Fusing Simultaneous EEG-fMRI by Linking Multivariate Classifiers

Bryan R. Conroy[1], Jordan Muraskin[1], and Paul Sajda[1]

Dept. of Biomedical Engineering, Columbia University, Columbia NY, USA

Abstract. Multivariate pattern analysis (MVPA) has typically been used in neuroimaging to draw inferences from a single modality, e.g., functional magnetic resonance imaging (fMRI) or electroencephalography (EEG). As simultaneous acquisition of different neuroimaging modalities becomes more common, one consideration is how to apply MVPA methods to analyze the resulting multimodal dataspaces. We present a multi-modal fusion technique that seeks to simultaneously train a multivariate classifier and identify correlated components across the two modalities. We validate our approach on a real simultaneous EEG-fMRI dataset.

Keywords: Multi-modal data fusion, simultaneous EEG-fMRI

1 Introduction

Multivariate pattern analysis (MVPA) has typically been used in neuroimaging to draw inferences from a single modality, e.g., functional magnetic resonance imaging (fMRI) or electroencephalography (EEG). As simultaneous acquisition of different neuroimaging modalities becomes more common, one consideration is how to apply MVPA methods to analyze the resulting multimodal dataspaces. Currently, a popular approach is to apply multivariate techniques in one domain, such as EEG, and then utilize the results in a univariate regression over the other modality, e.g., fMRI [1–3]. This of course is not optimal, since it biases one modality over the other, when in fact both might rightfully best be analyzed in a multivariate space. Other approaches have used factorization methods such as ICA to identify correlated components across the two modalities [4], but these methods are usually exploratory and completely unsupervised.

In this paper we describe a principled approach for MVPA on simultaneously acquired EEG-fMRI data, which we pose as a multi-objective optimization problem. Given a set of experimental trial labels, e.g., task difficulty, we derive an objective function which seeks to maximize classification performance in predicting these labels, while simultaneously forcing the modality subcomponents of this classifier to be correlated. This combines aspects of supervised learning (classification) with exploratory data analysis (correlated components). When analyzed in the joint space, the method has a simple interpretation as a regularized classification method.

2 Methods

We assume that the data acquired during a simultaneous EEG-fMRI experiment have been grouped into a set of n trials. This produces a multi-modal dataset $(E_1, F_1, y_1), \ldots, (E_n, F_n, y_n)$, where for each trial i we have p_E EEG features $E_i \in \mathbb{R}^{p_E}$, p_F fMRI features $F_i \in \mathbb{R}^{p_F}$, and a trial label y_i that encodes information about the stimulus condition. Here, we focus on the binary discrimination case, where $y_i \in \{0, +1\}$. We do not impose any restrictions on the types of features extracted from each modality. For example, E_i may be electrode voltages over a given time-window, or spectral power from particular frequency bands, and F_i may be the BOLD response from voxels from a region-of-interest (ROI). We standardized all of the features to have zero mean and unit norm, and assembled the EEG and fMRI trial data into columns of the matrices E and F.

We have a balance of two goals: (a) to design a classifier that aggregates activity across both modalities in order to predict the trial labels; and (b) identify within-modality components of this classifier that share similar patterns of trial-to-trial variability, i.e., the predictions output from each modality should be correlated. Specifically, we seek to learn linear projections of the EEG data $v \in \mathbb{R}^{p_E}$ and of the fMRI data $w \in \mathbb{R}^{p_F}$ such that the joint component across modalities $E^T v + F^T w$ is predictive of the trial label vector y, and also the modality subcomponents $E^T v$ and $F^T w$ are correlated. We propose to achieve this by minimizing an objective of the form:

$$(v^*, w^*) = \arg\min_{v,w} \ell(y, E^T v + F^T w) + \lambda \text{dissim}(E^T v, F^T w) \tag{1}$$

where $\ell(\cdot, \cdot)$ is a classifier loss function, $\text{dissim}(\cdot, \cdot)$ measures the dissimilarity between modality components, and λ is a trade-off parameter.

We use logistic regression to measure prediction accuracy, in which case $\ell(\cdot, \cdot)$ is the negative log-likelihood of y given the joint component $E^T v + F^T w$:

$$\ell(y, E^T v + F^T w) = -\sum_{i=1}^{n} y_i (E_i^T v + F_i^T w) - \log\left(1 + \exp(E_i^T v + F_i^T w)\right) \tag{2}$$

To maximize the similarity between the subcomponents, one could attempt to maximize the covariance between them: $\text{cov}(E^T v, F^T w) = v^T E F^T w$.[1] This, however, does not result in an objective that is jointly convex in (v, w), so we instead propose to minimize the following:

$$\text{dissim}_\alpha(E^T v, F^T w) = ||v||^2 + ||w||^2 - \frac{2\alpha}{\sigma_{\max}} v^T E F^T w \tag{3}$$

where α is a parameter and $\sigma_{\max}$ is the maximum singular value of EF^T. The above is convex for $\alpha \in [0, 1]$.

[1] Since the features are assumed to have zero mean, $\text{cov}(E^T v, F^T w) = v^T E \left(I - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right) F^T w = v^T E F^T w$.

Although correlation is not transitive, the subcomponents will tend to be correlated to each other simply because they are both discriminative of the trial label vector y. As a result, the inter-modality correlation objective could simply be locking on to patterns that both relate to y. To account for this, we can instead drive our algorithm to maximize the partial covariance of $E^T v$ and $F^T w$ given y by replacing EF^T in (3) with $E\left(I - yy^T/||y||^2\right)F^T$. As we show in the results, however, this does not seem to be necessary as we achieve significant cross-validated partial correlations even without this modification.

To better understand how the two objectives interact with each other, it is beneficial to represent (1) in the joint space. Specifically, let $z = (v, w)$ and $X = [E, F]$. Then our objective is equivalent to:

$$z^* = \arg\min_z \ell(y, X^T z) + \lambda z^T L_\alpha z \tag{4}$$

$$L_\alpha = \begin{bmatrix} I & -\frac{\alpha}{\sigma_{\max}} EF^T \\ -\frac{\alpha}{\sigma_{\max}} FE^T & I \end{bmatrix} \tag{5}$$

We may interpret (4) as a regularized version of logistic regression in the joint space, with L_α serving as the regularization matrix. If $\alpha = 0$, then $L_\alpha = I$ and our objective reduces to simple l2-regularized logistic regression, which ignores linking between the modalities. For $\alpha > 0$, the regularization matrix L_α "penalizes" components differentially depending on the degree of inter-modality covariance. Let $EF^T = \sum_i \sigma_i v_i w_i^T$ be a singular value decomposition, with singular values given by $\sigma_i = \mathrm{cov}(E^T v_i, F^T w_i)$. Then the penalty incurred by $z = (v_i, w_i)$ is given by $2(1 - \alpha\sigma_i/\sigma_{\max})$.

The optimization in (4) is carried out in a very high-dimensional space – the dimension of z is the sum of the feature space dimensions of the two modalities. In most neuroimaging experiments, however, the brain data is low-rank since the number of imaging sensors greatly exceeds the number of trials acquired over the course of the experiment (i.e., $p_E, p_F \gg n$). This assumption can be exploited to greatly improve computational performance. Let Q_E and Q_F denote orthogonal matrices whose columns span the n-dimensional ranges of E and F, respectively. Then it can be shown that the solution to (4) can be expressed as:

$$z^* = \begin{bmatrix} Q_E & 0 \\ 0 & Q_F \end{bmatrix} \beta^* \tag{6}$$

for some $2n$-dimensional vector β^*. Under this change of variables, we obtain an optimization problem over $2n \ll p_E + p_F$ parameters. This allows us to solve (4) very efficiently using iteratively reweighted least squares (IRLS).

3 Results

We applied our method to a simultaneous EEG-fMRI dataset of an event-related three-alternative forced choice (3-AFC) visual discrimination task. On each trial, an image of either a face, car, or house was presented and subjects were instructed to respond with the category of the image by pressing one of three

buttons on an MR compatible button controller. The difficulty of the task was modulated ("easy" or "hard") by systematically affecting the salience of the image via randomization of image phase [5]. The task difficulty was observed behaviorally both in reaction time and accuracy (mean reaction time: 646ms (777ms) for easy (hard) trials; mean accuracy: 92% (57%) for easy (hard) trials). Over 4 runs, a total of 720 trials were acquired (240 of each category with 120 "easy" trials) with a random interstimulus interval (ISI) sampled uniformly between 2-2.5s. Twenty subjects participated in this experiment, and all data were collected on a Philips 3T Achieva MR scanner with EEG simultaneously recorded using a custom built 43 channel MR compatible EEG system [6–8]. Standard preprocessing were applied to both imaging datasets [9].

To format the multimodal dataset, we first associated each trial with a set of data from each modality. For the EEG data, this was achieved by taking the channel time-series data over the 600ms period after each stimulus onset. At a sampling frequency of 250Hz, this resulted in 6493 EEG features (43 channels $\times$ 151 time points). The data were assembled into a 6493×720 matrix E, where each row corresponds to the voltage level of a single channel at some time relative to stimulus onset, and columns index trials.

Associating fMRI data to each trial is more challenging for two main reasons: (a) the temporal dynamics of the hemodynamic response function (HRF) evolve over a longer time-scale than the mean ISI of the event-related design, resulting in overlapping responses between adjacent trials; and (b) the ISI was random for each trial so that the fMRI data was not acquired at a common lag relative to stimulus onset. To overcome these issues, we employed the LS-S deconvolution method proposed in [10]. For every trial, the time-series of each voxel is regressed against a "signal" regressor and a "noise" regressor. The "signal" regressor is the modeled HRF response due to that trial (a delta function centered at stimulus onset convolved with a canonical HRF), while the "noise" regressor is the modeled HRF response due to all other trials (superimposed linearly). The resulting regression coefficients of the "signal" regressor represent the estimated voxel activations due to that trial. These voxel activations can then be organized into a single brain volume per trial. We extracted 58697 voxels from a common group mask at 3mm^3 spatial resolution that excluded white matter and CSF and assembled the resulting voxel activations into rows of the data matrix F.

We were interested in identifying EEG-fMRI components that correlated with task difficulty regardless of the presented image category, so we assigned our trial label vector y to categorize the difficulty of each trial (0=difficult, 1=easy). We then trained our fusion algorithm on each subject separately and tuned our parameters using 5-fold cross-validation over a grid of 100 λ values on a log scale and 11 α values linearly spaced between 0 and 1. Within each fold, the regularization matrix L_α and the maximum singular value parameter $\sigma_{\max}$ were computed using only the training EEG and fMRI data.

Typically, cross-validation is used to select the model that maximizes some objective of interest, e.g., classification accuracy. In our case, model selection is not well-defined because we have two objectives: (i) classification accuracy,

as measured by cross-validated area under the ROC curve (Az); and (ii) cross-validated inter-modality correlation (ρ). The candidate set of "best" models are those that lie on the Pareto frontier, but selecting a single winner from this set depends on a specific tradeoff between the two objectives. For our purposes, the best model was selected that minimized the distance to the optimal $(\mathrm{Az}, \rho) = (1, 1)$ point. We have used similar concepts previously in fMRI analysis for evaluating prediction accuracy vs. reproducibility tradeoffs (see, e.g., [11]).

To evaluate the effect of modality linking, we used the criterion described above to select the best model in both the unlinked ($\alpha = 0$) and linked ($\alpha > 0$) cases. Table 1 presents the average cross-validated classification accuracy, inter-modality correlation ρ, as well as the average inter-modality partial correlation given the trial labels y. Using a Wilcoxon signed-rank test, the Az values were not significantly different, while the linked method showed a significant increase in both ρ and ρ given y ($p < 0.003$ and $p < 0.02$, respectively).

Method	Az	ρ	ρ given y
Unlinked ($\alpha = 0$)	0.65 ± 0.02	0.09 ± 0.02	0.05 ± 0.01
Linked ($\alpha > 0$)	0.62 ± 0.02	0.14 ± 0.02	0.10 ± 0.02

Table 1. Cross-validated results for the unlinked ($\alpha = 0$) and linked ($\alpha > 0$) cases. Values are averages over 20 subjects ($\pm$ standard error).

The significant inter-modality partial correlation suggests that the algorithm is identifying patterns across the modalities whose correlation is not fully explained by mean category activation differences. Instead, the correlation is also driven by common trial-to-trial modulations that may arise due to a latent brain state, e.g., attention, which is not measured by the experimental paradigm.

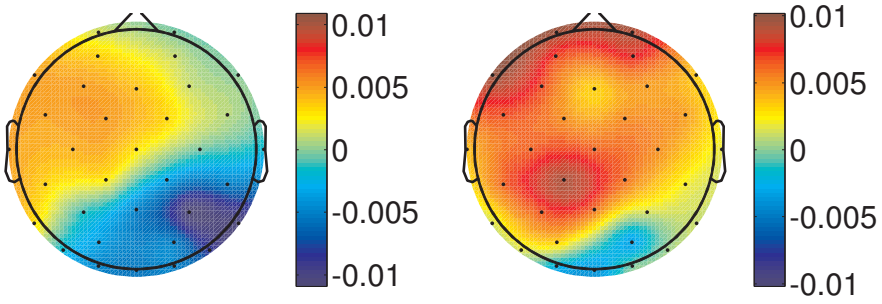
We also evaluated the spatio-temporal EEG-fMRI maps learned by the linking method. We converted feature weights to z-scores by averaging the maps over subjects and dividing by the standard error. We then thresholded the resulting maps at $p < 0.005$ (uncorrected). Due to the orientation of the trial label vector y, positive feature weights indicate higher mean activation on easy trials.

In the EEG case, significant weights were largely contained to an early temporal window between 150-240ms post-stimulus, and a later window between 400-600ms post-stimulus. Scalp map plots of the mean weights of both components are given in Figure 1. In the fMRI case, we found large positive clusters, including one that extended from Temporal Fusiform Cortex and Parahippocampal Gyrus to Lateral Occipital Cortex (LOC). We also found large positive clusters in Frontal Orbital Cortex and the Right Superior Frontal Gyrus (see Figure 2(a)). Negative clusters were identified in frontal areas commonly associated with decision making, including Paracingulate Gyrus and bilateral Insular Cortex (see Figure 2(b)).

With the exception of the Right Superior Frontal Gyrus, the unlinked method identified very similar areas. This suggests that the increased inter-modality correlation is explained less by identifying new areas, and more by altering the fine-grained patterns within the networks associated with task difficulty.

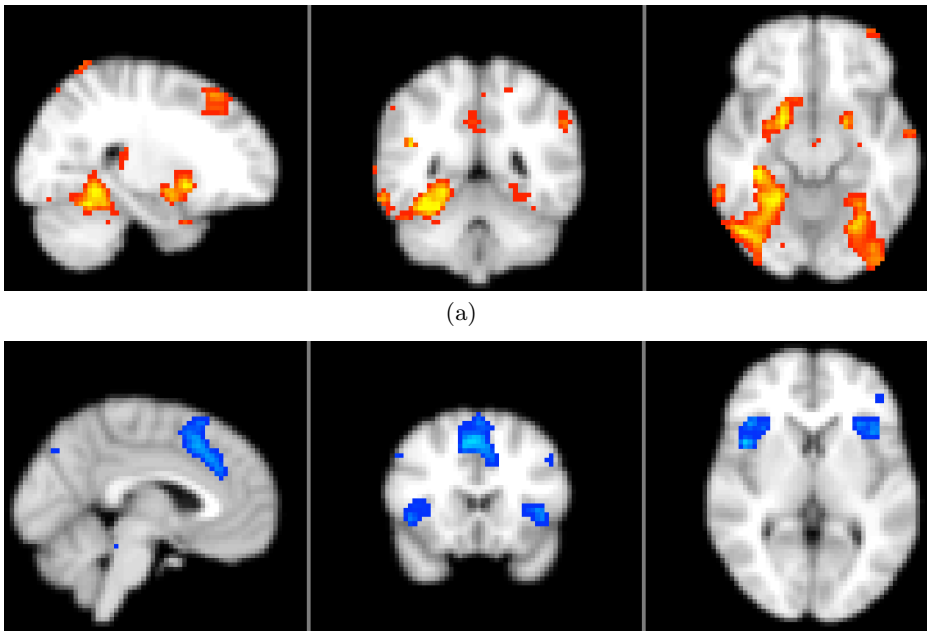
4 Conclusion

We formulated multi-modal data fusion as a multi-objective optimization problem that seeks to identify components with high classification accuracy and high inter-modality correlation. We cross-validated our approach on a real simultaneous EEG-fMRI dataset and identified EEG and fMRI components that were predictive of task difficulty in a visual perceptual decision-making task.



(a) (b)

Fig. 1. Scalp map plots of EEG component in (a) the early 150-240ms time window; and (b) the late 400-600ms time window.



(a)

(b)

Fig. 2. Brain maps of the fMRI component. (a) Positive weight clusters (MNI coordinates: (66,135,60)); (b) Negative weight clusters (MNI coordinates: (93,144,75)).

References

1. S. Debener, M. Ullsperger, M. Siegel, K. Fiehler, D. von Cramon, and A.K. Engel. Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 25(50):11730–11737, December 2005.
2. T. Eichele, K. Specht, M. Moosmann, M.L.A. Jongsma, R.Q. Quiroga, H. Nordby, and K. Hugdahl. Assessing the spatiotemporal evolution of neuronal activation with single-trial event-related potentials and functional MRI. *Proceedings of the National Academy of Sciences of the United States of America*, 102(49):17798–17803, December 2005.
3. R.I. Goldman, C.-Y. Wei, M.G. Philiastides, A.D. Gerson, D. Friedman, T.R. Brown, and P. Sajda. Single-trial discrimination for integrating simultaneous EEG and fMRI: identifying cortical areas contributing to trial-to-trial variability in the auditory oddball task. *NeuroImage*, 47(1):136–147, August 2009.
4. M. Moosmann, T. Eichele, H. Nordby, K. Hugdahl, and V.D. Calhoun. Joint independent component analysis for simultaneous EEG–fMRI: Principle and simulation. *International Journal of Psychophysiology*, 67(3):212–221, March 2008.
5. S.C. Dakin. What causes non-monotonic tuning of fMRI response to noisy images? *Curr Biol*, 12:476–477, 2002.
6. M. Dyrholm, R. Goldman, P. Sajda, and T.R. Brown. Removal of bcg artifacts using a non-kirchhoffian overcomplete representation. *IEEE Transactions on Biomedical Engineering*, 56:200–204, 2009.
7. R.I. Goldman, C.Y. Wei, M.G. Philiastides, A.D. Gerson, D. Friedman, T.R. Brown, and P. Sajda. Single-trial discrimination for integrating simultaneous EEG and fMRI: Identifying cortical areas contributing to trial-to-trial variability in the auditory oddball task. *NeuroImage*, 47:136–147, 2009.
8. P. Sajda, R.I. Goldman, M.G. Philiastides, A.D. Gerson, and T.R. Brown. A system for single-trial analysis of simultaneously acquired EEG and fMRI. *3rd International IEEE EMBS Conference on Neural Engineering*, 2007.
9. J.M. Walz, R.I. Goldman, J. Muraskin, T.R. Brown, and P. Sajda. Simultaneous EEG-fMRI Reveals a Superposition of Task-Dependent and Default-Mode Networks During a Simple Target Detection Task. *NeuroImage (In Press)*, 2013.
10. J.A. Mumford, B.O. Turner, F.G. Ashby, and R.A. Poldrack. Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analysis. *NeuroImage*, pages 2636–2643, 2012.
11. B.R. Conroy, J.M. Walz, and P. Sajda. Fast bootstrapping and permutation testing for assessing reproducibility and interpretability of multivariate fmri decoding models. *PLoS ONE*, 8:e79271.