Predicting Motor Tasks in fMRI Data with Support Vector Machines

S. LaConte[1], S. Strother[2], V. Cherkassky[2], X. Hu[1]

[1]Emory University/Georgia Tech, Atlanta, GA, United States, [2]University of Minnesota, Minneapolis, MN, United States

SYNOPSIS The support vector machine (SVM) is introduced to fMRI as a method for classifying temporal scans. The SVM is found to perform comparably to canonical variates analysis (CVA) in terms of misclassification error. We examine the interpretation of the SVM model in the context of fMRI, and find that removing model-related scans enhances the statistical difference between those scans.

INTRODUCTION This study focuses on temporally predictive models of fMRI data. Specifically, we introduce the support vector machine (SVM), a universal supervised learning method arising out of the statistical learning theory of Vapnik [1,2]. Our focus on temporal prediction is motivated by three key observations: i) Temporal information about an fMRI experiment is critical for either directly obtaining summary maps, or indirectly interpreting "data-driven" results. ii) As recently demonstrated with CVA, classification of temporal scans can be used in model validation [3]. iii) The unique formulation of the SVM, which can find non-linear decision boundaries and handle high dimensional problems, makes it a promising technique for fMRI. We compare the classification accuracy of the SVM to that of canonical variates analysis (CVA) using cross-validation on data from a visually guided static force task. A key issue that needs to be addressed for the neuroscientific applicability of the SVM to fMRI is the physical meaning associated with a given support vector model. We begin to address this interpretation problem by looking at the SVM model of a simple bi-manual finger opposition task.

THEORY Here we summarize only the salient concepts for SVM-based classification (see [3,4] for a description of CVA). Training data, consisting of input vectors, $\vec{x}_i$, and their corresponding class labels, y_i, are used to obtain a SVM model. For fMRI, each $\vec{x}_i$ and y_i are the measured brain voxels and experimental design value for the time point, i, respectively. For simplicity, we examine the binary classification problem ($y_i = \pm 1$), corresponding to e.g. task vs. rest. The input vectors are mapped to high dimensional feature space via a non-linear mapping function $\vec{z} = g(\vec{x})$. The SVM algorithm attempts to find linear decision boundaries (separating hyperplanes) in the feature space, formalized by the decision function $D(\vec{z}_i) = (\vec{w} \cdot \vec{z}_i) + w_0$, where $\vec{w}$ defines the linear decision boundaries. This hyperplane satisfies $y_i[(\vec{w} \cdot \vec{z}_i) + w_0] \geq 1$, and is optimal when $\|\vec{w}\|^2$ is minimized. For data that cannot be separated without error, slack variables, ζ, are defined as the values of the error between the true class labels and the decision functions. In this case, the hyperplane is defined by $y_i[(\vec{w} \cdot \vec{z}_i) + w_0] \geq 1 - \zeta_i$, and is optimal when $C/n \sum_{i=1}^{n} \zeta_i + \frac{1}{2}\|\vec{w}\|^2$ is minimized. The free parameter, C, affects the trade-off between complexity and number of non-separable samples. An important result, not derived here, is that the solution of the optimal $\vec{w}$ is a linear combination of a subset of the training vectors, $\vec{x}_i$, termed support vectors. Thus the trained model consists of the predefined non-linear function $g(\cdot)$, the support vectors, and the support vector class labels.

METHODS The SVM algorithm used was SVMlight [5]. Comparing SVM with CVA Here we used a subset of the data described earlier [3]. Eight right-handed volunteers performed a static force task alternating six rest and five force periods/run (44 s/period; 200-100 g randomized forces with thumb and forefinger). Data were collected on a Siemens 1.5T scanner (EPI BOLD: TR/TE=3986/60 msec, slices=30, voxel=3.44x3.44x5 mm). After transformation to Talairach space, data from the eight subjects were divided into two equal parts (using all possible combinations) to create 70 sets of training and independent test data. Principal components analysis (PCA) was used as input to both CVA and SVM classification techniques. For several levels of model complexity, resampled estimates of percent misclassification error was used as the index for comparison. For CVA this complexity was varied with the number of PCs used. For the SVM, a polynomial mapping function was used, and the degree of the polynomial and a range of C values determined model complexity. Interpreting the SVM model A single subject performed a sequential finger opposition task performing 20 s each of left-hand, right-hand, and rest. This series was repeated 4 times (for a total of 4 min.). Two oblique, axial, EPI slices were collected on a 3T Siemens Trio (TR/TE = 200/32 msec, slices=2, voxel=3.4×3.4×5 mm). A support vector machine model was trained to classify left vs. right scans using a linear mapping function with C=2000. A simple voxel-wise two-sided t-test was performed on left vs. right scans, with and without the support vector data.

RESULTS AND DISCUSSION Comparing SVM with CVA Similar misclassification results were obtained with both CVA and SVM as shown in Fig. 1. We did not observe SVM sensitivity to the C parameter. The average number of support vectors for each basis set (degree 1-4) was 186, 300, 346, and 357 out of a maximum possible of 364 (the number of PCs used). The degree 1 and degree 2 models performed similarly, although the former used dramatically fewer support vectors. Interpreting the SVM model Conceptually, the support vectors, which determine the decision boundary, are most difficult to classify. While not guaranteed by this property, the enhancement of the t-statistics in Fig. 2 corroborates this explanation. Shown are the distributions of t-scores from the two t-tests (all left-right scans 400/400 and the subset of non-support vector scans 273/317). The distribution from the subset has enhanced the t-map – small t-scores are smaller, evinced by the higher peak, while large t-scores are larger, evinced by the slight elongation of the tails. Note the increase in t-values occurs even with a simultaneous decrease in degrees of freedom.

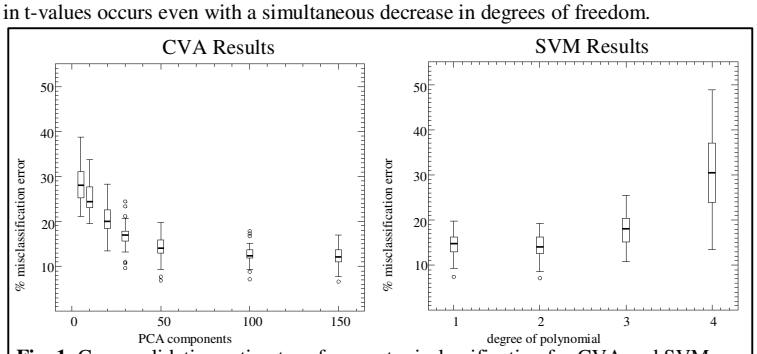

Fig. 1. Cross validation estimates of percent misclassification for CVA and SVM models for various levels of model complexity

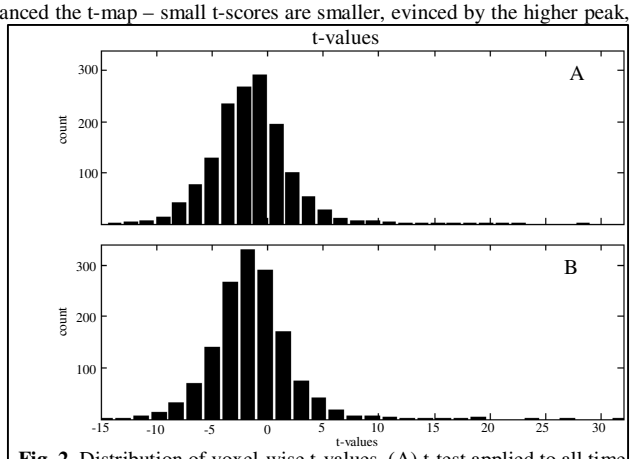

Fig. 2. Distribution of voxel-wise t-values. (A) t-test applied to all time points. (B) times corresponding to non-support vectors.

SUMMARY We have demonstrated the applicability of SVM for building predictive models of fMRI data. For the conditions studied, SVM performed comparably to CVA in terms of misclassification error. Neuroscientific interpretation of the SVM model is an important issue for further study. Here we demonstrated that one characteristic of the support vectors is that they tend to represent scans that are difficult to classify. We have observed a relatively large number of support vectors in all of our models, perhaps indicating that learning problems in fMRI are relatively difficult.

REFERENCES [1] Vapnik, V. *The Nature of Statistical Learning Theory*, New York: Springer Verlag, 1995. [2] Cherkassky, V. and F. Mulier. *Learning from Data: Concepts, Theory and Methods*, New York: John Wiley & Sons, 1998. [3] LaConte, S., et al. NeuroImage, in press. [4] Strother, S.C., et al. NeuroImage 15:747-771, 2002. [5] Joachims, T. *Advances in Kernel Methods - Support Vector Learning*, B. Schölkopf and C. Burges and A. Smola (ed.), MIT-Press, 1999.

ACKNOWLEDGEMENT Work supported by the Whitaker Foundation, the Georgia Research Alliance, and NIH grants MH57180,RO1MH55346 and RO1EB00321.