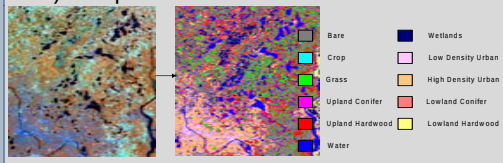
Land Use and Land Cover Classification

Ranga Raju Vatsavai, Budhendra L. Bhaduri,
Eddie Bright, and Nagendra Singh
GIST/CSED, Oak Ridge National Laboratory,
Oak Ridge, TN 37831, USA

Goo Jun and Joydeep Ghosh
Department of Electrical and Computer Engineering
University of Texas, Austin, USA

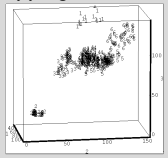
Problem

- Given a image, and set of class labels and training samples, objective is to classify (assign label) each pixel.



Challenges

1. Thematic Classes are Overlapping

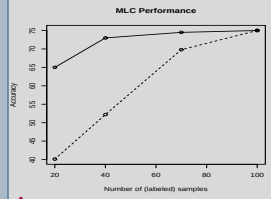


Example: Upland Classes vs. Lowland Classes

- Need ancillary data to separate such classes
- Not straight forward with statistical pattern recognition techniques, but ancillary data can be incorporated into Maximum Likelihood Classifier via 'a priori' term [1].

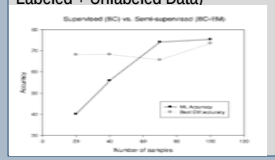
2. Small training Samples

- Number of training samples ~ (10-30) * Number of dimensions
  - Costly ~ $500-$800 per plot
  - Accessibility (difficult terrain, privacy, ...)
  - Real-time (Emergency situations)



MLC Performance as the size of training data increases
(Figure also shows variance in accuracy for different subsets of training samples from |TD| = 100)

- Recent studies [2] showed that including unlabeled samples may improve classification accuracy
- Semi-supervised Approach (Expectation Maximization using Labeled + Unlabeled Data)



Semi-supervised Performance (Fixed unlabeled, increased labeled).
Main advantage: Very few labeled plots.

Semi-supervised Learning

- Semi-supervised Learning
  - Gaussian Mixture Model (GMM)

$$p(x_i \mid \theta) = \sum_{j=1}^{M} \alpha_j p_j(x_i \mid \theta_j)$$

- Estimate GMM parameters using EM by posing unlabeled sample as missing data [3]
  - EM essentially consists of following two steps:
  - E-Step:

$$e_{ij} = \frac{\left|\hat{\Sigma}_j^k\right|^{-1/2} \exp\left\{-\frac{1}{2}\left(x_i - \hat{\mu}_j^k\right)^T \hat{\Sigma}_j^{-1,k}\left(x_i - \hat{\mu}_j^k\right)\right\}}{\sum_{l=1}^{M}\left|\hat{\Sigma}_l^k\right|^{-1/2} \exp\left\{-\frac{1}{2}\left(x_i - \hat{\mu}_l^k\right)^T \hat{\Sigma}_l^{-1,k}\left(x_i - \hat{\mu}_l^k\right)\right\}}$$

- M-Step

$$\alpha_j = \frac{\sum_{i=1}^{N} e_{ij}}{N}, \qquad \hat{\mu}_j^{k+1} = \frac{\sum_{i=1}^{N} e_{ij} x_i}{\sum_{i=1}^{N} e_{ij}},$$

$$\text{and} \quad \hat{\Sigma}_j^{k+1} = \frac{\sum_{i=1}^{N} e_{ij}\left(x_i - \hat{\mu}_j^{k+1}\right)\left(x_i - \hat{\mu}_j^{k+1}\right)^T}{\sum_{i=1}^{N} e_{ij}}$$

Challenge: How to get ancillary data into this framework

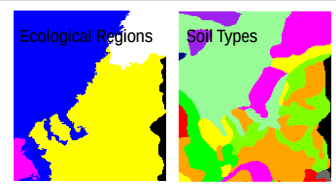
- SSL for Multisource Data
  - Mixture of Mixture Distributions

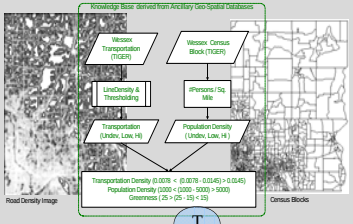
$$p(x_i \mid \theta) = \sum_{j=1}^{M} \alpha_j \prod_{l=1}^{2} p_{jl}(x_{il} \mid \theta_{jl})$$

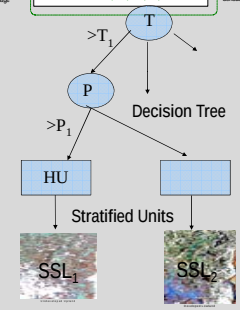
- Two models (l=2)
  - Continuous Distribution (for image data)
  - Discrete distribution (for categorical variables – ancillary geospatial data)

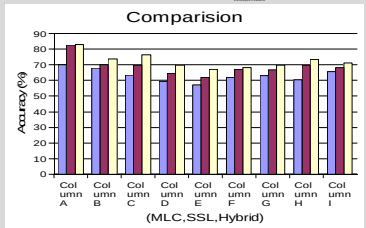
- For getting ancillary data into the semi-supervised framework, we used a knowledge based algorithm to stratify the image into three broad categories
  - Uplands
  - Lowlands
  - Developed

Multisource Semi-supervised Learning





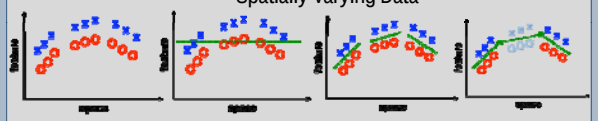




Gaussian Process Learning

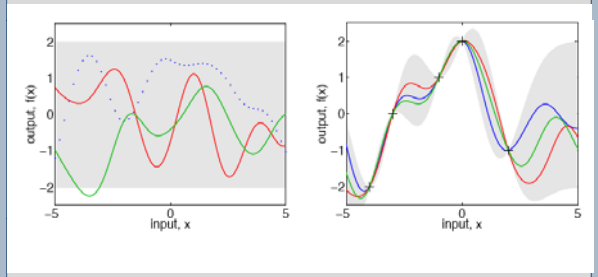
- Signature Extension is a problem when dealing with large geographic region classification
- In practice, it is generally not possible to have data from all possible locations
  - Train separate models for each location – needs lot of data
  - Train one model, use on all – poor performance
  - Train on some data, adapt for others – needs adaptive model

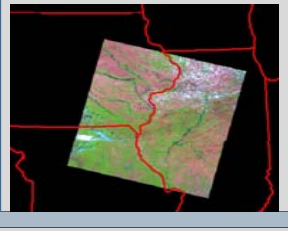
Spatially Varying Data

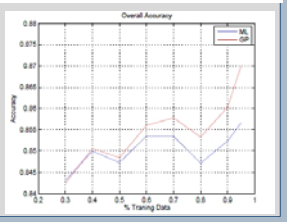


- Change of distribution over space is modeled by

$$p(x \mid y) \sim N(\mu, \Sigma) \qquad p(x(s) \mid y) \sim N(\mu(s), \Sigma(s))$$







References
[1] A.H. Strahler, "The use of prior probabilities in maximum likelihood classification of remote sensing data," Remote Sensing of Environment, vol. 10, pp. 135–163, 1980.
[2] Ranga Raju Vatsavai, Shashi Shekhar, and Thomas E. Burk, "A semi-supervised learning method for remote sensing data mining," in ICTAI, 2005, pp. 207–211.
[3] J. Bilmes. A gentle tutorial on the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. Technical Report, University of Berkeley, ICSI-TR-97-021, 1997., 1997.
[4] Goo Jun, Ranga Raju Vatsavai, and Joydeep Ghosh. Spatially Adaptive Classification and Active Learning of Multispectral Data with Gaussian Processes. IEEE ICDM International Workshop on Spatial and Spatiotemporal Data Mining, 2009.

Acknowledgements: Prepared by Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, Tennessee 37831-6285, managed by UTBattelle, LLC for the U. S. Department of Energy under contract no. DEAC05-00OR22725. Research was supported through LDRD program

Contact: Ranga Raju Vatsavai (vatsavairr@ornl.gov)