ABSTRACT

Zhang, Zhou Ph.D., Purdue University, August 2017. Spectral-Spatial Classification of Hyperspectral Data using Active Learning. Major Professor: Melba M. Crawford.

Augmenting spectral data with spatial information for hyperspectral data classification has recently gained significant attention, as classification accuracy can often be improved by spatial information from neighboring pixels. However, the resulting high-dimensional feature space often imposes significant challenges in developing robust supervised classifiers, especially when the quantity of labeled samples is limited. Active learning (AL), which aims to find the most informative training set, has become an effective approach for dealing with the limited availability of labeled samples. However, most AL methods presented in the literature for classification of hyperspectral data deal solely with spectral information. Furthermore, common AL strategies work in a fixed feature space, and the feature space is not updated as AL proceeds with potentially suboptimal performance. In this dissertation, three adaptive active learning strategies are proposed for hyperspectral data classification using both spectral and spatial features:

- Single view based active learning with adaptive spatial information

  Since hyperspectral data is defined in both spectral and spatial domains, combining spectral and spatial features for its classification helps reduce the labeling uncertainty when utilizing only the spectral information. A new strategy that combines spectral and spatial information is developed. First, a new supervised strategy is proposed for segmentation hierarchy pruning that allows a best map to be determined and updated at each iteration of the AL process, thus exploiting informative labeled samples provided by the user. Then the resulting map is used to incorporate the spatial information via two strategies: 1) enrich spec-

tral signatures with spatial features (adding features); 2) add pseudo-labeled samples through a new semi-supervised approach (adding samples).

- Multi-view based active learning with view updating

  Besides the limited number of labeled samples, high dimensionality of the input data resulting from stacking different types of features is another challenge for supervised classification. To address this issue, the proposed single view active learning framework is extended to a multi-view setting. In particular, the hyperspectral data is partitioned into multiple disjoint views, and for each view, the spatial features are extracted from the corresponding best segmentation map and combined with the spectral features. For AL, a maximum disagreement query strategy is adopted. Moreover, a dynamic view is generated at each AL iteration to further increase the disagreement level which is important for AL query.

- Regularized multi-metric active learning with updated regularizer

  To simultaneously address the issues of high dimensionality and limited labeled samples, a new strategy is proposed which combines feature reduction for multi-type feature data and batch-mode AL into a unique framework. To accomplish this, a regularized multi-metric learning approach was developed to jointly learn a specific metric for each type of feature, and the regularizer is refined by exploiting the increased labeled information provided by the user as AL proceeds. In addition, a batch mode AL query strategy considering both uncertainty and diversity criteria was developed in order to select the most informative samples in the final low-dimensional feature space.