Real-Time functional MRI Classification of Brain States Using Markov-SVM Hybrid Models:

Peering inside the rt-fMRI black box

Machine Learning (ML) methods applied to real-time functional MRI (rt-fMRI) data provide the ability to predict and detect online any changes in cognitive states. Applications based on rt-fMRI require appropriate selection of features, preprocessing routines, and models in order to both be practical to implement and deliver interpretable results. In the paper, we evaluate blind MVPA feature selection methods against *a priori* defined spatial IC maps, as well as the role of non-stationarity in real-time prediction. We also compare the performance of offline modeling and prediction with online, and the effectiveness of blind machine learning algorithms such as SVM with hybrid Bayesian-SVM models that utilize history of states to predict future occurrences. Collectively, we explore what is inside the "black-box" of real-time fMRI, and examine both the advantages and shortcomings when ML methods are applied to predict and interpret cognitive states in the real-time context.

Ariana Anderson, Dianna Han, Pamela K. Douglas, Jennifer Bramen, Mark S. Cohen

UCLA

iNt Laboratory of Integrative Neuroimaging Technology

Blood **O**xygenation **L**evel **D**ependent functional MRI (BOLD-fMRI) detects changes in brain energy consumption related to neuronal signaling.

fMRI for Detection of Cognitive State

fMRI experiments seek to "read out" brain or cognitive state based on reverse inference from the measured brain signals. More recent experiments seek to detect these state changes in real-time, assigning brain states to individual time points. Traditionally, these assumed a one-to-one, linear, correspondence between signal in a brain region and an independent measure, such as subjective pain (deCharms, 2005). Multiple brain regions contribute to functional brain activity, however, as subcomponents of systems distributed across the brain.

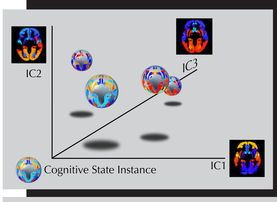
Brain Reading with Machine Learning

Machine Learning may classify brain states with high accuracy based on weak priors. Haxby showed good classification accuracy to visual stimuli by looking at a small number of individual pixel intensities constrained simply to be adjacent in a small brain volume (Haxby, 2001).

Interpretable Features

Machine methods generate a "hidden" decision layer. To support discovery, the decision layer ideally is interpretable within the natural architecture of the system under test.

Independent Component Analysis applied to fMRI time series factors the time data into spatial topographies whose time courses are statistically independent. The resulting ICs often reflect functional systems (Smith, 2009). We use these ICs as *interpretable* features to separate cognitive state classes (Anderson, 2011). In essence, we describe a cognitive state as a linear superposition of activities in functional subsystems.



Schematically, we can imagine a classifier that operates on just three ICs as features. Any given cognitive state is seen as having an identifiable location in this three dimensional space.

Blood Oxygenation Level Dependent (BOLD) fMRI signal drifts due to:

- Subject Motion
- Instrument Instability
- Physiological Changes
- Changes in Cognitive Background State
- Brain Changes inducted by Feedback

The goal of our work presented here is to create an adaptive classification process to mitigate problems in data drift, while retaining the interpretability of our IC-based classifier.

Data and General Methods

MRI scans of 51 subjects scanned while trying to control their drug craving in response to provocative cues (described fully in Brody, 2007).

We studied two feature sets:
1. ROI features extracted as the average signal in each of 110 atlas ROIs.
2. IC/functional correlations *extracted as follows:*
a) create a dictionary of common ICs (methods from Anderson, 2011)
-- spatial alignment of ICs from 279 single sessions
-- project these onto the same lower dimensional atlas subspace of 110 regions.
-- Identify 20 clusters of ICs using k-means and bootstrapping
b) Create feature vector from correlation of Vt at time t

$$\vec{x}_t = \left(r^2(IC_1, V_t), r^2(IC_2, V_t), \ldots, r^2(IC_{20}, V_t)\right)$$

We wish to learn the model, g, that optimally predicts the cognitive state class association of $\vec{x}_t$ at time t to the set of N possible cognitive states:

$$g : \vec{x}_t \rightarrow \mathcal{C}$$

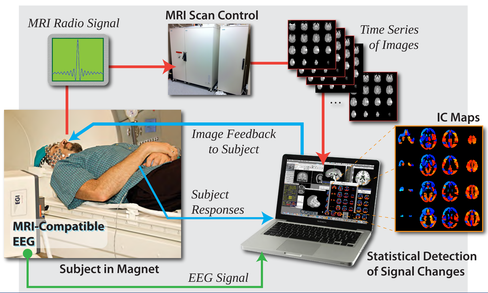
We evaluate classifier and model drift by assessing the effects of demeaning the data within each feature on the classification accuracy.

Machine Methods

All classifiers used variations on Support Vector Machines (SVM). We augment the SVM model with a transition matrix, **A**, (similar to Garczarek, 2002). Let $C_{i,t}$ denote the cognitive state i of system $C \in \mathcal{C}$ at time t.
The *states form a discrete-time Markov chain,* if for any states: $\{j, i, i_{t-1}, \ldots, i_0\}$,

$$P(C_{t+1} = j \mid C_t = i, C_{t-1} = i_{t-1}, \ldots, C_0 = i_0) = P(C_{t+1} = j \mid C_t = i).$$

The rows $\vec{a}_i$ of transition matrix **A** contain the transition probabilities to all states: $j \in 1, \ldots, N$ given the previous state i. Each element $a_{i,n} \in \vec{a}_i$ gives the probability of transitioning to state $n \in N$ given the previous state i.



In Real-Time fMRI data are recorded from the subject immediately during scanning, then analyzed rapidly enough for immediate re-presentation to the human subject. In our unique implementation both functional MRI and electroencephalographic data (EEG) are collected and analyzed jointly.

Models Tested

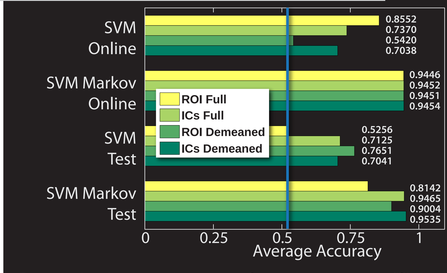
Model A - SVM Online: Train SVM model g on data from time (1,t-1) and test on time t. For N possible states at t, g outputs current likelihood of state $\vec{c}_t$ given the prior data, such that $\vec{c}_t = g(x_t \mid \vec{x}_{t-1}, \vec{x}_{t-2}, \ldots, \vec{x}_0)$, where $\vec{c}_t = \left(p(C_{t,1}), p(C_{t,2}), \ldots, p(C_{t,N})\right)$. The model is updated at each time point but is readily computable because there are few EVs

Model B - SVM Markov Online: Update SVM class probabilities using Markov transition matrix, **A**, estimated at each time point given history of the process. Predicted class label C_j at time point t, given state is decided by $C_j = \max_{n \in N}\left\{a_{i,n} \vec{c}_t\right\}$.

Model C - SVM Test: Train an SVM model, g, offline, then test it on a new scan from the same subject.

Model D - Markov Test: Train a new model, g, offline using a training scan, and test it online using the testing scan. The offline model updates the SVM probabilities with the Markov transition matrix, **A**, learned from the training data.

Results



- Demeaning increased accuracy for online, but not offline tests
- Online models slightly better (4%) than offline models
- Adding a Markov transition matrix increased accuracy by about 23%
- **ROI Features** accuracy much more sensitive than **IC/functional correlations** to demeaning and online/offline choice

Discussion

The basic SVM models ended up being highly biased, with as much as 30% difference between the predicted and obtained accuracy. Cognitive state changes are probably not random events; harnessing the information contained in autocorrelations among successive observations can improve classification accuracy.

The Markov transition matrix removed unlikely state transitions. For example, in a blocked design, the transition probabilities to many states approaches zero. This effectively embodies information of the experimental design itself. In a neurofeedback experiment, Markov models of the drift may be useful in monitoring therapy and functional plasticity. It is likely that this Markov transition model will apply broadly to real-world data beyond neuroimaging. By incorporating the temporal dependencies into the model we leverage known structure to classify an unknown outcome.

Our IC dictionary is a lower dimensional feature space, as compared to the **ROI Features** method, the **IC/functional correlations** approach results in better statistical performance, supporting the model and hypothesis that the functional architecture of the brain prominently includes interactions among regions. It is a further step away from the neophrenological concept of strict localization to a more holonomic view of the brain.

Overall, this work emphasizes the importance of incorporating prior knowledge into both feature selection and the machine algorithms. Peering into the rt-fMRI black box can be more illuminating than observing what it produces.

This work is supported by NIH DA026109 to M.S.C.

deCharms, R. C., F. Maeda, et al. (2005). "Control over brain activation and pain learned by using real-time functional MRI." *Proc Natl Acad Sci U S A* 102(51): 18626-18631

Haxby, J. V., M. I. Gobbini, et al. (2001). "Distributed and overlapping representations of faces and objects in ventral temporal cortex." *Science* 293(5539): 2425-2430.

Smith, S. M., et al. "Correspondence of the Brain's Functional Architecture During Activation and Rest." *Proc Natl Acad Sci U S A* 106.31 (2009): 13040-5

Anderson, Ariana, et al. "Large Sample Group-ICA of fMRI Using Anatomical-Atlas Based Reduction, Bagging and Clustering." *International Journal of Imaging Systems and Technology* 21.2 (2011): 223-31

Brody, A. L., et al. "Neural Substrates of Resisting Craving During Cigarette Cue Exposure." *Biological Psychiatry* 62.6 (2007): 642-51

Garczarek UM. "Classification rules in standardized partition spaces." *Doctoral Dissertation: University of Dortmund*, 2002