x}|y_i)$ is usually assumed to be multivariate Gaussian: $p(\mathbf{x}|y_i) \sim N(\boldsymbol{\mu_i}, \Sigma_i)$, where $\boldsymbol{\mu_i}$ is the mean vector and Σ_i is the covariance matrix of the i-th class. For simple notation, lets focus for now on a single class and omit i. Typically, both $\boldsymbol{\mu}$ and Σ are considered to be constant over the entire image. Instead we model $\mathbf{x}(\mathbf{s})$ as a random process indexed by a spatial coordinate $\mathbf{s} = (s_1, s_2)$ with a mean function $\boldsymbol{\mu}(\mathbf{s}) = (\mu_1(\mathbf{s}), \mu_2(\mathbf{s}), ..., \mu_d(\mathbf{s}))^T$ and a spatial covariance function $K_j(\mathbf{s}_m, \mathbf{s}_n)$ for band j according to the Gaussian process model. We assume each band is spatially independent of each other, hence neglecting cross-correlation of $x_j(\mathbf{s}_m)$ and $x_k(\mathbf{s}_n)$ for $j \neq k$ and $\mathbf{s}_m \neq \mathbf{s}_n$. This assumption implies that Σ is assumed to be constant without spatial variation. Modeling multiple correlated target variables has been studied in spatial statistics, and it is called *cokriging* [4]. It is impractical and too demanding, however, to model hyperspectral data directly by cokriging [3], since cokriging requires solving $(n+1) \cdot d$ linear equations for n data points with d dimensions, and it makes the matrix so big that the system becomes sensitive to noise and inaccurate parameters. The popular squared exponential covariance function is employed [5]. The covariance function is assumed to be identical over all classes, and over all bands except for a scalar factor, *i.e.*,

$$K_j(\mathbf{s}_m, \mathbf{s}_n) = \sigma_j^2 \exp\left(-\frac{||\mathbf{s}_m - \mathbf{s}_n||^2}{2\lambda^2}\right) \quad \sigma_j^2 = var(x_j) = \Sigma(j,j), \quad j = 1, 2, ...d. \tag{1}$$

Assume now that we have a set of labeled data points from this class, $X = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m)$, located at corresponding spatial coordinates $S = (\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_m)$, and let $\mathbf{x}_j$ be a vector consists of j-th bands in X, $\mathbf{x}_j = (x_{1j}, x_{2j}, ...x_{mj})^T$. Then the predicted distribution of the j-th band of a new data point $\mathbf{x}^*$ at coordinate s^* is Gaussian with the mean $\boldsymbol{\mu}^*$:

$$\mu_j^* = K_j(\mathbf{s}^*, S)[Cov(S,S) + \sigma_j^2 I]^{-1}\mathbf{x}_j \quad .$$

$p(\mathbf{x}(\mathbf{s})|y_i)$ is estimated by using the predicted $\boldsymbol{\mu}^*$ for each class to find y_i that maximizes the posterior probability $p(y_i|\mathbf{x}(\mathbf{s}))$.

This work was supported by NSF Grant IIS-0705815.

3.1. Fitting Hyperparameters

In the Gaussian process model, a covariance function determines the nature of the process, and the covariance function is characterized by hyperparameters. In (1), we have one hyperparameter, λ, which is the length parameter that determines how fast the correlation between two points decreases as the distance between the points increases. One way to find the best hyperparameter is to use cross-validation. A random sampling of training data to construct training and validation sets turned out to be inappropriate, however, since randomly sampled training and test data points are too close to each other. In this case, the obtained length parameter tends to be too big because there is not enough statistical difference between training and validation sets. The situation is opposite to a conventional cross-validation setup, where homogeneity between training and validation data is desirable. We divided the training data into spatially disjoint cross-validation sets, and searched for λ that provides highest classification accuracies. As a result, we set $\lambda = 530$.

3.2. Spatially Localized Priors

So far we have only considered spatial modeling of the mean vectors. In hyperspectral images, however, prior probability of each class also varies spatially[3]. We applied self-training [6], a method of semi-supervised learning, to estimated localized prior probabilities, since most of the map is unlabeled. The entire map is divided into equally sized tiles, and a classifier is applied to each tile with global priors that are estimated from the training set. For GP-ML classification, we estimate the mean vector for the center of the tile using the proposed algorithm and assume that this mean vector is common for all pixels in the tile. Classified results for each tile are then used to estimate localized prior probability distribution $P_i(\mathbf{s})$. Using the localized priors and class-conditional distributions, we run maximum-likelihood (ML) classifiers once more for each point of interest to obtain the final class label.

4. EXPERIMENTS AND RESULTS

A Hyperion hyperspectral image taken from Okavango Delta, Botswana in May 2001 is used for experiments. Each pixel contains 145 spectral bands. The size of the map is 1476 by 256 with a spatial resolution of 30m. We used two spatially disjoint class maps from the same geographical region, and there are 9 classes in total. Best-bases feature extraction algorithm [7] is used to aggregate highly correlated adjacent bands, which is beneficial for a spatially independent band model. The number of bands for best bases algorithm is 40, as determined by cross-validation. Fisher's feature extractor is also applied after best-bases extraction. The first class map is used as a training set, from which we obtain global model $p(\mathbf{x}|y_i)$ and spatially adaptive model $p(\mathbf{x}(\mathbf{s})|y_i)$, and the second map is used as a test set. For prior estimation, we used a 32×32 tile size. Table 1 shows classification accuracies from ML and GP-ML classifiers, before and after localized estimation of prior probabilities. Our baseline of 86.68% is the ML result without localized priors. As can be seen in the table, our algorithm with Gaussian process model and localized prior shows the best result, achieving 92.40% accuracy for a nine-class problem.

	ML	GP-ML	ML with $P_i(\mathbf{s})$	GP-ML with $P_i(\mathbf{s})$
Accuracy	86.68%	89.82%	89.54%	**92.40**%

Table 1. Classification results from ML and GP-ML algorithms, before and after spatially localized priors

5. CONCLUSION

We have proposed a novel method for classification of hyperspectral data with a spatially adaptive approach that models class-conditional distributions by Gaussian processes, and estimates spatially localized prior probabilities with semi-supervised learning. Preliminary experimental results show that the proposed method shows significant improvements over the baseline algorithm where no spatial information is considered.

6. REFERENCES

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

[2] Yangchi Chen, M.M. Crawford, and J. Ghosh, "Knowledge based stacking of hyperspectral data for land cover classification," *Computational Intelligence and Data Mining, 2007. CIDM 2007. IEEE Symposium on*, pp. 316–322, 1 2007-April 5 2007.

[3] P. Goovaerts, "Geostatistical incorporation of spatial coordinates into supervised classification of hyperspectral data," *Journal of Geographical Systems*, vol. 4, no. 1, pp. 99–111, 2002.

[4] N. Cressie, *Statistics for Spatial Data*, Wiley, New York, 1993.

[5] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, The MIT Press, 2005.

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

[7] S. Kumar, J. Ghosh, and M. M. Crawford, "Best-bases feature extraction algorithms for classification of hyperspectral data," *IEEE Trans. on Geosci. and Remote Sens.*, vol. 39, no. 7, pp. 1368–1379, 2001.