Active Learning for Hyperspectral Image Classification.

Wei Di, Ph.D., Purdue University, October 2011. Major Professor: Melba M. Crawford.

Abstract: Obtaining labeled data for supervised classification of remotely sensed imagery is expensive and time consuming. Further, manual selection of the training set is often subjective and tends to induce redundancy into the supervised classifier, thus considerably slowing the training phase. Active learning (AL) integrates data acquisition with the classifier design by ranking the unlabeled data to provide advice for the next query which has the highest training utility. Thus, it explores the maximum potential of the learner toward both the labeled and unlabeled data, and the training set can be maintained as small as possible and focused on the most informative samples for the learning task. This potentially leads to greater exploitation of the information in the data, while significantly reducing the cost of data collection. Despite its promising advantage, although active learning has been widely studied in many other areas, such as document retrieval and natural language learning, related research has been very limited in remote sensing. In this thesis, three major active learning frameworks are proposed for hyperspectral image classification, including:

• Multi-view Adaptive Maximum Disagreement based Active Learning.

Hyperspectral image data contain hundreds of narrow bands over a wide range of the electromagnetic spectrum, making it feasible to obtain multi-view information that is intrinsically embedded in the original data. Instead of drawing samples from the data space to construct the ensembles, the multi-view information is explored to construct the ensembles from the feature space and their disagreement is used to measure the uncertainty for a given unlabeled sample. A Multi-view Adaptive Maximum Disagreement based active learning framework is proposed. It further incorporates a quantitative disagreement measure that aims to compensate for the risk of view insufficiency and avoid the inflation of the contention pool. Moreover, an extensive investigation on view generation, which is key to constructing multi-view based ensembles, is presented. Interesting results are observed, which clearly reveal the highly redundant and complex characteristics of hyperspectral image.

• Local Proximity Data Regularization based Active Learning.

Based on the "*Consistency Assumption*", local proximity is explored to build the query strategy. Samples violate this assumption have higher possibility that lie on the border of classes, which can help to improve the classification hyperplane for the discriminative classifier. Two kinds of approaches are proposed. The first approach is based on Laplacian graph and Gaussian Heat kernel, which incorporates the spatial information through a composite kernel. The second approach combines the multi-view disagreement information and the local inconsistency to build a co-regularizer, whereby the local inconsistency is measured on the low dimensional manifold structure, which is generated from the spatial or spectral space and is assumed to be able to enforce the consistency between similar samples.

• Critical Class oriented Active Learning.

In order to focus on the hard classes in a multi-class classification task, a critical class oriented query strategy is proposed, which combines the "guided learning" and "active learning". In conjunction with SVM classifier, hard pair classes are first identified based on the instability of the classification

hyperplane, whereby a category level guidance for which class should be queried next is given and then feed the active querier. Samples with higher possibility belong to these classes as evaluated by the current learner are queried first. Based on different critical class set generation and query strategies, several methods are developed, including: the single-view based critical class, and its margin sampling variation, which server as the base-line methods; three multi-view based weighted critical class query strategies. Further, the proposed methods are extended to batch- mode learning, which aims to alleviate the computational cost in the single-sample query.

Experiments were conducted on both the AVIRIS and Hyperion hyperspectral data set. Results are compared to passive learning method: Random Sampling (RS), and the state-of-the-art active learning method SVM based simple margin sampling. Superior performance is obtained with interesting some observations, which especially involved in the multi-view learning and critical class oriented learning.