Graph-Based Decoding Model for Functional Alignment of Unaligned fMRI Data

Weida Li,[1] **Mingxia Liu,**[1,2*] **Fang Chen,**[1] **Daoqiang Zhang**[1*]
[1]College of Computer Science and Technology & MIIT Key Laboratory of Pattern Analysis and Machine Intelligence, Nanjing University of Aeronautics and Astronautics, Nanjing, China
[2]School of Information Science and Technology, Taishan University, Taian, China
[*]Corresponding Authors: {mingxialiu, dqzhang}@nuaa.edu.cn

Abstract

Aggregating multi-subject functional magnetic resonance imaging (fMRI) data is indispensable for generating valid and general inferences from patterns distributed across human brains. The disparities in anatomical structures and functional topographies of human brains warrant aligning fMRI data across subjects. However, the existing functional alignment methods cannot handle well various kinds of fMRI datasets today, especially when they are not temporally-aligned, i.e., some of the subjects probably lack the responses to some stimuli, or different subjects might follow different sequences of stimuli. In this paper, a cross-subject graph that depicts the (dis)similarities between samples across subjects is used as a priori for developing a more flexible framework that suits an assortment of fMRI datasets. However, the high dimension of fMRI data and the use of multiple subjects makes the crude framework time-consuming or unpractical. To address this issue, we further regularize the framework, so that a novel feasible kernel-based optimization, which permits nonlinear feature extraction, could be theoretically developed. Specifically, a low-dimension assumption is imposed on each new feature space to avoid overfitting caused by the high-spatial-low-temporal resolution of fMRI data. Experimental results on five datasets suggest that the proposed method is *not only* superior to several state-of-the-art methods on temporally-aligned fMRI data, *but also* suitable for dealing with temporally-unaligned fMRI data.

Introduction

Functional Magnetic Resonance Imaging (fMRI) is an imaging technology used to measure neural activity by using the blood-oxygen-level-dependent (BOLD) contrast as an indicator for cognitive states (Logothetis 2002). The informative patterns encoded in fMRI enable investigators to study how the human brain works (Haxby, Connolly, and Guntupalli 2014). Specifically, the use of multi-subject fMRI data is indispensable for accessing the validity and generality of the findings across subjects (Talairach and Tournoux 1988; Watson et al. 1993). From another angle, aggregating multi-subject fMRI data is also critical due to the high-spatial-low-temporal (HSLT) resolution of fMRI, i.e., the number of samples (time points or volumes) is generally much smaller than the number of features (voxels) per subject. However, such aggregation is facing a challenge that both anatomical structure and functional topography vary across subjects (Haxby et al. 2011). Hence, inter-subject alignment is an indispensable step in fMRI analysis.

Existing studies for inter-subject alignment include anatomical alignment and functional alignment, which can work in unison. In fact, anatomical alignment is usually used as a preprocessing step for fMRI analysis, by aligning anatomical features based on structural MRI images across subjects. Typical examples include Talairach alignment (Talairach and Tournoux 1988), cortical surface alignment (Fischl et al. 1999) and so on. However, anatomical alignment generated limited accuracy since the size, shape and anatomical location of functional loci differ across subjects (Watson et al. 1993; Rademacher et al. 1993). In contrast, functional alignment tries to directly align functional responses across subjects (Sabuncu et al. 2009; Conroy et al. 2009). As more radical approaches of functional alignment, Hyperalignment (Haxby et al. 2011) and Shared Response Model (SRM) (Chen et al. 2015) learn implicitly shared patterns across subjects, which are closely related to multi-view Canonical Correlation Analysis (CCA). Though both of them have been extensively studied and extended to an assortment, the existing related studies assume that the given fMRI datasets should be temporally-aligned across subjects (Chen et al. 2015; Turek et al. 2018; Xu et al. 2012). In other words, the sequential fMRI time points of each subject have to correspond to the same sequence of stimuli, like all subjects watching a movie together. Such a demand makes them not flexible enough as fMRI datasets today could be not temporally-aligned. For example, some subjects probably lack the responses to some stimuli, or different subjects may respond to different sequences of stimuli. Even though this problem could be somewhat solved by reordering and truncating (or downsampling) the dataset to generate an aligned version (Chen et al. 2015), these processes may lead to an inevitable loss of information. A recent study tries to extend SRM into a semi-supervised one by exploiting labeled samples, the unlabeled samples are, however, required to be temporally-

aligned (Turek et al. 2017).

In this paper, we aim to develop an adaptable functional alignment framework by using a cross-subject graph that depicts the (dis)similarities between all samples as a priori. Such a graph can be generated according to samples' category labels or through inference (De Sa et al. 2010). From this perspective, we can then focus on the (dis)similarity, which is encoded in a graph, between any two samples rather than merely caring about if the given fMRI dataset is temporally-aligned. However, the crude framework is unpractical as the related matrices are too large to be used, which is caused by the high dimension of fMRI data and the use of multiple subjects. To address this problem, the unrefined framework is regularized so that a novel feasible kernel-based optimization, which allows for non-linear feature extraction, could be theoretically set up. With such a regularization, the optimal solution is, sometimes, unique. Nevertheless, the high-spatial-low-temporal (HSLT) resolution of fMRI data causes that the generated optimal solution could indicate overfitting, i.e., it aligns all aligning samples perfectly. In a specific case, the culprit is that the dimension of the subspace spanned by the aligning samples equals to the number of them. Therefore, a low-dimension assumption, which agrees with that the number of informative features is generally less than the number of voxels, is imposed on each new feature space to avoid overfitting. The refined framework, together with the proposed optimization method, is referred to as Graph-based Decoding Model (GDM) in this paper. Notably, the objective function of Hyperalignment is the same as that of our GDM (with an evident graph). The main contributions of this paper are summarized as follows:

i) Unlike previous studies that rely on temporally-aligned data, GDM does not require temporal alignment for fMRI data. Once the prior information of the (dis)similarities among samples is available or can be inferred, one can employ our GDM to solve fMRI-based problems at hand.

ii) Different from the conventional naive kernel implementation, the computational time of our proposed kernel-based optimization (naturally accompanied by a low-dimension assumption) is faster on the number of samples, making it suitable for processing large-scale dataset.

iii) The feasible kernel-based optimization method with the low-dimension assumption is equipped with some theoretical guarantees.

In the following, we first briefly review related works, and concisely mention the notation and problem statement. Then, the proposed method GDM will be introduced in detail. We further introduce the materials used in this work, experimental setup, competing methods, and experimental results achieved by different methods on both aligned and unaligned datasets, which is followed by a Conclusion section. Related proofs and additional experimental results are given in the *Supplementary File*.

Related Works

The initial Hyperalignment (HA) method aims to seek implicitly shared features across subjects (Haxby et al. 2011), which is based on the orthogonal Procrustes problem. It is the first that links functional alignment and multi-view CCA. The performance of Hyperalignment on fMRI analysis is dramatically increased compared with any other anatomical alignment methods. To tackle the singularity caused by the HSLT resolution of fMRI, Regularized Hyperalignment (RHA) was developed by Xu *et al.* (Xu et al. 2012).

However, neither of HA nor RHA can handle full-brain data. To address this issue, there have been several works: Chen *et al.* developed a Singular Vector Decomposition Hyperalignment (SVDHA), which firstly carries out a joint-SVD by grouping all subjects' fMRI data for dimension reduction across subjects (Chen et al. 2014). Later, Chen *et al.* introduced a Shared Response Model (SRM) which can be modeled from probabilistic perspective by assuming that each sample from the latent common space has undergone a Gaussian noise disturbance (Chen et al. 2015). Solely linear feature extraction was considered until that Kernel Hyperalignment (KHA) was formulated by Lorbert and Ramadge (Lorbert and Ramadge 2012). Since fMRI dataset may partially contain labels, a semi-supervised scheme based on SRM was studied by Turek *et al.* (Turek et al. 2017).

On the other hand, a Searchlight approach, which takes functional alignment method as a module, was established to enhance functional alignment further by assuming that any voxel is only in connection with voxels in its anatomical vicinity (Guntupalli et al. 2016). Recently, a Robust SRM that accounts for individual variations was developed by Turek *et al.* (Turek et al. 2018).

Notation and Problem Statements

Notation In this paper, the bold letters are reserved for matrices (upper) or vectors (lower), whereas the plain are for scalars. Given any sequence of matrices $\{\mathbf{A}_i\}_{i=1}^M$, let $\mathbf{A}_*$ be the corresponding block diagonal matrix whose diagonal matrices are $\{\mathbf{A}_i\}_{i=1}^M$ from the top left to the bottom right. Plus, for any matrix $\mathbf{A}$, $\mathbf{a}_i$ refers to its i-th column vector, A_{ij} is its (i,j)-th entry, $\mathrm{R}(\mathbf{A})$ denotes the subspace spanned by the columns of $\mathbf{A}$ and $\mathrm{N}(\mathbf{A})$ is the null space of $\mathbf{A}$, i.e., $\{\mathbf{x} \,|\, \mathbf{A}\mathbf{x} = \mathbf{0}\}$. Moreover, any vector is treated as a column vector and the subscript of $\mathbf{A}_{I\times J}$ indicates its shape.

Let $\{\mathbf{X}_i \in \mathbb{R}^{V_i \times T_i}\}_{i=1}^M$ be an fMRI dataset where T_i and V_i are the number of samples (time points or volumes) and features (voxels) of the i-th subject, respectively, and M is the total number of subjects. Due to the HSLT resolution of fMRI, $T_i \ll V_i$. To develop a kernel-based method, we introduce a column-wised non-linear map Φ_i that maps each sample, e.g., each column of $\mathbf{X}_i$, of the i-th subject into a new feature space $\mathcal{H}_i$, which is a Hilbert space. Unlike Kernel Hyperalignment (Lorbert and Ramadge 2012), subject-specific kernels are allowed. Here, different kernels could be thought to account for different structures of human brain. For simplicity, denote $\mathbf{\Phi}_i$ by setting $(\boldsymbol{\phi}_i)_j = \Phi_i((\mathbf{x}_i)_j)$ for $1 \le j \le T_i$, and let $\mathbf{K}_i$ be $\mathbf{\Phi}_i^T\mathbf{\Phi}_i$. Plus, let K denote the number of the shared features across subjects.

Assumption for Theoretical Development Generally, the dimension of $\mathcal{H}_i$ could be infinite. For example, the reproducing kernel Hilbert space of Gaussian kernel is isomorphic to a subspace of $l_2(\mathbb{N})$ (Steinwart and Christmann 2008). For clarity in the development of the optimization, we assume that $\mathcal{H}_i$ is a finite dimensional real Hilbert space throughout the paper. The general lengthy proofs are left in the *Supplementary File*. Thus, $\Phi_i : \mathbb{R}^{V_i} \mapsto \mathbb{R}^{N_i}$ and $\mathbf{\Phi}_i \in \mathbb{R}^{N_i \times T_i}$ where N_i is the dimension of $\mathcal{H}_i$.

The goal is to learn aligning maps $\{f_i : \mathbb{R}^{V_i} \mapsto \mathbb{R}^K\}_{i=1}^M$ for each subject such that they map populations of subjects' fMRI responses into a shared space in which the disparities between subjects' brains are eliminated. Here, we aim to learn linear aligning maps $\{h_i : \mathbb{R}^{N_i} \mapsto \mathbb{R}^K\}$ with good generalization. Therefore, $f_i = h_i \circ \Phi_i$ and $h_i((\boldsymbol{\phi}_i)_j) = \mathbf{W}_i^T(\boldsymbol{\phi}_i)_j$ for $1 \leq j \leq T_i$ where $\mathbf{W}_i \in \mathbb{R}^{N_i \times K}$.

Proposed Method

Formulation

Cross-Subject Graph A graph about the (dis)similarities among all samples are mostly available. For example, the part of temporally-aligned samples, the category of each sample, or the distances between samples tell which samples are closely related or distinctive. To describe such (dis)similarities, let $\mathbf{G} \in \mathbb{R}^{T \times T}$ be a cross-subject graph matrix where $T = \sum_{i=1}^M T_i$ and G_{ij} indicates the (dis)similarity of the i-th and j-th samples, and thus $\mathbf{G}^T = \mathbf{G}$. Here, i or j could refer to any sample from any subject.

Objective Function Let $\mathbf{W}^T$ be $\begin{pmatrix}\mathbf{W}_1^T & \cdots & \mathbf{W}_M^T\end{pmatrix}$ and $\mathbf{Y}$ be $\mathbf{W}^T\mathbf{\Phi}_* = \begin{pmatrix}\mathbf{W}_1^T\mathbf{\Phi}_1 & \cdots & \mathbf{W}_M^T\mathbf{\Phi}_M\end{pmatrix}$. Since $\mathbf{Y} \in \mathbb{R}^{K \times T}$ contains all samples, the objective function can be expressed as

$$\underset{\mathbf{W}}{\operatorname{argmin}} \frac{1}{2}\sum_{i=1}^T\sum_{j=1}^T G_{ij}\left\|\mathbf{y}_i - \mathbf{y}_j\right\|_F^2 = \operatorname{tr}\left(\mathbf{Y}\mathbf{L}\mathbf{Y}^T\right) \quad (1)$$

where $\mathbf{L} = \mathbf{D} - \mathbf{G}$ is the Laplacian matrix of the graph matrix $\mathbf{G}$ (Chung and Graham 1997) and $\mathbf{D}$ is a diagonal matrix with $D_{ii} = \sum_{j=1}^T G_{ij}$. This objective function tries to separate the transformed samples $\mathbf{y}_i$ and $\mathbf{y}_j$ when $G_{ij} < 0$ but attempts to make them close when $G_{ij} > 0$.

Constraint Given a stimulus, suppose $\{\mathbf{z}_i \in \mathbb{R}^{V_i}\}_{i=1}^M$ are subjects' corresponding fMRI responses and the authentic aligning maps $\{f_i : \mathbb{R}^{V_i} \mapsto \mathbb{R}^K\}_{i=1}^M$ are already there. Since each subject's fMRI responses to the same stimulus behave like a random variable, $\{f_i(\mathbf{z}_i)\}_{i=1}^M$ are expected to be from the same shared random variable. In other words, we do not require that $f_i(\mathbf{z}_i) = f_j(\mathbf{z}_j)$ for any i, j. Therefore, the statistical constraint $\mathbf{Y}\mathbf{Y}^T = \mathbf{I}$ can be applied directly even if some samples are expected to be from the same latent response. The constraint means that each extracted shared feature is on the same scale and they are restricted be as uncorrelated as possible. The crude framework is

$$\begin{aligned}&\underset{\mathbf{W}}{\operatorname{argmin}} \operatorname{tr}\left(\mathbf{W}^T\mathbf{\Phi}_*\mathbf{L}\mathbf{\Phi}_*^T\mathbf{W}\right)\\ &\text{subject to } \mathbf{W}^T\mathbf{\Phi}_*\mathbf{\Phi}_*^T\mathbf{W} = \mathbf{I}\,.\end{aligned} \quad (2)$$

Relationship between GDM and Hyperalignment The Hyperalignment (HA) method is based on temporally-aligned dataset, assuming that $T_i = T_0$ for $i = 1, 2, \ldots, M$. Define a graph $\mathbf{G}_{HA}$ by setting $G_{ij} = 1/M$ when the i-th and j-th samples are aligned and $G_{ij} = 0$ otherwise. Then, $\left\|\mathbf{W}_i^T\mathbf{X}_i - \mathbf{S}\right\|_F^2$, which is the objective function of HA, is equal to $\operatorname{tr}\left(\mathbf{W}^T\mathbf{X}_*\left(\mathbf{I}_{MT_0\times MT_0} - \mathbf{G}_{HA}\right)\mathbf{X}_*^T\mathbf{W}\right)$ since

$$\begin{aligned}&\sum_{i=1}^M \|\mathbf{W}_i^T\mathbf{X}_i - \mathbf{S}^*\|_F^2\\ =&\sum_{i=1}^M \|\mathbf{W}_i^T\mathbf{X}_i\|_F^2 + \left(M\|\mathbf{S}^*\|_F^2 - 2\langle\sum_{i=1}^M \mathbf{W}_i^T\mathbf{X}_i, \mathbf{S}^*\rangle\right)\\ =&\operatorname{tr}\left(\mathbf{W}^T\mathbf{X}_*\mathbf{X}_*^T\mathbf{W}\right) - \operatorname{tr}\left(\mathbf{W}^T\mathbf{X}_*\mathbf{G}_{HA}\mathbf{X}_*^T\mathbf{W}\right)\\ =&\operatorname{tr}\left(\mathbf{W}^T\mathbf{X}_*\left(\mathbf{I}_{MT_0\times MT_0} - \mathbf{G}_{HA}\right)\mathbf{X}_*^T\mathbf{W}\right)\,.\end{aligned}$$

where the optimal $\mathbf{S}^*$ is $1/M\sum_{i=1}^M \mathbf{W}_i^T\mathbf{X}_i$.

Computational Cost Problem (2) is a generalized eigenvalue problem, which has been studied extensively. However, with linear kernel, the size of $\mathbf{X}_*\mathbf{L}\mathbf{X}_*^T$ or $\mathbf{X}_*\mathbf{X}_*^T$, which is $\left(\sum_{i=1}^M V_i\right)^2$, is too tremendous to be used. For example, the dataset DS001 used in our experiment includes 16 subjects with 19174 features per subject, and then it requires at least 350 GB to store $\mathbf{X}_*\mathbf{L}\mathbf{X}_*^T$ or $\mathbf{X}_*\mathbf{X}_*^T$ of shape $(16\times 19174)\times(16\times 19174)$, which is not affordable. Thus, an efficient feasible optimization is needed. The Proposition below is helpful for solving such an issue.

Proposition 1 If $\mathbf{W}$ is one solution for problem (2), then there must be another solution that belongs to $\mathrm{R}(\mathbf{\Phi}_*)$, and has the same objective value as $\mathbf{W}$.

Proof. $\mathbf{W}$ can be decomposed uniquely as $\mathbf{W} = \mathbf{W}_R + \mathbf{W}_N$ where $\mathbf{W}_R \in \mathrm{R}(\mathbf{\Phi}_*)$ and $\mathbf{W}_N \in \mathrm{N}(\mathbf{\Phi}_*^T)$. Since

$$\mathbf{W}^T\mathbf{\Phi}_* = \mathbf{W}_R^T\mathbf{\Phi}_* + \mathbf{W}_N^T\mathbf{\Phi}_* = \mathbf{W}_R^T\mathbf{\Phi}_* + \mathbf{0} = \mathbf{W}_R^T\mathbf{\Phi}_*\,,$$

Plugging $\mathbf{W}_R$ into problem (2) leads to that $\mathbf{W}_R$ satisfies the constraint and shares the same objective value with $\mathbf{W}$. □

Regularized Framework In Proposition 1, the trivial part $\mathbf{W}_N$ exists due to the HSLT resolution of fMRI, i.e., $T_i \ll V_i$. Such a trifling part indicates that it does not help produce a better solution, and thus there are many optimal solutions. If the trivial part is excluded by constraint, the optimal solution, sometimes, become unique, and a feasible optimization will be there. More details about the uniqueness are in-

cluded in the *Supplementary File*. In a nutshell, the regularized framework is expressed as

$$\begin{aligned} &\underset{\mathbf{W}}{\operatorname{argmin}} \operatorname{tr}\left(\mathbf{W}^T\mathbf{\Phi}_*\mathbf{L}\mathbf{\Phi}_*^T\mathbf{W}\right) \\ &\text{subject to } \mathbf{W}^T\mathbf{\Phi}_*\mathbf{\Phi}_*^T\mathbf{W} = \mathbf{I} \\ &\mathbf{w}_i \in \mathrm{R}(\mathbf{\Phi}_*) \text{ for } 1 \le i \le K\,. \end{aligned} \tag{3}$$

Optimization

Naive Kernel-Based Optimization A simple way to solve GDM in Eq. (3) is to let $\mathbf{W}_i$ be $\mathbf{X}_i\mathbf{B}_i$, where $\mathbf{B}_i$ is a new variable. Then, $\mathbf{W}^T\mathbf{X}_*\mathbf{L}\mathbf{X}_*^T\mathbf{W}$ becomes $\mathbf{B}^T\mathbf{K}_*\mathbf{L}\mathbf{K}_*\mathbf{B}$, where $\mathbf{B}$ is constructed like $\mathbf{W}$. The optimal solution of GDM can be achieved by solving a generalized eigenvalue problem. However, with any kernel, the complexity in terms of $\{T_i\}_{i=1}^M$ will be at least $O(T^3)$ where $T = \sum_{i=1}^M T_i$, meaning that it heavily depends on the number of samples. In the following, we propose a more efficient kernel-based optimization algorithm.

Proposed Kernel-Based Optimization Here are some tricks to solve problem (3). For each i, by spectral decomposition, $\mathbf{K}_i = \mathbf{V}_i\mathbf{D}_i\mathbf{V}_i^T$ where zero eigenvalues of $\mathbf{K}_i$ are excluded. With $\mathbf{U}_i = \mathbf{\Phi}_i\mathbf{V}_i\mathbf{D}_i^{-\frac{1}{2}}$, it leads to a Singular Vector Decomposition (SVD) of $\mathbf{\Phi}_i$ as

$$\mathbf{\Phi}_i = \mathbf{U}_i\mathbf{D}_i^{\frac{1}{2}}\mathbf{V}_i^T\,. \tag{4}$$

As shown in the *Supplementary File*, $\mathbf{\Phi}_i$ can be decomposed similarly when the dimension of $\mathcal{H}_i$ is infinite. Thus, the development below is without loss of generality. With Eq. (4), $\mathbf{\Phi}_* = \mathbf{U}_*\mathbf{D}_*^{\frac{1}{2}}\mathbf{V}_*^T$ and then problem (3) is equivalent to

$$\begin{aligned} &\underset{\mathbf{Q}}{\operatorname{argmin}} \operatorname{tr}\left(\mathbf{Q}^T\mathbf{V}_*^T\mathbf{L}\mathbf{V}_*\mathbf{Q}\right) \\ &\text{subject to } \mathbf{Q}^T\mathbf{Q} = \mathbf{I}\,. \end{aligned} \tag{5}$$

To see this, denote the shape of $\mathbf{D}_*$ by $S \times S$. Let $\mathcal{S}$ be $\{\mathbf{W} : \mathbf{W}^T\mathbf{\Phi}_*\mathbf{\Phi}_*^T\mathbf{W} = \mathbf{I}$ and $\mathbf{w}_i \in \mathrm{R}(\mathbf{\Phi}_*)$ for $1 \le i \le K\}$ and $\mathcal{T}$ be $\{\mathbf{Q} \in \mathbb{R}^{S\times K} : \mathbf{Q}^T\mathbf{Q} = \mathbf{I}\}$. Denote a map $g : \mathcal{S} \mapsto \mathcal{T}$ by setting $g(\mathbf{W}) = \mathbf{D}_*^{\frac{1}{2}}\mathbf{U}_*^T\mathbf{W}$. Since each column of $\mathbf{W}$ belongs to $\mathrm{R}(\mathbf{\Phi}_*) = \mathrm{R}(\mathbf{U}_*)$, $\mathbf{U}_*\mathbf{D}_*^{-\frac{1}{2}}\mathbf{D}_*^{\frac{1}{2}}\mathbf{U}_*^T\mathbf{W} = \mathbf{W}$, which in turn leads to that g is a bijection between $\mathcal{S}$ and $g(\mathcal{S}) = \mathcal{T}$. Plugging $\mathbf{W} = \mathbf{U}_*\mathbf{D}_*^{-\frac{1}{2}}\mathbf{Q}$ into problem (3) leads to problem (5).

Proposition 2 Using spectral decomposition, $\mathbf{V}_*^T\mathbf{L}\mathbf{V}_* = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^T$ where all eigenvalues of $\mathbf{V}_*^T\mathbf{L}\mathbf{V}_*$ along the diagonal of $\mathbf{\Lambda}$ from the top left to the bottom right are in ascending order. Denote the shape of $\mathbf{V}_*^T\mathbf{L}\mathbf{V}_*$ by $S \times S$. If $K \le S$, the first K columns of $\mathbf{E}$ is optimal for problem (5).

Proof. Firstly, problem (5) is equivalent to

$$\begin{aligned} &\operatorname{argmin} \operatorname{tr}\left(\mathbf{R}^T\mathbf{\Lambda}\mathbf{R}\right) \\ &\text{subject to } \mathbf{R}^T\mathbf{R} = \mathbf{I} \end{aligned} \tag{6}$$

where $\mathbf{R} = \mathbf{E}^T\mathbf{Q}$. As $\mathbf{R}^T\mathbf{R} = \mathbf{I}$ infers $\sum_{i=1}^S\sum_{j=1}^K R_{ij}^2 = K$ and $\sum_{j=1}^K R_{ij}^2 \le 1$ for each i, there is

$$\operatorname{tr}\left(\mathbf{R}^T\mathbf{\Lambda}\mathbf{R}\right) = \sum_{i=1}^S \Lambda_{ii}\sum_{j=1}^K R_{ij}^2 \ge \sum_{i=1}^K \Lambda_{ii}\,.$$

Let $\mathbf{R}^*$ denote $\begin{pmatrix}\mathbf{I}_{K\times K} & \mathbf{0}_{K\times(S-K)}\end{pmatrix}^T$. Since $\operatorname{tr}\left((\mathbf{R}^*)^T\mathbf{\Lambda}\mathbf{R}^*\right) = \sum_{i=1}^K \Lambda_{ii}$, $\mathbf{R}^*$ is optimal. Therefore, an optimal solution $\mathbf{Q}^* = \mathbf{E}\mathbf{R}^*$ for problem (5) is indeed the first K columns of $\mathbf{E}$. □

An Optimal Solution for Regularized Framework and Its Uniqueness Let $\hat{\mathbf{E}}$ denote the first K columns of $\mathbf{E}$ and take Eq. (4) into consideration, then an optimal solution for problem (3) is

$$\mathbf{W}^* = \mathbf{U}_*\mathbf{D}_*^{-\frac{1}{2}}\hat{\mathbf{E}} = \mathbf{\Phi}_*\mathbf{V}_*\mathbf{D}_*^{-1}\hat{\mathbf{E}}\,. \tag{7}$$

Since each $\mathbf{W}_i$ is separable from $\mathbf{W}$, an optimal solution for subject i is

$$\mathbf{W}_i^* = \mathbf{\Phi}_i\mathbf{V}_i\mathbf{D}_i^{-1}\hat{\mathbf{E}}_i, \tag{8}$$

where $\{\hat{\mathbf{E}}_i\}_{i=1}^M$ are block matrices of $\hat{\mathbf{E}}$, which is cut along the first dimension according to the dimensions of block matrices in $\mathbf{D}_*$.

By the equivalences above, if $K > S$, there is no solution satisfying the constraint in problem (3) or (6) as there is no $\mathbf{R}$ satisfying $\mathbf{R}^T\mathbf{R} = \mathbf{I}$. If $K = S$, or $K < S$ with $\Lambda_{KK} < \Lambda_{(K+1)(K+1)}$, the optimal solution of problem (3) is unique except being rotated. In other words, if $\mathbf{W}^{(1)}$ and $\mathbf{W}^{(2)}$ are two optimal solutions, there is an orthogonal matrix $\mathbf{P}$ such that $\mathbf{W}^{(1)} = \mathbf{W}^{(2)}\mathbf{P}$. By the definition of $\mathbf{W}$, it implies that the shared feature space is unique except being rotated. More details are given in the *Supplementary File*.

Low-Dimension Assumption

Potential Overfitting of GDM Suppose the dataset $\{\mathbf{X}_i\}_{i=1}^M$ is temporally-aligned, which means that $T_i = T_j = T_0$ for any i, j. Construct a graph matrix $\mathbf{G}$ by setting $G_{ij} = 1$ if the i-th and j-th samples are temporally-aligned, and $G_{ij} = 0$ otherwise. With this graph matrix, the objective function of problem (3) with linear kernel becomes

$$\underset{\mathbf{W}_i}{\operatorname{argmin}} \frac{1}{2}\sum_{i=1}^M\sum_{j=1}^M \left\|\mathbf{W}_i^T\mathbf{X}_i - \mathbf{W}_j^T\mathbf{X}_j\right\|_F^2\,. \tag{9}$$

Assume that each $\mathbf{X}_i \in \mathbb{R}^{V_i\times T_0}$ is full-column rank. Let $\mathbf{P}_{K\times T_0}$ ($K \le T_0$) be any matrix such that $\mathbf{P}\mathbf{P}^T = \mathbf{I}$ and take Eq. (4) into consideration where $\mathbf{\Phi}_i$ is replaced by $\mathbf{X}_i$. With $\mathbf{W}_i^* = M^{-1}\mathbf{U}_i\mathbf{D}_i^{-\frac{1}{2}}\mathbf{V}_i^T\mathbf{P}^T$ and $(\mathbf{W}^*)^T = \begin{pmatrix}(\mathbf{W}_1^*)^T & \cdots & (\mathbf{W}_M^*)^T\end{pmatrix}$, $\mathbf{W}^*$ satisfies the constraints in problem (3). However, $(\mathbf{W}_i^*)^T\mathbf{\Phi}_i = M^{-1}\mathbf{P}$ for each i, which implies that the generated optimal solution (8) aligns each aligning sample perfectly. The culprit is the full-column rank assumption of each $\mathbf{X}_i$, which is almost the case due to the HSLT resolution of fMRI, i.e., $T_0 \ll V_i$.

Therefore, we impose a low-dimension assumption over each new feature space $\mathcal{H}_i$, which conforms with that the number of informative features is usually much less than the number of voxels. Suppose the low-dimension in $\mathcal{H}_i$ is L_i, then we try to fit the data in $\mathcal{H}_i$ by an L_i dimensional affine subspace[1], i.e.,

$$\underset{\substack{\mathbf{m}_i\in\mathbb{R}^{N_i}\\ \mathbf{F}_i\in\mathbb{R}^{N_i\times L_i}}}{\operatorname{argmin}} \sum_{j=1}^{T_i} \left\|\mathbf{F}_i\mathbf{F}_i^T((\boldsymbol{\phi}_i)_j-\mathbf{m}_i)-((\boldsymbol{\phi}_i)_j-\mathbf{m}_i)\right\|_F^2$$
$$\text{subject to } \mathbf{F}_i^T\mathbf{F}_i=\mathbf{I}\,. \tag{10}$$

An optimal solution is $\mathbf{m}_i^*=T_i^{-1}\sum_{j=1}^{T_i}(\boldsymbol{\phi}_i)_j$ and $\mathbf{F}_i^*$ be the first L_i columns of $\mathbf{U}_i$ in Eq. (4) where $(\boldsymbol{\phi}_i)_j \leftarrow (\boldsymbol{\phi}_i)_j-\mathbf{m}_i^*$ for $1\le j\le T_i$. The general proof for any Hilbert space is left in the *Supplementary File*.

Centralizing over Gram Matrices To generate and apply $\mathbf{F}_i^*$, it is necessary to centralize all data by the mean of the aligning data, i.e., $(\boldsymbol{\phi}_i)_j \leftarrow (\boldsymbol{\phi}_i)_j-\mathbf{m}_i^*$. Suppose $\mathbf{Z}_i\in\mathbb{R}^{V_i\times E_i}$ is extra fMRI data for the i-th subject. Denote all-one matrices by $\mathbf{J}$. For subject i, the centralizing can be applied on the Gram matrices directly since

$$\begin{aligned}&\left(\Phi_i(\mathbf{Z}_i)^T-T_i^{-1}\mathbf{J}_{E_i\times T_i}\boldsymbol{\Phi}_i^T\right)\left(\boldsymbol{\Phi}_i-T_i^{-1}\boldsymbol{\Phi}_i\mathbf{J}_{T_i\times T_i}\right)\\ &=\Phi_i(\mathbf{Z}_i)^T\boldsymbol{\Phi}_i+T_i^{-2}\mathbf{J}_{E_i\times T_i}\boldsymbol{\Phi}_i^T\boldsymbol{\Phi}_i\mathbf{J}_{T_i\times T_i}\\ &-T_i^{-1}\mathbf{J}_{E_i\times T_i}\boldsymbol{\Phi}_i^T\boldsymbol{\Phi}_i-T_i^{-1}\Phi_i(\mathbf{Z}_i)^T\boldsymbol{\Phi}_i\mathbf{J}_{T_i\times T_i}\,.\end{aligned} \tag{11}$$

From now on, suppose all Gram matrices have been centralized. As is provided in Eq. (4), $\boldsymbol{\Phi}_i=\mathbf{V}_i\mathbf{D}_i^{\frac{1}{2}}\mathbf{U}_i^T$, which is an SVD. Denote the number of the (non-zero) singular values in $\mathbf{D}_i$ by s_i. Assume the singular values in $\mathbf{D}_i^{\frac{1}{2}}$ are in descending order and the first L_i ($L_i\le s_i$) singular values approximately contains $p_i\%$ ($p_i\in(0,100]$) energy , i.e., $\sum_{j=1}^{L_i}(D_i)_{jj}^{\frac{1}{2}}/\sum_{j=1}^{s_i}(D_i)_{jj}^{\frac{1}{2}}\approx p_i\%$. By this way, the low dimension L_i is controlled by $p_i\%$. Therefore, the corresponding low-dimensional representation of $\Phi_i(\mathbf{Z}_i)$ would be $\hat{\mathbf{U}}_i\hat{\mathbf{U}}_i^T\Phi_i(\mathbf{Z}_i)$ where $\hat{\mathbf{U}}_i$ is the first L_i columns of $\mathbf{U}_i$. Generally, with only Gram matrices, there is

$$\begin{aligned}\Phi_i(\mathbf{Z}_i)^T\hat{\mathbf{U}}_i\hat{\mathbf{U}}_i^T\hat{\mathbf{U}}_i\hat{\mathbf{U}}_i^T\boldsymbol{\Phi}_i&=\Phi_i(\mathbf{Z}_i)^T\hat{\mathbf{U}}_i\hat{\mathbf{U}}_i^T\boldsymbol{\Phi}_i\\ &\neq\Phi_i(\mathbf{Z}_i)^T\boldsymbol{\Phi}_i\,.\end{aligned}$$

Nevertheless, the equality holds with the help of $\hat{\mathbf{V}}_i$ that is defined by the first L_i columns of $\mathbf{V}_i$

Proposition 3

$$\Phi_i(\mathbf{Z}_i)^T\boldsymbol{\Phi}_i\hat{\mathbf{V}}_i=\Phi_i(\mathbf{Z}_i)^T\hat{\mathbf{U}}_i\hat{\mathbf{U}}_i^T\boldsymbol{\Phi}_i\hat{\mathbf{V}}_i\,. \tag{12}$$

Proof. Since $\boldsymbol{\Phi}_i\hat{\mathbf{V}}_i=\mathbf{U}_i\mathbf{D}_i^{\frac{1}{2}}\mathbf{V}_i^T\hat{\mathbf{V}}_i=\hat{\mathbf{U}}_i\boldsymbol{\Lambda}_i$ where $\boldsymbol{\Lambda}_i$ is the upper left $L_i\times L_i$ submatrix of $\mathbf{D}_i^{\frac{1}{2}}$, there is $\hat{\mathbf{U}}_i\hat{\mathbf{U}}_i^T\boldsymbol{\Phi}_i\hat{\mathbf{V}}_i=\hat{\mathbf{U}}_i\hat{\mathbf{U}}_i^T\hat{\mathbf{U}}_i\boldsymbol{\Lambda}_i=\hat{\mathbf{U}}_i\boldsymbol{\Lambda}_i=\boldsymbol{\Phi}_i\hat{\mathbf{V}}_i$. □

[1]An L dimensional affine subspace in $\mathbb{R}^N$ is $\mathcal{V}+\mathbf{c}$ where $\mathcal{V}$ is an L dimensional subspace and $\mathbf{c}\in\mathbb{R}^N$.

Algorithm 1 Graph-Based Decoding Model (GDM)

Input: Aligning data $\{\mathbf{X}_i\in\mathbb{R}^{V_i\times T_i}\}_{i=1}^M$, the number of the shared features K, the energy $\{p_i\%\}_{i=1}^M$ to be kept, a specific Laplacian matrix $\mathbf{L}$ and kernel functions for each subject.

For each i, standardize $\mathbf{X}_i$ such that it has zero mean along the second dimension and the variance of each feature, i.e., voxel, is 1.
Generate $\{\mathbf{K}_i\}_{i=1}^M$ via specified kernel functions.
Centralize Gram matrices: $\mathbf{K}_i \leftarrow \mathbf{K}_i + T_i^{-2}\mathbf{J}_{T_i\times T_i}\mathbf{K}_i\mathbf{J}_{T_i\times T_i}-T_i^{-1}\mathbf{J}_{T_i\times T_i}\mathbf{K}_i-T_i^{-1}\mathbf{K}_i\mathbf{J}_{T_i\times T_i}$
for $i\leftarrow 1$ **to** M **do**
    $\mathbf{K}_i=\mathbf{V}_i\mathbf{D}_i\mathbf{V}_i^T$ by spectral decomposition. The eigenvalues in $\mathbf{D}_i$ is in descending order.
    Find L_i such that the first L_i diagonal elements of $\mathbf{D}_i^{\frac{1}{2}}$ contains approximately $p_i\%$ energy.
    Let $\hat{\mathbf{V}}_i$ be the first L_i columns of $\mathbf{V}_i$.
    Let $\hat{\mathbf{D}}_i$ be the top left $L_i\times L_i$ submatrix of $\mathbf{D}_i$.
end for
By spectral decomposition, $\hat{\mathbf{V}}_*^T\mathbf{L}\hat{\mathbf{V}}_*=\mathbf{E}\boldsymbol{\Sigma}\mathbf{E}^T$ where the diagonal elements of $\boldsymbol{\Sigma}$ is ascending.
Let $\hat{\mathbf{E}}$ be the first K columns of $\mathbf{E}$ and then cut $\hat{\mathbf{E}}$ along the first dimension such that $\hat{\mathbf{E}}_i\in\mathbb{R}^{L_i\times K}$.
For $1\le i\le M$, $\mathbf{W}_i^*\leftarrow\boldsymbol{\Phi}_i\hat{\mathbf{V}}_i\hat{\mathbf{D}}_i^{-1}\hat{\mathbf{E}}_i$.

Therefore, the proposed kernel-based optimization can easily incorporate the low-dimension assumption over each new feature space. It will be shown in our experiments that this is essential for getting useful results. The overall optimization procedure of GDM is summarized in Algorithm 1.

Complexity Analysis

The shape of $\hat{\mathbf{V}}_*^T\mathbf{L}\hat{\mathbf{V}}_*$ is $L\times L$ where $L=\sum_{i=1}^M L_i$ and L_i is the low-dimension of the i-th subject. Suppose the Gram matrices are given, and $K<T_i$ for each i, i.e., the number of the shared features is smaller than the sample size, the complexity of GDM is $O((\sum_{i=1}^M T_i^3)+L(L^2+T^2+LT))$ where $T=\sum_{i=1}^M T_i$. If each low-dimension L_i is fixed, the complexity thus becomes $O((\sum_{i=1}^M T_i^3)+T^2)$. Notably, it could be reduced into $O(\max_{1\le i\le M}T_i^3+T^2)$ by using parallel programming. By contrast, the naive kernel scheme cannot be parallelized, and the complexity of employing it is $O(T^3)$. Therefore, our proposed kernel method is more efficient compared to the naive kernel scheme. Besides, different from methods based on iterative optimization algorithms, one can obtain the optimal solution of GDM directly.

Experiments

Materials We utilize five datasets shared by openfmri.org and Chen *et al.* (Chen et al. 2015). The relevant information about each dataset is outlined in Table 1. Raw datasets are preprocessed by using FSL (fsl.fmrib.ox.ac.uk), following a standard process (i.e., slice timing, anatomical alignment, normalization, and smoothing). The default parameters in

Table 1: The brief information and parameter settings for each dataset. Here, K is the number of the shared features, energy $p\%$ is set for all subjects, ν is related to ν-SVM.

Dataset	#subject	#sample/subject	#feature	#category	K	energy($p\%$)	ν	#subject left out
DS105WB	6	994	19174	8	10	82	0.8	1
DS105ROI	6	994	2294	8	10	82	0.8	1
DS011	14	271	19174	2	10	82	0.3	2
DS001	16	485	19174	4	10	82	0.5	4
DS232	10	1691	9947	4	10	82	0.8	2
Raider.Movie	10	2203	1000	—	20	35	—	—
Raider.Image	10	56	1000	7	—	—	0.5	2

Table 2: The performance on temporally-aligned datasets is measured by BSC accuracy. The larger the better. Each performance is reported by averaging accuracies over all folds with standard deviation. The bold denotes the best result on each dataset.

Dataset(#class)	ν-SVM	HA	KHA	SVDHA	SRM	RSRM	RHA	GDM (Ours)
DS105WB(8)	11.67 ± 1.80	39.70 ± 3.90	39.22 ± 4.50	30.48 ± 3.52	39.69 ± 3.95	40.01 ± 3.84	52.50 ± 4.28	$\mathbf{60.68 \pm 5.23}$
DS105ROI(8)	13.06 ± 2.93	48.05 ± 3.93	48.22 ± 3.34	41.33 ± 4.19	48.14 ± 3.17	48.51 ± 3.80	57.63 ± 5.55	$\mathbf{62.22 \pm 4.23}$
DS011(2)	51.80 ± 3.73	85.39 ± 3.52	85.79 ± 3.82	74.42 ± 4.40	85.47 ± 3.53	85.58 ± 3.89	91.80 ± 2.65	$\mathbf{92.49 \pm 2.24}$
DS232(4)	25.89 ± 2.46	69.34 ± 3.22	69.38 ± 3.16	56.77 ± 4.52	69.18 ± 3.27	69.25 ± 3.20	77.64 ± 2.75	$\mathbf{82.47 \pm 1.45}$
DS001(4)	34.32 ± 2.08	56.74 ± 1.63	57.10 ± 1.97	51.99 ± 1.87	56.83 ± 1.54	57.20 ± 1.30	57.87 ± 0.61	$\mathbf{62.68 \pm 1.53}$
Raider(7)	26.61 ± 3.80	60.48 ± 3.68	60.71 ± 3.23	58.99 ± 4.19	60.65 ± 4.16	62.38 ± 3.48	59.82 ± 4.10	$\mathbf{64.52 \pm 3.28}$

FSL were taken when the dataset does not provide. The description of each dataset is as follows:

(1) **DS105**: The fMRI data were measured while six subjects viewed gray-scale images of faces, houses, cats, bottles, scissors, shoes, chairs, and nonsense images (Haxby et al. 2001). Hence, there are totally 8 categories in this dataset. Here, DS105WB contains the whole-brain fMRI data while the data in DS105ROI are based on a region of interest.

(2) **DS011**: Fourteen subjects participated in a single task (weather prediction). In the first phase, they learned to predict weather outcomes (rain or sun) for two different cities. After learning, they predicted weather (Foerde, Knowlton, and Poldrack 2006). Thus, there are two cognitive states.

(3) **DS001:** Sixteen subjects were instructed to inflate a control balloon or a reward balloon on a screen. For a control balloon, subjects had merely one choice whereas they could choose to pump or cash out for another case. After opting to pump, the balloon may explode or expand (Schonberg et al. 2012). Hence, there are four different cognitive states.

(4) **DS232**: Ten subjects were instructed to respond to images of faces, scenes, objects and phrase-scrambled versions of the scene images (Carlin and Kriegeskorte 2017).

(5) **Raider**: As a commonly-used one, it collected data from 10 subjects participating in two experiments. Firstly, 10 subjects watched a movie Raiders of the Lost Ark (2203 TRs). The data of movie dataset does not contain any label. In the next experiment, the same 10 subjects were shown 7 classes of images (female face, male face, monkey face, dog face, house, chair and shoes) (Chen et al. 2015).

It's worth noting that, except for Raider, all four other datasets are not temporally-aligned. To compare GDM with other temporal-alignment-based methods, following the previous study in (Chen et al. 2015), these datasets (i.e., DS105, DS011, DS001, and DS232) are reordered and truncated, or downsampled, to be aligned according to their categories.

Experimental Setup We follow the experiment setup with a cross-validation strategy in previous studies (Chen et al. 2014; Haxby et al. 2011), as illustrated by Fig. S1 and Fig. S2 in the *Supplementary File*. Specifically, except for the Raider, each subject's data is equally divided into two parts with each category being equally split, one is for alignment whereas the other is for training or testing a classifier. Switching the roles of the two parts and leave-k-subject-out strategy are adopted for cross-validation. For instance, if there are 16 subjects, leave-4-subject-out leads to $16 \div 4 \times 2 = 8$ folds for cross-validation. For Raider dataset, the movie data is taken for alignment while the image data is for classification. Here, the first $2,202$ time points of movie data are used for alignment. Then it is equally divided into threes parts with each part having 734 samples for cross-validation. Since a leave-2-subject-out is used in this dataset, there are a total of $10 \div 2 \times 3 = 15$ folds when using Raider. As shown in Fig. S1 in the *Supplementary File*, the experiment on each dataset contains two stages: 1) aligning phase, and 2) classification phase. At the first phase, one part of all subjects' data are fed into a functional alignment method to yield the corresponding aligning maps $\{f_i : \mathbb{R}_i^V \mapsto \mathbb{R}^K\}_{i=1}^M$. At the classification phase, the remaining part of data are first mapped to the shared feature space via the learned aligning maps and then used for classification model construction. Note that those data used at the aligning phase will not be used at the classification phase in our experiments.

Since each dataset (or part of it) used in this paper includes labels, the performance of alignment is assessed by testing how well a trained classifier can generalize to new

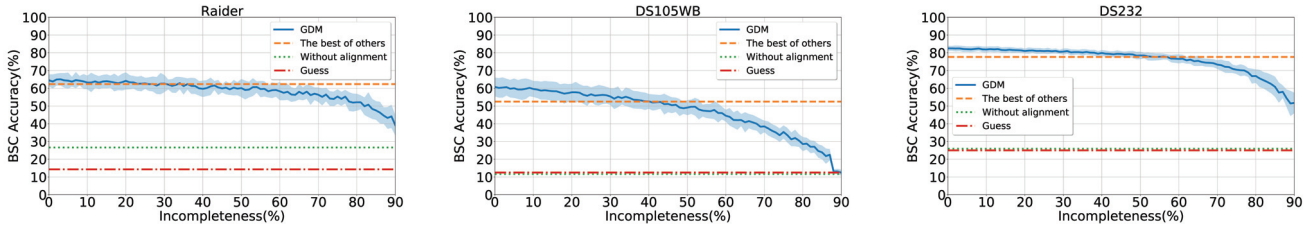

Figure 1: Performance of GDM on incomplete or unaligned datasets. Here, $q\%$ incompleteness means that $q\%$ of the aligning data are randomly removed per subject. The term "Without alignment" denotes the ν-SVM method without any alignment. "The best of others" refers to the best result of six competing methods with complete data, and "Guess" denotes the randomly guess method.

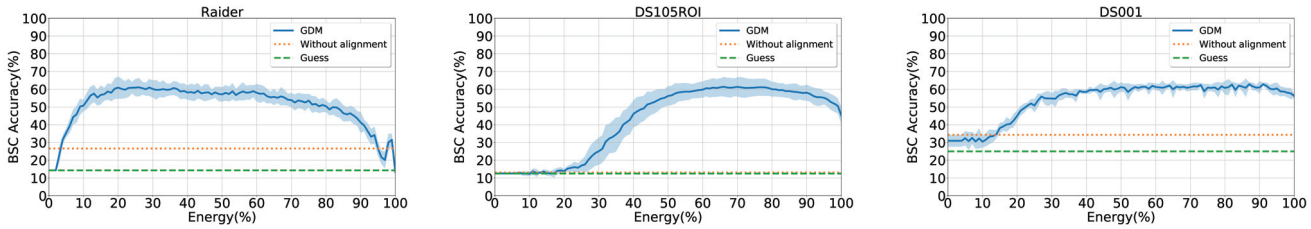

Figure 2: The necessity of low-dimension assumption for GDM. Here, $p\%$ energy shows how much energy is nearly kept per subject. The term "Without alignment" denotes the ν-SVM method without any alignment.

subjects, i.e., between-subject classification (BSC) accuracy (Haxby et al. 2011). Like previous studies, ν-SVM is used for classification (Chang and Lin 2011).

Competing Methods The proposed GDM method is compared with six state-of-the-art methods in the experiments, including (1) Hyperalignment (HA) (Haxby et al. 2011), (2) Regularized Hyperalignment (RHA) (Xu et al. 2012), (3) Kernel Hyperalignment (KHA) (Lorbert and Ramadge 2012), (4) SVD-Hyperalignment (SVDHA) (Chen et al. 2014), (5) Shared Response Model (SRM) (Chen et al. 2015), and (6) Robust SRM (RSRM) (Turek et al. 2018). All methods are implemented by ourselves in Python.

The parameter settings for each dataset are briefly listed in Table 1. For a fair comparison, the parameter ν in ν-SVM (with a linear kernel) is fixed for all methods on each dataset. For six competing methods, we choose the optimal hyperparameters according to their original papers. For our GDM model, a linear kernel is fixed, while the influence of different kernels are shown in Figs. S3-S8 in the *Supplementary File*. For the Raider dataset, we set $G_{ij} = 1$ if the i-th and j-th samples are temporally aligned; and $G_{ij} = 0$, otherwise. For the other datasets, we set $G_{ij} = 1$ if the i-th and j-th samples are in the same category; and $G_{ij} = -1$, otherwise.

Results on Aligned Datasets On the temporally-aligned datasets, we report the BSC accuracy values achieved by eight different methods in Table 2. As can be seen from Table 2, for each aligned dataset, the proposed GDM method consistently outperforms the competing methods in terms of BSC accuracy. For example, GDM achieves the improvement of $> 8\%$ compared to the second best result (i.e., 52.50 of RHA) on the DIS105WB dataset.

Results on Unaligned Datasets To assess the performance of GDM when dealing with unaligned data, we randomly remove some data from each aligned dataset. Here, the term $q\%$ incompleteness means that q percent of aligning data are randomly removed per subject. The corresponding results are shown in Fig. 1. Notably, six competing methods (i.e., HA, RHA, KHA, SVDHA, SRM, and RSRM) cannot be applied to such incomplete datasets since they are designed for aligned data. More results can be found in Figs. S3-S5 in the *Supplementary File*. From Fig. 1, one can observe that our GDM is able to preserve a dominant BSC accuracy with incompleteness up to at least 20%. For the DS232 dataset, the performance of GDM still beats others with 50% incompleteness. These results further validate the superiority of GDM in handling unaligned data.

Necessity of Low-Dimension Assumption To evaluate the influence of low-dimension assumption on GDM, we perform another group of experiments to study the BSC values achieved by GDM with different energy ratios kept on three datasets, with results reported in Figure 2. This figure suggests that the best results are not achieved by GDM with 100% energy on each dataset, thus verifying the importance of the low-dimension assumption. Besides, on the Raider dataset, GDM still achieves a good result with around 20% energy kept. We conjecture that it results from the fact that the movie data contain much richer information than the visual data generated from simple objects. More results can

be found in Figs. S6-S8 in the *Supplementary File*.

Conclusion

As an essential step in fMRI analysis, functional alignment removes the differences between subjects' brains so that multi-subject fMRI data can be aggregated to make valid and general inferences. However, the existing methods cannot well handle unaligned fMRI datasets. In this paper, a flexible framework is developed on a cross-subject graph that depicts the (dis)similarities among all samples. To reduce the computational cost, the framework is regularized so that a novel feasible kernel-based optimization is analytically developed. To avoid overfitting caused by the HSLT resolution of fMRI, a low-dimension assumption is made over each new feature space, and we also propose a way to incorporate such an assumption into our proposed optimization. Experimental results attest to the superiority of GDM. In the future, we plan to study how to construct an informative graph matrix in different situations.

Acknowledgments

This work was in part supported by the National Natural Science Foundation of China (Nos. 61876082, 61732006, 61861130366, 61703301), the National Key R&D Program of China (Nos. 2018YFC2001600, 2018YFC2001602), the Taishan Scholar Program of Shandong Province in China, and the Shandong Natural Science Foundation for Distinguished Young Scholar in China (No. ZR2019YQ27).

References

Carlin, J. D., and Kriegeskorte, N. 2017. Adjudicating between face-coding models with individual-face fMRI responses. *PLoS Computational Biology* 13(7):e1005604.

Chang, C.-C., and Lin, C.-J. 2011. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2(3):27.

Chen, P.-H.; Guntupalli, J. S.; Haxby, J. V.; and Ramadge, P. J. 2014. Joint svd-hyperalignment for multi-subject fMRI data alignment. In *2014 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 1–6. IEEE.

Chen, P.-H. C.; Chen, J.; Yeshurun, Y.; Hasson, U.; Haxby, J.; and Ramadge, P. J. 2015. A reduced-dimension fMRI shared response model. In *Advances in Neural Information Processing Systems (NIPS)*, 460–468.

Chung, F. R., and Graham, F. C. 1997. *Spectral graph theory*. Number 92. American Mathematical Soc.

Conroy, B.; Singer, B.; Haxby, J.; and Ramadge, P. J. 2009. fMRI-based inter-subject cortical alignment using functional connectivity. In *Advances in Neural Information Processing Systems (NIPS)*, 378–386.

De Sa, V. R.; Gallagher, P. W.; Lewis, J. M.; and Malave, V. L. 2010. Multi-view kernel construction. *Machine Learning* 79(1-2):47–71.

Fischl, B.; Sereno, M. I.; Tootell, R. B.; and Dale, A. M. 1999. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping* 8(4):272–284.

Foerde, K.; Knowlton, B. J.; and Poldrack, R. A. 2006. Modulation of competing memory systems by distraction. *Proceedings of the National Academy of Sciences* 103(31):11778–11783.

Guntupalli, J. S.; Hanke, M.; Halchenko, Y. O.; Connolly, A. C.; Ramadge, P. J.; and Haxby, J. V. 2016. A model of representational spaces in human cortex. *Cerebral Cortex* 26(6):2919–2934.

Haxby, J. V.; Gobbini, M. I.; Furey, M. L.; Ishai, A.; Schouten, J. L.; and Pietrini, P. 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293(5539):2425–2430.

Haxby, J. V.; Guntupalli, J. S.; Connolly, A. C.; Halchenko, Y. O.; Conroy, B. R.; Gobbini, M. I.; Hanke, M.; and Ramadge, P. J. 2011. A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* 72(2):404–416.

Haxby, J. V.; Connolly, A. C.; and Guntupalli, J. S. 2014. Decoding neural representational spaces using multivariate pattern analysis. *Annual Review of Neuroscience* 37:435–456.

Logothetis, N. K. 2002. The neural basis of the blood–oxygen–level–dependent functional magnetic resonance imaging signal. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 357(1424):1003–1037.

Lorbert, A., and Ramadge, P. J. 2012. Kernel hyperalignment. In *Advances in Neural Information Processing Systems (NIPS)*, 1790–1798.

Rademacher, J.; Caviness Jr, V.; Steinmetz, H.; and Galaburda, A. 1993. Topographical variation of the human primary cortices: implications for neuroimaging, brain mapping, and neurobiology. *Cerebral Cortex* 3(4):313–329.

Sabuncu, M. R.; Singer, B. D.; Conroy, B.; Bryan, R. E.; Ramadge, P. J.; and Haxby, J. V. 2009. Function-based intersubject alignment of human cortical anatomy. *Cerebral Cortex* 20(1):130–140.

Schonberg, T.; Fox, C. R.; Mumford, J. A.; Congdon, E.; Trepel, C.; and Poldrack, R. A. 2012. Decreasing ventromedial prefrontal cortex activity during sequential risk-taking: An fMRI investigation of the balloon analog risk task. *Frontiers in Neuroscience* 6:80.

Steinwart, I., and Christmann, A. 2008. *Support vector machines*. Springer Science & Business Media.

Talairach, J., and Tournoux, P. 1988. Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system: an approach to cerebral imaging.

Turek, J. S.; Willke, T. L.; Chen, P.-H.; and Ramadge, P. J. 2017. A semi-supervised method for multi-subject fMRI functional alignment. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1098–1102. IEEE.

Turek, J. S.; Ellis, C. T.; Skalaban, L. J.; Turk-Browne, N. B.; and Willke, T. L. 2018. Capturing shared and individual information in fMRI data. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 826–830. IEEE.

Watson, J. D.; Myers, R.; Frackowiak, R. S.; Hajnal, J. V.; Woods, R. P.; Mazziotta, J. C.; Shipp, S.; and Zeki, S. 1993. Area v5 of the human brain: evidence from a combined study using positron emission tomography and magnetic resonance imaging. *Cerebral Cortex* 3(2):79–94.

Xu, H.; Lorbert, A.; Ramadge, P. J.; Guntupalli, J. S.; and Haxby, J. V. 2012. Regularized hyperalignment of multi-set fMRI data. In *2012 IEEE Statistical Signal Processing Workshop (SSP)*, 229–232. IEEE.