SPARSE BAND SELECTION FOR SUPPORT VECTOR DATA DESCRIPTION APPLICATIONS

Amit Banerjee, Joshua Broadwater, and Philippe Burlina

The Johns Hopkins University
Applied Physics Laboratory
11100 Johns Hopkins Rd.
Laurel, MD 20723 USA

1. INTRODUCTION

Support Vector Data Description (SVDD) methods have been successfully applied to tasks such as hyperspectral anomaly detection [1] and spectral unmixing [2] [3]. Unfortunately, the performance of SVDD methods suffers when noisy or non-informative bands are present in the data. If a set of sparse bands could be identified for these techniques, the resulting data may improve SVDD performance while enjoying the benefits of decreased processing overhead. Although band selection has been investigated in previous efforts, this work builds on recent research that has resulted in the development of a theoretical framework for signal classification with sparse representation using L_1 measures. This data-driven approach combines the classification power of the discriminative methods with the reconstruction property and a sparse representation that enables one to deal with signal corruptions: noise, missing data and outliers.

2. ALGORITHM DESCRIPTION

Given the large set of spectral bands present in hyperspectral imagery, a critical challenge is to identify those bands that maximize the ability of the SVDD algorithm to separate different classes of materials either for detection or unmixing applications. In this effort, we propose to extend a L_1 framework to automatically select a sparse set of spectral bands that provides similar or better SVDD performance than the entire set of wavelengths. This is a critical issue for kernel SVDD methods, since they suffer from the "curse of dimensionality", where the presence of noisy or redundant bands can reduce their performance.

Sparse representations of signals over an overcomplete basis (dictionary) for band selection have been successfully exploited for applications such as spectral estimation [4] [5]. These approaches also exploit the redundancy and correlations of the high-dimensional data, especially when the feature vector is sparse in another transform domain. Hyperspectral imagery is an example of such data and thus can be analyzed via overcomplete basis functions.

Searching for the sparse representation of a signal over an overcomplete dictionary is achieved by optimizing an objective function. In general, the function contains two terms, one that measures the signal reconstruction error and another that measures the sparsity. This objective function works well in applications where signals need to be reconstructed, like coding and denoising.

To find a sparse set of spectral bands, we propose to use discriminative criteria, such as Fisher's criterion (or LDA), as they are suited for classification tasks. Recent research [5] [6] has resulted in the development of a theoretical framework for signal classification with sparse representation. The approach combines the discrimination power of the discriminative methods with the reconstruction property and the sparsity of the sparse representation that enables one to deal with signal corruptions: noise, missing data and outliers. The proposed approach is therefore capable of robust classification with a sparse representation of signals and may provide a sparse set of bands that reduce the "curse of dimensionality" for kernel unmixing methods.

For signal classification, one can adopt a class separability criterion for the cost function. Optimizing the resulting cost function will produce a sparse subset of spectral bands that maintains similar class separation as with the original full spectrum. For example, given C classes of spectra with mean vectors $\mathbf{m}_i$, i=1,…,C, define the within-class scatter matrix as

$$S_W = \sum_{i=1}^{C} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T \tag{1}$$

and the between-class scatter matrix as

$$S_B = \sum_{i=1}^{C} (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T \; , \tag{2}$$

where $\mathbf{m}$ is the mean spectra of entire set of spectra. Fisher's criterion for class separability is then defined as $F(\mathbf{x}) = \mathrm{tr}(S_B S_W^{-1})$. The objective function can be re-written as

$$J(\mathbf{x}; \lambda) = F(\mathbf{x}) - \lambda \|\mathbf{x}\|_1 \; , \tag{3}$$

to find a sparse set of bands with good class separability.

3. EXPERIMENTAL RESULTS

Using real-world hyperspectral imagery, a set of experiments is provided to show the applicability of the proposed algorithm to SVDD anomaly detection. Based on previous work [1], receiver operating characteristic (ROC) curves will be generated for both the full set of bands, set of bands with known atmospheric effects removed (e.g. water absorption bands at 1.4 and 1.9 microns), and the set of bands chosen using the proposed algorithm. Results show the ability of the new algorithm to provide improved performance while reducing the computational processing overhead.

4. REFERENCES

[1] A. Banerjee, P. Burlina, and C. Diehl, "A support vector method for anomaly detection in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 8, pp. 2282-2291, Aug. 2006.

[2] A. Banerjee, P. Burlina, and J. Broadwater, "A machine learning approach for finding hyperspectral endmembers," in *IEEE International Geoscience and Remote Sensing Symposium*, Barcelona, Spain, 2007, pp. 3817-3820.

[3] J. Broadwater, A. Banerjee, and P. Burlina, "Kernel Methods for Unmixing Hyperspectral Imagery," in *Kernel Methods for Remote Sensing Data Analysis*, G. Camps-Valls and L. Bruzzone, Eds. West Sussex, UK: John Wiley & Sons, 2009, ch. 11, pp. 249-269.

[4] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397-3415, Dec. 1993.

[5] K. Huang and S. Aviyente, "Sparse Representation for Signal Classification," in *Advances in Neural Information Processing Systems 19*, B. Scholkopf, J. Platt, and T. Hoffman, Eds. Cambridge, MA, USA: MIT Press, 2007, pp. 609-616.

[6] S. Fidler, D. Skocaj, and A. Leonardis, "Combining reconstructive and discriminative subspace methods for robust classification and regression by resampling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 3, pp. 337-350, Mar. 2006.