Final Report

Hyperspectral Data Classification

Jun Chen, Dan Wang, Zhou Zhang and Penghao Wang
Purdue University
chen1241@purdue.edu, wang1766@purdue.edu
zhan1553@purdue.edu, wang1128@purdue.edu

December 11, 2015

1 Background and Motivation

Conservation tillage management has been advocated for the purpose of soil preservation and sustainable crop production. Conservational tillage practice induces less surface disturbance and leaves more crop residues, which can decrease runoff rate, improve soil and water quality, and increase organic matter. The demand for the mapping of crop tillage practices has been brought up for precision agricultural management and appraisal. However, current methods for mapping massive crop tillage practices are mainly done with field investigations, which are labor costing, time consuming, subjective, and make it difficult to generate widely distributed survey data. Remote sensing technology provides a more rapid, accurate, and objective solution. Moreover, the vast data from remote sensing in agriculture require more efficient approaches in data analytics, including tillage mapping, in support of management decisions.

Recently, hyperspectral remote sensing has gained attention in the remote sensing application community. Hyperspectral imaging generates hundreds of images corresponding to different wavelength channels for the same area on the surface of the earth. A hyperspectral image is a 3-D cube of data with the width and length of the array corresponding to the spatial dimensions and the continuous spectrum of each point as the third dimension, which enables discrimination of materials based on their spectral characteristics. One of the most important applications of hyperspectral data is image classification, where pixels are labeled to one of the classes based on their spectral characteristics.

2 Related work

The overall objective of hyperspectral image classification procedures is to automatically categorize all pixels in image into land cover classes (Lu and Weng, 2007 [5]). Based on the spectral features, several classification algorithms have been used in this field.

- **Maximum likelihood classification.**
  Maximum likelihood decision rule is based on Gaussian estimate of the probability density function of each class. Maximum likelihood classifier evaluates both the variance and covariance of the spectral response patterns in classifying an unknown pixel.

- **Neural networks classifier.**
  Neural networks (Atli.J et al., 1995 [1]) have been applied successfully in various fields. Neural networks are networks which need a long training time but are relatively fast data classifier. For very high dimensional data, the training time of a neural network can be very long and the resulting neural network can be very complex. This leads to the importance of feature reduction mechanisms for neural networks.

- **Decision trees.**
  Decision tree classifier breaks a complex classification problem into multiple stages of simpler decision-making processes (Safavian and Landgrebe, 1991 [10]). Decision trees are trees that classify instances by sorting them based on feature values. Each node in a decision tree represents a feature in an instance to be classified, and each branch represents a value that the node can assume (Murthy,1998 [7]).

- **Support Vector Machine (SVM).**
  Specific attention has been dedicated to support vector machines for the classification of remotely sensed images recently (Hermes et al., 1999; Roli and Fumera, 2001; Hung et al., 2002 [4]). The interest in growing Support Vector Machines (Vapnik, 1998; Burges, 1998; http://www.kernal-Machines.org/tutorial.html [11]) is confirmed by their successful implementation in numerous other pattern recognition applications like biomedical applications (El-Naqa et al., 2002 [2]), image compression (Robinson and Kecman,2003 [9]), and three dimensional object recognition (Pontil and Verri, 1998 [8]).

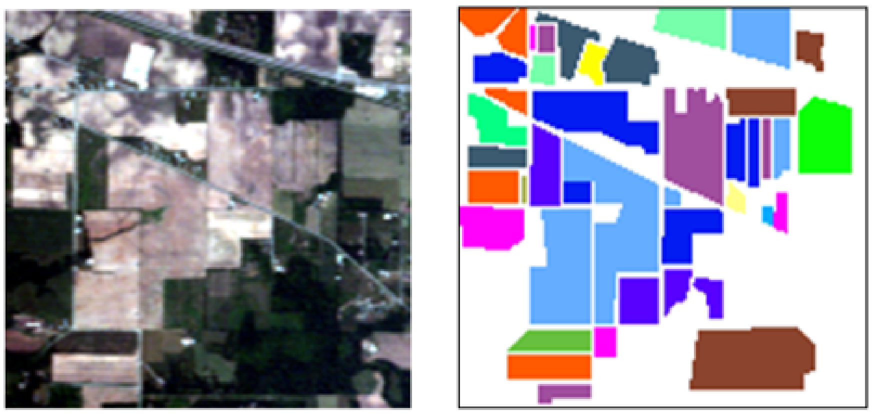
- **Knowledge based Classifiers.**
  Different kinds of ancillary data, such as digital elevation model, soil map, housing and temperature are readily available; they may be incorporated into a classification procedure in different ways.

- **Contextual Classifiers.**
  In contextual classifiers, the spatially neighboring pixel information is used. Contextual classifiers are developed to cope with the problem of intraclass spectral variations (Gong and Howarth, 1992 [3]). To improve the classification results, it exploits spatial information among neighboring pixels (Magnussen et al., 2004 [6]).

3 Dataset Description

The Indian Pines dataset was acquired by the NASA Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over the Indian Pines agricultural site in northwestern Indiana in June 1992, at 20-m spatial resolution and 10-nm spectral resolution over the range of 400-2500 nm.

There are 220 spectral bands in the image. Ground reference data were comprised of 10,366 labeled samples assigned to 16 classes. The crop-tillage related classes relate to the management practices used for tilling in a particular field. For example, Soybeans-no till refers to a field in which soybeans are planted, but no tillage was performed after the last harvest, thereby leaving a significant quantity of crop residue. Soybeans-min till refers to tillage which breaks up the soil, reducing the quantity of remaining residue, while Soybeans-clean till refers to fields that are fully ploughed after the previous harvest, leaving essentially no residue. Corn-tillage related classes are defined similarly. True-color composite and relative ground reference maps are shown in Fig. 1, while the details related to the number of labeled samples is listed in Fig. 2.



(a) (b)

Figure 1: Indian Pines dataset (a) True-color composite image; (b) Ground reference map

4 Problem Formulation and Data Preprocessing

The hyperspectral image classification task is perfectly suitable to be modeled as a machine learning problem. The problem is challenging because the feature space is high dimensional with limited number of labeled data. Furthermore, it is a multi-class classification problem. Each pixel of the hyperspectral image is a sample in this problem. For each pixel, it contains 220 spectral bands, so there are 220 features for each sample. To conduct the classification task, we did the following things to preprocess the data.

- Remove noisy spectral bands (features), and use 200 out of 220 as final features.
- Ignore minor classes with very labeled samples, and focus on 12 out of 16 classes.
- For each class, randomly select 100 labeled samples for training and 100 for testing.
- Normalize the data for each feature.

	Class name	# labeled samples
	C1: Corn-no till	1434
	C2: Corn-min till	834
	C3: Corn	234
	C4: Soybeans-no till	968
	C5: Soybeans-min till	2468
	C6: Soybeans-clean till	614
	C7: Alfalfa	54
	C8: Grass/Pasture	497
	C9: Grass/Trees	747
	C10: Grass/Pasture-mowed	26
	C11: Hay-windrowed	489
	C12: Oats	20
	C13: Wheat	212
	C14: Woods	1294
	C15: Building-Grass-Trees-Drive	380
	C16: Stone-Steel towers	95
	Total	10366

Figure 2: Number of labeled samples for Indian Pines dataset

5 Selected Models

Considering the high dimensionality of the data, we used the following three classifiers:

- Kernel SVM: Gaussian kernel, cross-validation for parameter selection
- Random Forest: Use cross validation to select the number of trees and the number of features for each split
- Adaboost: Use Decision tree as basic classifier and cross-validation for parameter selection

Among the above three models, Kernel SVM implement kernel functions to decrease the bias of the classification model. Random forest is a decision tree based method with bagging resampling technique, while Adaboost is a decision tree based method with boosting resampling technique. Both of them attempt to decrease the bias of the model without sacrificing the variance, in other words, avoid overfitting.

The above three models were chosen so that the effects of the three popular method to improve the performance of a classification model: kernel function, bagging and boosting could be easily compared under the scope of hyperspectral data classification.

6 Results and Conclusion

A total of 1200 examples are used for model training. For each of the three models, hyper parameters were chosen based on the result of 5-fold cross validation. The result of the hyper parameter selection is shown in Table 1.

Model	Hyper-parameter	Trained Values	Best Value
SVM	Kernel type	Linear, Polynomial, RBF, sigmoid	RBF
SVM	Weight for regularizer	0:10 (step=0.5)	6.5
SVM	Gamma	0:10 (step=0.5)	0.5
RF	Number of Trees	50, 100, 150, 200	150
RF	Max splitting features	2:20	5
ADA	Learning rate	0.1, 0.5, 1, 2	1
ADA	Max depth	1:13	9

To evaluation the performances of the three models, we have calculated the overall accuracies and class specific accuracies on the testing set. The results are shown in the following table.

Table 1: Accuracies on testing set

	Kernel SVM	Random Forest	Adaboost
Overall Accuracy	**0.85**	0.80	0.82
Corn-no till (C1)	**0.72**	0.68	0.71
Corn-min till (C2)	**0.67**	0.60	0.55
Corn (C3)	**0.91**	0.84	0.88
Soybeans-no till (C4)	**0.82**	0.77	0.76
Soybeans-min till (C5)	0.77	0.77	**0.80**
Soybeans-clean till (C6)	**0.78**	0.66	0.70
Grass/Pasture (C7)	**0.95**	0.91	**0.95**
Grass/Trees (C8)	**0.94**	0.91	**0.94**
Hay-windrowed (C9)	**0.99**	**0.99**	**0.99**
Wheat (C10)	**0.99**	**0.99**	**0.99**
Woods (C11)	0.83	**0.84**	0.82
Building-Grass-Trees-Drive (C12)	**0.78**	0.65	**0.78**

From the above results, we can see that kernel SVM performs the best among the three approaches. This is because SVM with Gaussian kernel project the original data to a much higher feature space in which the data are linearly separable. Using the cross-validation allow us to select such a feature space. For the other two ensemble approaches, boosting is a little better than bagging for this data.

References

[1] Atli,J. Benediktsson, Johan Nes R, Arnason,S.K (1995),Classification and feature extraction of AVIRIS data, IEEE Trans. On Geo Science and Remote Sensing, 33,

[2] El-Naqa.I , Yongi Y, Wernick M.N, Galatsanos .P.N, Nishikawa.M.R ,(2002), A support vector machine approach for detection of micro calcifications, IEEE Transactions on medical Image,21,12, 1552-1563.

[3] Gong.P and Howaeth P.J, (1992), Frequency-based contextual classification and gray level vector reduction for land use identification, Photogrammetric Engineering and Remote sensing, 58, 423-437.

[4] Hermes .L, D.Frieauff, J.Puzicha and J.M Buhmann, (1999). Support Vector Machines for land usage classification in land set TM imagery, in Proc. IGARSS, Hamburg, Germany, 348-350

[5] Lu .D and Weng.Q, (2007), A Survey of Image Classification methods and techniques for improving classification Performance, International Journal of Remote Sensing, 28, 5, 823-870.

[6] Magnussen.S, Boudewyn.P and Wulder.M, (2004), Contextual Classification of Landsat TM images to forest inventory cover types. International Journal of Remote Sensing. 24 2421-2440

[7] Murthy,(1998), Automatic Construction of Decision Trees from Data. A Multi-Disciplinary Survey, Data Mining and Knowledge, 345-389

[8] Pontil.M, Verri.A (1998), Support Vector Machines for 3D Object Recognition, IEEE Transaction pattern Anal. Machine Intel, 20, 637-646.

[9] Robinson, Kecman.V, (2003),Combining Support Vector Machine learning with the discrete cosine transform in Image Compression, IEEE Trans. Neural Networks, 14, 950-958.

[10] Safavian.S.R and D. Landgrebe, (1991), A Survey of decision tree Classifier Methodology, IEEE Transaction on Systems, Man and Cybernetics ,660-674.

[11] Vapnik. V.N, (1998), Statistical Learning Theory, Newyork, Wiley.