Manifold Learning for Spatial Analysis of Hyperspectral Data

Wonkook Kim

Degree : Doctor of Philosophy in Civil Engineering

Academic advisor : Dr. Melba M. Crawford

Accurate land cover mapping is essential for ecological studies where the land cover changes play important role in modeling the interactions between ecosystems and the climate. The advent of hyperspectral sensor offers improved capability for more accurate land cover mapping via the greatly improved spectral resolution. However, constructing a robust classification framework that produces consistent mapping results for varying acquisition conditions, poses two major challenges: (1) "population drift" across different image data sets, and (2) instability of learning algorithms for high dimensional data, which is referred to as "Hughes phenomena".

First, sample distributions of remote sensing data often varies across data sets acquired in different time and space, due to the environmental factors such as soil type, local atmospheric condition, and topography, and exhibit significant differences for the same land cover type. We call the drift in generating populations *population drift*. Direct classification approach which does not consider the population drift usually produces poor classification results. However, the sources and processes that cause the population drift is difficult to model since the variation is highly localized and the acquisition of corresponding reference data is difficult.

Second challenge for the successful land cover classification using hyperspectral data is the problems induced by the high dimensionality of data. Hughes phenomena states that for a fixed number of samples, the performance of a classifier increase as the dimensionality increases, but starts to decrease after an optimal number of dimension. Dimension reduction approach has been used to mitigate the problem through, for example, feature extraction and feature selection methods, prior to the classification. However, many of previous dimension reduction methods do not consider the nonlinearity in hyperspectral data, which causes great loss in information.

In this dissertation, we investigate (1) nonlinear feature extraction methods, namely *manifold learning* and (2) adaptive classification framework, for robust land cover classification, where the two aforementioned issues are treated effectively. We first investigated on the performance of several manifold learning algorithms that are widely used in machine learning community, specifically focusing

on their performance on hyperspectral data. The investigation shows that the manifold learning gives consistently enhanced representation of hyperspectral data compared to other linear methods such as PCA and MNF. In a following study, we tackled the difficulties in applying manifold learning to large volume remote sensing data by constructing a multi-resolution manifold learning framework, to mitigate its heavy computation cost. The experiment results show that comparable classification accuracies are obtained by using the proposed framework with reduced computational complexity. We further develop a kernel-based representation scheme for spatial coherence of remote sensing data to solve the limitation in the image segmentation method used in the previous study. The proposed representation scheme allows more flexible way to consider the spatial homogeneity than image segmentation results where the segment boundaries are fixed for a given level. Finally, we develop an adaptive classification framework using semi-supervised kernel machines without employing feature extraction step prior to the classification. In the experiment, a classifier trained with labeled samples in one location is adapted and applied to samples in spatially disjoint area that exhibit significantly different distributions. The experiment results show that the proposed approach improved the classification accuracy compared to other kernel based semi-supervised classification methods.