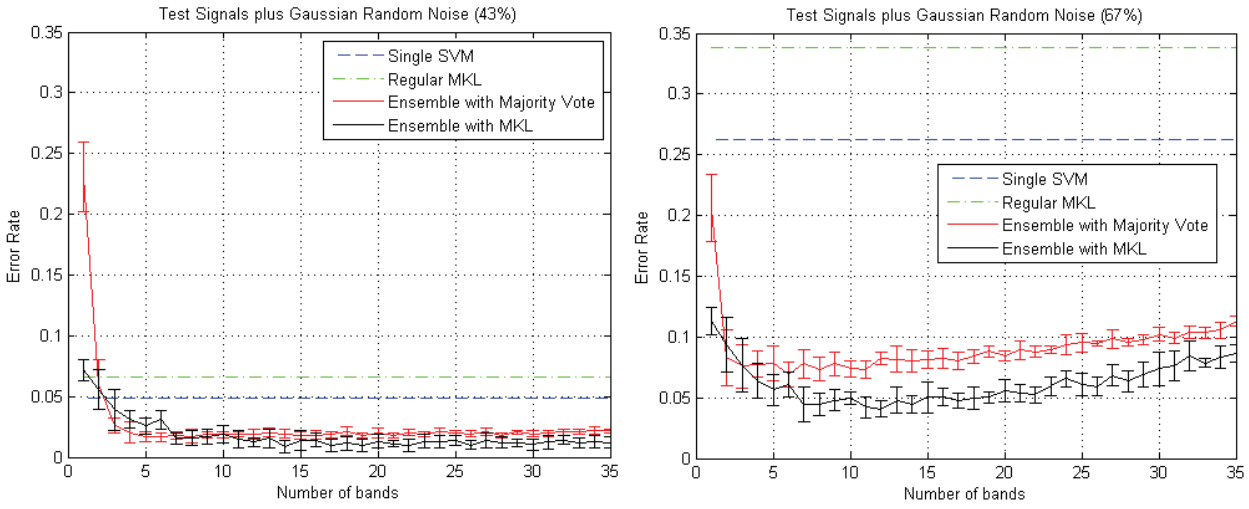
A FULL DIAGONAL BANDWIDTH GAUSSIAN KERNEL SVM BASED ENSEMBLE LEARNING FOR HYPERSPECTRAL CHEMICAL PLUME DETECTION

Prudhvi Gurram, Heesung Kwon

Army Research Lab, ATTN: RDRL-SES-E

2800 Powder Mill Rd., Adelphi, MD 20783

Support Vector Machines (SVMs) have been popular tools for discriminative classification. The kernel trick has made SVMs more popular than ever due to its ability to deal with non-linear and non-separable data [1]. To improve the generalization performance of SVMs, it is customary to use multiple kernels to build weak classifiers and combine them using some kind of aggregating technique. Recently, a feature-based Ensemble Learning (EL) technique has been proposed to be used with SVM for hyperspectral chemical plume detection [2]. In this technique, feature subsets are randomly selected from the high-dimensional hyperspectral data and are used to train the SVMs. Many aggregating strategies like majority voting and weighted average can be used to combine the weak classifiers obtained from the SVMs trained using the random subsets of features. It has been shown in [2], that this technique has a better generalization performance than bagging [3] or boosting [4, 5] algorithms which use subsets of training samples. In this direction, an iterative L2-L1 norm optimization based Multiple Kernel Learning (MKL) technique [6] has been used to combine the ensemble of random feature subset weak classifiers and found to have further improved the generalization properties of the classifier [7]. This is illustrated in Figures 1(a) and (b), where the error rates have been plotted with respect to number of bands selected. This figure compares different aggregating techniques for two different noise levels in test data while no noise has been added to the training data.



(a) 43% Noise level (b) 67% Noise level

Figure 1. Comparison of different aggregating strategies for feature-based ensemble learning used for hyperspectral chemical plume detection

But in all these experiments, a single width (σ^2) has been used in the Gaussian kernel functions for all the features selected randomly to build the weak classifiers as shown in Equation 1. This results in a spherical Gaussian kernel.

$$\kappa(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (1)$$

In this paper, a method to use a full diagonal bandwidth Gaussian kernel in the weak classifiers is proposed. The kernel is shown in Equation 2.

$$\kappa(x_i, x_j) = \exp\left(-\frac{1}{2}(x_i - x_j)^T C^{-1}(x_i - x_j)\right) \quad (2)$$

Here the matrix C is given by (where N is the number of features being used in the classifier)

$$C = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & \sigma_N^2 \end{bmatrix} \quad (3)$$

In the existing algorithm, various single bandwidth Gaussian kernels are used to train different sets of weak classifiers. The single bandwidth (the ensemble of weak classifiers) which gives the best classification results is chosen from this set using cross-validation technique or MKL technique itself. In the proposed algorithm, the full bandwidth parameters for each kernel are obtained by minimizing the upper bound of a generalization error estimate like Leave-One-Out error estimate using gradient descent technique. The bound used in this paper is called the Radius-Margin bound [8].

Classification results will be compared for different aggregating techniques like majority voting and MKL with feature-based ensemble learning using the full diagonal bandwidth Gaussian kernel for chemical plume detection. A significant improvement in the results is expected when a full diagonal bandwidth Gaussian kernel is used in the weak classifiers rather than a single bandwidth Gaussian kernel.

References:

[1] Scholkopf, B., and A. J. Smola, *Learning with Kernels*, The MIT Press, Massachusetts, 2002.

[2] Kwon, H., and P. Rauss, "Feature-based Ensemble Learning for Hyperspectral Chemical Plume Detection," under review for publication in the *International Journal of Remote Sensing*.

[3] Brieman, L., "Bagging Predictors," *Machine Learning*, vol. 24, no. 2, pp. 123 – 140, 1996.

[4] Schapire, R. E., "The Strength of Weak Learnability," *Machine Learning*, vol. 5, no. 2, pp. 197 – 227, 1990.

[5] Freund, Y., and R. E. Schapire, "A Decision-theoretic Generalization of On-line Learning and An Application to Boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119 – 139, 1997.

[6] Rakotomamonjy, A., F. R. Bach, S. Canu, and Y. Grandvalet, "SimpleMKL," *Journal of Machine Learning Research X*, pp. 1-34, 2008.

[7] Gurram, P., and H. Kwon, "Ensemble Learning Based on Multiple Kernel Learning for Hyperspectral Chemical Plume Detection," to be published in *Proceedings of SPIE Defense, Security and Sensing Symposium*, Orlando, Fl, 2010.

[8] Chapelle, O., V. Vapnik, O. Bousquet, and S. Mukherjee, "Choosing Multiple Parameters for Support Vector Machines," Machine Learning, vol. 46, pp. 131-159, 2002.