NeuroImage 96 (2014) 183–202

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Disease prediction based on functional connectomes using a scalable and spatially-informed support vector machine

Takanori Watanabe [a,*], Daniel Kessler [c], Clayton Scott [a,b], Michael Angstadt [c], Chandra Sripada [c]

[a] Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA
[b] Department of Statistics, University of Michigan, Ann Arbor, MI, USA
[c] Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA

ARTICLE INFO

Article history:
Accepted 24 March 2014
Available online 1 April 2014

Keywords:
Classification
Feature selection
Structured sparsity
Resting state fMRI
Functional connectivity
Support vector machine

ABSTRACT

Substantial evidence indicates that major psychiatric disorders are associated with distributed neural dysconnectivity, leading to a strong interest in using neuroimaging methods to accurately predict disorder status. In this work, we are specifically interested in a multivariate approach that uses features derived from whole-brain resting state functional connectomes. However, functional connectomes reside in a high dimensional space, which complicates model interpretation and introduces numerous statistical and computational challenges. Traditional feature selection techniques are used to reduce data dimensionality, but are blind to the spatial structure of the connectomes. We propose a regularization framework where the 6-D structure of the functional connectome (defined by pairs of points in 3-D space) is explicitly taken into account via the fused Lasso or the GraphNet regularizer. Our method only restricts the loss function to be convex and margin-based, allowing non-differentiable loss functions such as the hinge-loss to be used. Using the fused Lasso or GraphNet regularizer with the hinge-loss leads to a structured sparse support vector machine (SVM) with embedded feature selection. We introduce a novel efficient optimization algorithm based on the augmented Lagrangian and the classical alternating direction method, which can solve both fused Lasso and GraphNet regularized SVM with very little modification. We also demonstrate that the inner subproblems of the algorithm can be solved efficiently in analytic form by coupling the variable splitting strategy with a data augmentation scheme. Experiments on simulated data and resting state scans from a large schizophrenia dataset show that our proposed approach can identify predictive regions that are spatially contiguous in the 6-D "connectome space," offering an additional layer of interpretability that could provide new insights about various disease processes.

Introduction

There is substantial interest in establishing neuroimaging-based biomarkers that reliably distinguish individuals with psychiatric disorders from healthy individuals. Towards this end, neuroimaging affords a variety of specific modalities including structural imaging, diffusion tensor imaging (DTI) and tractography, and activation studies under conditions of cognitive challenge (i.e., task-based functional magnetic resonance imaging (fMRI)). In addition, resting state fMRI has emerged as a mainstream approach that offers robust, sharable, and scalable ability to comprehensively characterize patterns of connections and network architecture of the brain.

Recently a number of groups have demonstrated that substantial quantities of discriminative information regarding psychiatric diseases reside in resting state functional connectomes (Castellanos et al., 2013; Fox and Greicius, 2010). In this article, we define the functional connectomes as the cross-correlation matrix that results from parcellating the brain into hundreds of distinct regions, and computing cross-correlation matrices across time (Varoquaux and Craddock, 2013). Even with relatively coarse parcellation schemes with several hundred regions of interest (ROIs), the resulting connectomes encompass hundreds of thousands of connections or more. The massive size of connectomes offers new possibilities, as patterns of connectivity across the entirety of the brain are represented. Nonetheless, the high dimensionality of connectomic data presents critical statistical and computational challenges. In particular, mass univariate strategies that perform separate statistical tests at each edge of the connectome require excessively stringent corrections for multiple comparisons. Multivariate methods are promising, but these require specialized approaches in the context where the number of parameters dominate the number of observations, a setting commonly referred to as the "large p small n problem," denoted $p \gg n$ (Bühlmann and van de Geer, 2011; West, 2003).

In the $p \gg n$ regime, it is important to leverage any potential structure in the data, and sparsity is a natural assumption that arises in many

* Corresponding author at: 1301 Beal Avenue, 4111 EECS, Ann Arbor, MI 48109, USA.
E-mail addresses: takanori@umich.edu (T. Watanabe), kesslerd@umich.edu (D. Kessler), clayscot@umich.edu (C. Scott), mangstad@med.umich.edu (M. Angstadt), sripada@umich.edu (C. Sripada).

http://dx.doi.org/10.1016/j.neuroimage.2014.03.067
1053-8119/

applications (Candes and Wakin, 2008; Fan and Lv, 2010). For example, in the context of connectomics, it is reasonable to believe that only a fraction of the functional connectome is impacted under a specific disorder, an assumption that has been supported in nearly all extant studies (see Castellanos et al., 2013). Furthermore, when sparsity is coupled with a linear classifier, the nonzero variables can be interpreted as pairs of brain regions that allow reliable discrimination between controls and patients. In other words, sparse linear classifiers have the potential of revealing *connectivity-based biomarkers* that characterize mechanisms of the disease process of interest (Atluri et al., 2013).

The problem of identifying the subset of variables relevant for prediction is called feature selection (Guyon and Elisseeff, 2003; Jain et al., 2000), which can be done in a univariate or a multivariate fashion. In the univariate approach, features are independently ranked based on their statistical relationship with the target label (e.g., two sample *t*-test, mutual information), and only the top features are submitted to the classifier. While this method is commonly used (Sripada et al., 2013b; Zeng et al., 2012), it ignores the multivariate nature of fMRI. On the other hand, multivariate approaches such as *recursive feature elimination* (Guyon and Elisseeff, 2003) can be used to capture feature interactions (Craddock et al., 2009; Dai et al., 2012), but these methods are computationally intensive and rely on suboptimal heuristics. However, a more serious shortcoming common to all the methods above is that outside of sparsity, no structural information is taken into account. In particular, we further know that functional connectomes reside in a structured space, defined by pairs of coordinate points in 3-D brain space. Performing prediction and feature selection in a spatially informed manner could potentially allow us to draw more neuroscientifically meaningful conclusions. Fortunately, *regularization methods* allow us to achieve this in a natural and principled way.

Regularization is a classical technique to prevent overfitting (James and Stein, 1961; Tikhonov, 1963), achieved by encoding prior knowledge about the data structure into the estimation problem. Sparsity promoting regularization methods, such as Lasso (Tibshirani, 1996) and Elastic-net (Zou and Hastie, 2005), have the advantage of performing prediction and feature selection jointly (Grosenick et al., 2008; Yamashita et al., 2008); however, they also have the issue of neglecting additional structure the data may have. Recently, there has been strong interest in the machine learning community in designing a convex regularizer that promotes *structured sparsity* (Chen et al., 2012; Mairal et al., 2011; Micchelli et al., 2013), which extends the standard concept of sparsity. Indeed, spatially informed regularizers have been applied successfully in task-based detection, i.e., *decoding*, where the goal is to localize in 3-D space the brain regions that become active under an external stimulus (Baldassarre et al., 2012; Gramfort et al., 2013; Grosenick et al., 2013; Jenatton et al., 2012; Michel et al., 2011). Connectomic maps exhibit a rich spatial structure, as each connection comes from a pair of localized regions in 3-D space, giving each connection a localization in 6-D space (referred to as "connectome space" hereafter). However, to the best of our knowledge, no framework currently deployed exploits this spatial structure in the functional connectome.

Based on these considerations, the main contributions of this paper are two-fold. First, we propose to explicitly account for the 6-D spatial structure of the functional connectome by using either the fused Lasso (Tibshirani et al., 2005) or the GraphNet regularizer (Grosenick et al., 2013). Second, we introduce a novel scalable algorithm based on the classical alternating direction method (Boyd et al., 2011; Gabay and Mercier, 1976; Glowinski and Marroco, 1975) for solving the non-smooth, large-scale optimization problem that results from these spatially-informed regularizers. Variable splitting and data augmentation strategies are used to break the problem into simpler subproblems that can be solved efficiently in closed form. The method we propose only restricts the loss function to be convex and margin-based, which allows non-differentiable loss functions such as the hinge-loss to be used. This is important, since using the fused Lasso or the GraphNet regularizer with the hinge-loss function leads to a structured sparse support vector machine (SVM) (Grosenick et al., 2013; Ye and Xie, 2011), where feature selection is *embedded* (Guyon and Elisseeff, 2003), i.e., feature selection is conducted jointly with classification. We demonstrate that the optimization algorithm we introduce can solve both fused Lasso and GraphNet regularized SVM with very little modification. To the best of our knowledge, this is the first application of structured sparse methods in the context of disease prediction using functional connectomes. Additional discussions of technical contributions are reported in the Optimization section. We perform experiments on simulated connectomic data and resting state scans from a large schizophrenia dataset to demonstrate that the proposed method identifies predictive regions that are spatially contiguous in the connectome space, offering an additional layer of interpretability that could provide new insights about various disease processes.

Notation

We let lowercase and uppercase bold letters denote vectors and matrices, respectively. For every positive integer $n \in \mathbb{N}$, we define an index set $[n] := \{1,\ldots, n\}$, and also let $\boldsymbol{I}_n \in \mathbb{R}^{n \times n}$ denote the identity matrix. Given a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times p}$, we let $\boldsymbol{A}^T$ denote its matrix transpose, and $\boldsymbol{A}^H$ denote its Hermitian transpose. Given $\boldsymbol{w}, \boldsymbol{v} \in \mathbb{R}^n$, we invoke the standard notation $\langle \boldsymbol{w},\boldsymbol{v} \rangle := \sum_{i=1}^{n} w_i v_i$ to express the inner product in $\mathbb{R}^n$. We also let $\|\boldsymbol{w}\|_p = (\sum_{i=1}^{n} w_i^p)^{1/p}$ denote the ℓ_p-norm of a vector, $p \geq 1$, with the absence of subscript indicating the standard Euclidean norm, $\|\cdot\| = \|\cdot\|_2$.

Material and methods

Defining functional connectomes

In this work, we produced a whole-brain resting state functional connectome as follows. First, 347 non-overlapping spherical nodes are placed throughout the entire brain in a regularly-spaced grid pattern, with a spacing of $18 \times 18 \times 18$ mm; each of these nodes represents a pseudo-spherical ROI with a radius of 7.5 mm, which encompasses 33 voxels (the voxel size is $3 \times 3 \times 3$ mm). For a schematic representation of the parcellation scheme, see Fig. 1. Next, for each of these nodes, a single representative time-series is assigned by spatially averaging the BOLD signals falling within the ROI. Then, a cross-correlation matrix is generated by computing Pearson's correlation coefficient between these representative time-series. Finally, a vector $\boldsymbol{x}$ of length $\binom{347}{2} = 60{,}031$ is obtained by extracting the lower-triangular portion of the cross-correlation matrix. This vector $\boldsymbol{x} \in \mathbb{R}^{60{,}031}$ represents the whole-brain functional connectome, which serves as the feature vector for disease prediction.

The grid-based scheme for brain parcellation used in this work provides numerous advantages. Of note, this approach has been validated in previous studies (Sripada et al., 2013a, 2013b, 2014). Furthermore, the uniformly spaced grid is a good fit with our implementation of fused Lasso and GraphNet, as it provides a natural notion of nearest-neighbor and ordering among the coordinates of the connectome. This property also turns out to be critical for employing our optimization algorithm, which will be discussed in the Optimization section. This is in contrast to alternative approaches, such as methods that rely on anatomical (Tzourio-Mazoyer et al., 2002; Zeng et al., 2012) or functional parcellation schemes (Dosenbach et al., 2010). Anatomical parcellations in particular have been shown to yield inferior performance to alternative schemes in the literature (Power et al., 2011). Additionally, grid-based approaches provide scalable density: there is a natural way to increase the spatial resolution of the grid when computational feasibility allows. In particular, to increase node density, one could reduce the inter-node distance and also reduce the node size such that suitable inter-node space remains. This scalable density property turns out to be quite important, as our grid-based scheme is considerably more dense than standard functional parcellations (e.g., Dosenbach et al.,