Semi-Spatiotemporal fMRI Brain Decoding

Mohammad Hadi Kefayati*, Hamid Sheikhzadeh†, Hamid R. Rabiee‡, Ali Soltani-Farani†
*Dept. Electrical Engineering, Amirkabir University of Technology, Tehran, Iran.
†Dept. Computer Engineering, Sharif University of Technology, Tehran, Iran.
mhkefayati@aut.ac.ir, hsheikh@aut.ac.ir, rabiee@sharif.edu, a_soltani@ce.sharif.edu

Abstract—Functional behavior of the brain can be captured using functional Magnetic Resonance Imaging (fMRI). Even though fMRI signals have temporal and spatial structures, most studies have neglected the temporal structure when inferring mental states (brain decoding). This has two main side effects: 1. Degradation in brain decoding performance due to lack of temporal information in the model, 2. Inability to provide temporal interpretability. Few studies have targeted this issue but have had less success due to the burdening challenges related to high feature-to-instance ratio. In this study, a novel model for incorporating temporal information while maintaining a low feature-to-instance ratio, is proposed. Experimental results show the effectiveness of the model compared to recent state of the art approaches.

Keywords—Brain decoding; Sparsity; Spatiotemporal; fMRI.

I. INTRODUCTION

Decoding mental states or more fashionably mind reading is the process of determining a person’s mental state form the observed neural activities in his/her brain. This application has drawn great attention recently and regardless of all its limitations, it is no longer a fantasy [1]. Functional MRI has made it possible to capture brain activities with a reasonable spatial and temporal resolution, leading to the ability to discriminate between mental states. In the machine learning literature, decoding mental states is accomplished by applying a classifier on fMRI data. Despite the simplicity of linear classifiers, the excessively high feature-to-instance ratio has made them the major choice for brain decoding, while the demand for model interpretability has advanced the use of sparse regularization in these studies [2]–[13].

There are two approaches to model the inherently spatiotemporal fMRI data for brain decoding: 1. Pure spatial models where each brain volume is considered as a feature, neglecting the temporal structure of the data [2]–[8]. 2. Spatiotemporal models where a temporal series of brain volumes are considered as a feature [10], [11], [14]. Most studies take a purely spatial view since the high feature-to-instance ratio challenge becomes less severe (∼10 times). Even though these models are more interpretable, this interpretability is limited to the spatial structure since the inherent temporal structure of the data is neglected. Denoting the number of volumes in each trial by n and the number of voxels in each volume by k, pure spatial methods use k-dimensional features while spatiotemporal methods concatenate all volumes in a trial and use nk-dimensional features. Hence, spatiotemporal models are highly burdened by the high feature-to-instance ratio. This can degrade the classification performance and will likely result in an overfitted classifier.

In this paper, we propose to use an optimal temporal linear combination of the brain volumes in a trial, thus reducing the number of effective parameters to n+k by separating temporal and spatial weights. The proposed method has two main advantages: 1. Any spatial regularizer can readily be used to incorporate spatial knowledge and the model provides a means for incorporating temporal prior knowledge (e.g. smoothness, sparsity) via defining a temporal regularization function. 2. By separating the temporal and spatial weights, the model provides both temporal and spatial interpretability. Our results support the effectiveness of this approach for temporal smooth regularization. Further evaluation of this method with other temporal regularization priors is under study.

The remainder of this paper is organized as follows. In Section II we review the brain decoding problem along with pure spatial and spatiotemporal techniques. We explain our model in Section III. Section IV presents evaluation results and we conclude the paper in Section V.

II. BRAIN DECODING PROBLEM

The linear classification model for brain decoding is \( y = f(Xw) + e \), where \( y \) is the vector of mental states and \( X \) is a matrix containing fMRI data samples in each row. The function \( f \) resembles any simple mapping to the label space. The goal is to determine the weight vector \( w \) to minimize the prediction error \( e = y - f(Xw) \), while preserving known or desired properties such as good interpretability. The problem of finding \( w \) may be formulated as:

\[
    w^* = \arg \min_w \frac{1}{2} \| y - Xw \|_2^2 + \lambda \Omega(w) \tag{1}
\]

where the first term is the model error. \( \Omega(w) \) includes the prior knowledge of the problem and regularizes the model. Based on wether each brain volume is considered as an instance (Pure Spatial) or a series of brain volumes are concatenated to form an instance (Spatiotemporal), the prior knowledge included in \( \Omega(w) \) is different. We next review the related works.

A. Pure Spatial Techniques

The most commonly used regularizers for brain decoding are sparsity inducing regularizers. The Least Absolute
Shrinkage and Selection Operator (LASSO) or $\ell_1$ penalty is a typical choice which solves (1) using $\Omega(w) = \lambda \|w\|_1$. This results in a sparse vector $w$ [2], [12], which improves model interpretability. However, the inherent noise in fMRI signals, usually results in an unstable model. To seek a more stable model, authors of [4] add an $\ell_2$ penalty to the regularizer ($\Omega(w) = \lambda_1 \|w\|_1 + \lambda_2 \|\Gamma w\|_2^2$). This model, called Elastic Net (EN), improves the generalization performance of LASSO but results in a less sparse weight vector. EN and LASSO lack the ability to incorporate any structure in the weight vector such as spatially smooth variation over the brain volume. Generalized Sparse Classifiers (GSC) [3] are an extension to EN, that enable structured classifier weight patterns (e.g. spatially smooth patterns). The regularization term for GSC is: $\Omega(w) = \lambda_1 \|w\|_1 + \lambda_2 \|\Gamma w\|_2^2$, where $\Gamma$ is a matrix similar to the graph laplacian matrix [15] that introduces smoothness in the model. The same approach has been applied to group sparse models and is known as Generalized Group Sparse Classifiers (GGSC) [6], which are more capable in modeling implicit structures of the classification weight vector $w$. Spatial smoothness has been extended to wider relations and connectivity constraints in [13]. Long range connectivities are derived from the Diffusion Tensor Imaging (DTI) data and are considered as prior information to obtain a more accurate GSC model. GraphNet [8] is a similar approach to GSC but with a Tikhonov smoothness penalty on the classifier’s weight pattern. An $\ell_1$ regularization along with total variation regularization is proposed in [7] to obtain sparse and spatially smooth classification weights. All these methods are built by considering a single fMRI volume as an instance and we refer to them as pure spatial techniques.

**B. Spatiotemporal Techniques**

Beside the spatial structure, the temporal structure of fMRI data can also provide discriminative information. The effectiveness of considering these structures have been demonstrated in [16], where only the temporal information was utilized for separating drug addicted subjects from healthy controls. Inspired by this work, researchers have tried to exploit the spatiotemporal nature of fMRI data to build their models [11]. To model the spatiotemporal structure of fMRI data and benefit from both discriminative qualities at once, [11] proposed a reformulation of GSC, by augmenting the GSC with a temporal smoothness prior. Even though this model is able to capture temporal variations, it still fails to provide a unified overview of the general temporal variations in the fMRI signal and therefore, lacks the ability to provide a single interpretable spatial weight vector that reflects the significance of each voxel, which was readily provided by pure spatial techniques.

**III. PROPOSED SEMI-SPATIOTEMPORAL MODEL**

To take the temporal structure of fMRI data in to account and yet preserve the feature dimension, we make two assumptions: 1. Regions of the brain that are not related to a particular activity, remain inactive over time. 2. Activated regions of the brain have a similar temporal activation pattern, which we model by assigning a weight to each brain volume in a trial. Although our second assumption may not always hold [17], the spatial adjacency of activated regions makes this aberration negligible. Hence, we propose the following model:

$$y = \sum_{i=1}^{n} \beta_i X_i w + c$$

(2)

where $\beta$ is the vector of temporal weights, and $\beta_i$ denotes the weight assigned to the $i^{th}$ brain volume. $X_i$ is a matrix containing the $i^{th}$ brain volume of each trial in a row, and similar to pure spatial techniques, $w$ is the vector of spatial weights. Based on the brain decoding formulation of (1), we propose the following optimization to find the effective parameters of our model:

$$\{w^*, \beta^*\} = \arg \min_{w, \beta} \frac{1}{2} \|y - \sum_i \beta_i X_i w\|_2^2 + \lambda \Omega(w) + \gamma \Psi(\beta)$$

(3)

where $\Omega(w)$ may be any of the spatial regularizers discussed in Section II-A, and $\Psi(\beta)$ includes knowledge about the temporal variation of the activated brain regions (e.g. temporal smoothness). $\lambda$ and $\gamma$ control the amount of regularization. Therefore, the proposed model enjoys the benefits of pure spatial models while providing temporal interpretability. We name this model semi-spatiotemporal to emphasize the assumptions used in defining the model and the difference of our model from the existing spatiotemporal models.

**A. Optimization Procedure**

The prediction error in (3) is biconvex in $w$ and $\beta$. Hence, we use block coordinate descent and iterate between updating $w$ and $\beta$, based on the following steps:

1) **Minimization with respect to $w$:** For a fixed $\beta$ define:

$$X_\beta = \sum_{i=1}^{n} \beta_i X_i$$

(5)

The size of $X_\beta$ is equal to the size of $X$ in (1) (i.e. a single volume). We can therefore rewrite Step 1 as:

$$w^* = \arg \min_w \frac{1}{2} \|y - X_\beta w\|_2^2 + \lambda \Omega(w)$$

(6)

Since this step is similar to (1) for pure spatial techniques, we may benefit from the results and optimization techniques used in any of those frameworks. Here, we focus mainly on the second step of (4).
2) Minimization with respect to $\beta$: To simplify the formulation of Step 2, let us define:

$$A = \begin{bmatrix} X_1w & X_2w & \cdots & X_nw \end{bmatrix}$$ \hspace{1cm} (7)

where $n$ is the number of brain volumes in a trial, equal to the length of $\beta$. We can now rewrite Step 2 as:

$$\beta^* = \arg \min_{\beta} \frac{1}{2} \|y - A\beta\|_2^2 + \gamma \Psi(\beta)$$ \hspace{1cm} (8)

The optimization technique may vary based on the chosen regularizer for $\beta$. Let us consider the following simple yet reasonable regularizer to induce smoothness over $\beta$:

$$\Psi(\beta) = \sum_{i=2}^{n} (\beta_i - \beta_{i-1})^2 = \|L\beta\|_2^2 \hspace{1cm} (9)$$

where $L$ is given by:

$$L = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\
0 & -1 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & -1 & 1 \end{bmatrix}$$ \hspace{1cm} (10)

We solve the optimization for this particular case, although other regularizers can be similarly used. For $\Psi$ in (9) this step can be further simplified to:

$$\beta^* = \arg \min_{\beta} \frac{1}{2} \|y - A\beta\|_2^2 + \gamma \|L\beta\|_2^2$$ \hspace{1cm} (11)

Setting the derivative of the objective function to zero, we arrive at the following closed form solution:

$$\beta^* = (A^\top A + 2\gamma L^\top L)^{-1} A^\top y$$ \hspace{1cm} (12)

IV. Evaluation

A. Dataset

The fMRI data used in this study is the publicly available StarPlus fMRI dataset [18]. Here, we provide a brief description of the data and leave further details to [14], and the dataset web page. The dataset includes 13 healthy subjects performing 40 trials of picture/sentence matching task (the data for only 6 subjects is available online and each stimuli is classified either as a sentence or a picture). On each trial, two sequential stimuli (sentence or picture) are presented and the subject determines whether the two different stimuli match or not. The paradigm is an event-related design, and the length of each stimulus is 4 sec. The stimuli and trials are separated by a 4 and 15 seconds blank screen, respectively. The data is acquired with high field (3 Tesla) EPI every 0.5 sec with the 3*3*5 mm voxel dimension. Data preprocessing includes motion-correction and temporal detrending for removing head motion, low frequency signal drifts, and other artifacts of data acquisition. Spatial normalization was not performed and this preprocessed data was used as input to the classifier.

B. Experiments

To evaluate the proposed model, we compared our classification results against the Spatiotemporal Generalized Sparse Classifier (ST-GSC) [11], the spatiotemporal setting with an $\ell_1$ penalty (L1) and an Elastic Net penalty (EN), and a linear support vector machine (SVM). For our method, we report results for an $\ell_1$ spatial regularizer (SST-L1) and the Elastic Net spatial regularizer (SST-EN). To show the significance of optimizing the temporal weights, we have also reported our results for $\beta_i = \frac{1}{n}$ (i.e. average of brain volumes), denoted by L1+Avg and EN+Avg for $\ell_1$ and Elastic Net penalties, respectively. For all the methods, the parameters were determined using a 10-fold nested cross validation (NCV). To account for the delay in the hemodynamic response, similar to [11], we considered the second half of a trial (resulting samples include 8 sequentially obtained brain volumes) in each sample. The range of lambda and gamma in NCV were $\max |X'y| \ast [10^{-5}, 10^{-4}, ..., 10^{-2}]$ and $\max |X'y| \ast [10^{-5}, 10^{-4}, ..., 10^{0}]$, respectively. The classification results are reported in Fig. 1, which also shows the obtained standard deviations from the mean accuracy. Spatial and temporal weights for a sample model that are obtained using SST-EN are depicted in Fig. 2.

C. Discussion

Our results showed improvement over the recently proposed spatiotemporal sparse brain decoding (ST-GSC) Scheme [11]. We would like to point out that the computational cost of our method is by far less than [11], since the number of parameters that need to be optimized have been reduced by a factor of $n$. One should also consider that we have not incorporated any spatial relation into our regularizer, although our model can support other spatial regularizers. Our results showed improvement over all other methods both in classification accuracy and standard deviation. The results depicted in Fig. 1, highlight the fact
that given spatiotemporal data, even naively averaging the brain volumes can significantly improve the accuracy. Other studies such as [6], have shown that the visual cortex is the main activated region of the brain for the activities in this dataset. The spatial weights illustrated in Fig. 2(a) confirm this observation.

V. CONCLUSION

In this paper, we presented a novel scheme to extend vastly used spatial methods for brain decoding to the spatiotemporal setting. To construct the proposed model, we made two assumptions: 1. Regions of the brain that are not related to a particular activity, remain inactive over time, and 2. Activated regions of the brain have a similar temporal activation pattern. Then, we proposed to use a temporal linear combination of the brain volumes in a trial. Our semi-spatiotemporal model has several advantages: 1. Compared to spatiotemporal techniques, the number of parameters that need to be optimized has been effectively reduced by a factor equal to the number of volumes in a trial, 2. Previously defined spatial regularizers can readily be used to incorporate spatial prior knowledge, 3. The model provides a means for incorporating temporal prior knowledge, and 4. By separating the temporal and spatial weights the model provides both temporal and spatial interpretability. Experimental results showed the effectiveness of the proposed model with an $\ell_1$ or Elastic Net spatial regularizer, and a temporal regularizer that induces smoothness over time. We are currently evaluating the model with more complex spatial and temporal regularizers.

ACKNOWLEDGMENT

The authors would like to thank Dr. Bernard Ng for helpful discussions on the spatiotemporal generalized sparse classifiers. This work was supported in part by the AICT Research Center.

REFERENCES


