Journal of Neuroscience Methods 276 (2017) 46–55

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

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

A robust sparse-modeling framework for estimating schizophrenia biomarkers from fMRI

Keith Dillon[a,b,*], Vince Calhoun[c,d], Yu-Ping Wang[a,b]

[a] Department of Biomedical Engineering, Tulane University, New Orleans, LA, USA
[b] Department of Global Biostatistics and Data Science, Tulane University, New Orleans, LA, USA
[c] The Mind Research Network & LBERI, Albuquerque, NM, USA
[d] Department of Electrical Engineering, University of New Mexico, New Mexico, USA

HIGHLIGHTS

- We consider robust components as constant over all possible unknown mechanisms.
- We derive a method to incorporate a preference for sparsity in the mechanism.
- Improvement in robustness is demonstrated with simulation.
- Application to fMRI demonstrates superior accuracy in classifying schizophrenia.

ARTICLE INFO

Article history:
Received 19 September 2016
Received in revised form 9 November 2016
Accepted 10 November 2016
Available online 17 November 2016

Keywords:
Schizophrenia
Functional MRI
Sparsity
PCA
Optimization

ABSTRACT

Background: Our goal is to identify the brain regions most relevant to mental illness using neuroimaging. State of the art machine learning methods commonly suffer from repeatability difficulties in this application, particularly when using large and heterogeneous populations for samples.
New method: We revisit both dimensionality reduction and sparse modeling, and recast them in a common optimization-based framework. This allows us to combine the benefits of both types of methods in an approach which we call unambiguous components. We use this to estimate the image component with a constrained variability, which is best correlated with the unknown disease mechanism.
Results: We apply the method to the estimation of neuroimaging biomarkers for schizophrenia, using task fMRI data from a large multi-site study. The proposed approach yields an improvement in both robustness of the estimate and classification accuracy.
Comparison with existing methods: We find that unambiguous components incorporate roughly two thirds of the same brain regions as sparsity-based methods LASSO and elastic net, while roughly one third of the selected regions differ. Further, unambiguous components achieve superior classification accuracy in differentiating cases from controls.
Conclusions: Unambiguous components provide a robust way to estimate important regions of imaging data.

1. Introduction

In this paper our goal is to find the most relevant brain regions given labeled neuroimaging data; the ultimate goal is to use those results to understand disease mechanisms, as well as to provide biomarkers to help diagnose (i.e., classify) patients as having disease or not. There is a significant need for techniques which can robustly extract information in such a problem. Neuroimaging, particularly functional neuroimaging, has provided a wealth of intriguing information regarding brain function, but has yet to show clear value to psychiatric diagnosis (Krystal and State, 2014). Despite this, impressive results have been achieved with machine learning techniques such as support vector machines, which demonstrate high classification accuracies (Orr et al., 2012). Reproducibility problems persist however (Buck, 2015), with an apparent trend towards poorer performance for larger studies (Schnack and Kahn, 2016).

The identification of meaningful components of the data is a key benefit of many feature selection techniques (Guyon and Elisseeff,

* Corresponding author at: Department of Biomedical Engineering, Tulane University, New Orleans, LA, USA.
E-mail address: kdillon1@tulane.edu (K. Dillon).

http://dx.doi.org/10.1016/j.jneumeth.2016.11.005
0165-0270/

2003), in addition to providing improvements in performance of subsequent classification stages (Chu et al., 2012). In a typical neuroimaging study there may be tens or hundreds of subjects, each with an image consisting of up to hundreds of thousands of voxels, resulting in an extremely underdetermined problem. A popular category of feature selection approaches is regularized regression techniques such as LASSO (Tibshirani, 1996) and related methods employing sparse models (Cao et al., 2014; Lin et al., 2014). Such supervised techniques impose task-specific information (the data labels), with a penalty term to incorporate prior knowledge. In the case of LASSO, the prior knowledge amounts to a presumption of sparsity on the relationship between image data and labels, i.e., that the underlying biological mechanism involves a limited number of the imaged voxels. Unfortunately, if the problem is both very underdetermined and very noisy, then the regularized solution may not be a particularly superior choice; many solutions may potentially be of similar or even equal probability to each other. For example in the underdetermined case, the LASSO solution may not be unique for highly-structured datasets (Tibshirani, 2013; Zhang et al., 2014). Along these lines, we simply may not have sufficient confidence in the validity of our prior knowledge formulation to presume that the most regular solution is preferable to those even moderately as regular.

From a different direction, dimensionality reduction techniques (Lemm et al., 2011) offer a more robust approach to feature selection in neuroimaging data. An example is principal component analysis (PCA) (Dunteman, 1989), which finds basis vectors for the space containing the data variation. This set of basis vectors can be viewed as robust in the sense that they are common to all solutions of any linear regression based on the data. In statistics a closely-related concept is estimable functions (Milliken and Johnson, 2009). We will refer to components with such a property as unambiguous, and examine this more formally in the next section. Of course such components only describe the data itself, not necessarily the aspects of the data pertinent to our application, such as for finding information most related to a disease phenotype. A common approach is to utilize PCA and related factoring methods in a supervised fashion by choosing a subset of factors which best correlate with the labels. Supervised factoring techniques such as the "Supervised PCA" of Barshan et al. (2011), and related methods, can be viewed as a more sophisticated version of this technique, finding a transformation of the data such that the correlation with the labels is maximized. However these techniques are not able to incorporate prior knowledge, such as sparsity of the mechanism, into this transformation. Techniques have been developed which do incorporate sparsity into unsupervised factoring techniques (e.g., sparse PCA of Zou et al., 2006) in a heuristic sense, though this differs from presuming sparsity of the underlying mechanism; the presumption of sparsity is applied to the structure of the component itself rather than to the unknown mechanism. Hybrid methods have been proposed which perform PCA following a pre-screening step which picks a subset of variables over a correlation threshold (Bair et al., 2006) or in known pathways (Ma and Dai, 2011). However we would prefer to incorporate multi-variable relationships in the screening component.

In this paper, we develop an approach which combines the benefits of both regularized estimates and dimensionality reduction by simultaneously enforcing unambiguity and prior knowledge in calculating components. We start by reviewing dimensionality reduction from the perspective of unambiguous components. Then we review related regularization methods and show how they motivate our approach to incorporate prior knowledge into unambiguous components. By maximizing the correlation with the mechanism, we calculate components which identify the most important regions in the data. We use a simulation to show how this component performs robustly in the face of inaccurate prior knowledge, by demonstrating that the correlation still remains controlled as the prior knowledge is relaxed. Finally, we show a successful application to biomarker identification where we identify features of fMRI data which relate to schizophrenia more accurately than other methods which utilize sparsity as prior knowledge.

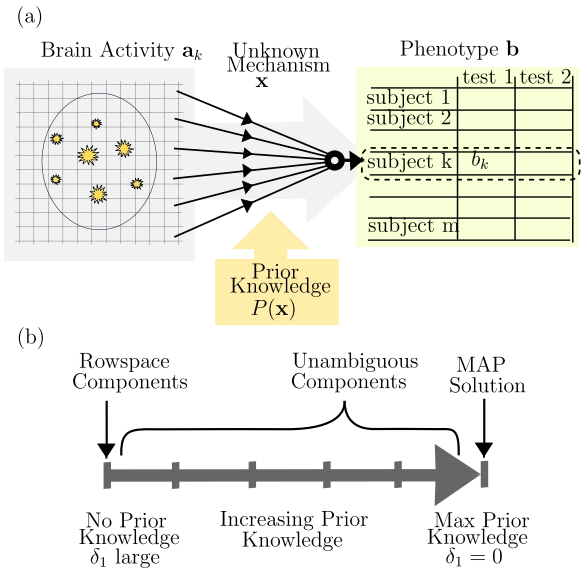

Fig. 1. (a) Mathematical model $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{x}$ describes the mechanism that relates brain activity to phenotype (psychiatric assessments). The contrast map for a single subject, $\mathbf{a}_k$, provides the kth row of $\mathbf{A}$. As there are still many unknown biological variables, the problem is underdetermined and $\mathbf{x}$ cannot be found uniquely; instead we must settle for a probable result such as a maximum likelihood solution which utilizes prior knowledge, or an estimable component of $\mathbf{x}$. (b) Continuum between rowspace components and most probable solution, based on increasing confidence in the prior knowledge, which we control by the relaxation parameter δ_1.

2. Materials and methods

We will consider the linear model $\mathbf{Ax} = \mathbf{b} + \mathbf{n}$ where $\mathbf{A}$ is a $m \times n$ data matrix with $n > m$, containing samples as rows, and variables as columns; $\mathbf{b}$ is the phenotype encoded into a vector of labels such as case or control; the solution $\mathbf{x}$ is the unknown model parameters that relate $\mathbf{A}$ to $\mathbf{b}$; and $\mathbf{n}$ is a noise vector about which we have only statistical information. We will also assume the means have been removed from $\mathbf{b}$ and the columns of $\mathbf{A}$ to simplify the presentation. The rows of $\mathbf{A}$ are provided by the contrast images from individual study subjects, so a predictor $\mathbf{x}$ selects a weighted combination of voxels (i.e., columns of $\mathbf{A}$) which relates the imagery to the case–control status. By examining the weightings in this combination we hope to learn more about the spatial distribution of causes or effects of the disease, which we will term the "mechanism" in this paper. The model is depicted in Fig. 1(a), where we depict the true solution $\mathbf{x}$ as the mechanism whereby brain activity relates to the measured phenotypes. Of course there are far more unknown variables than samples, hence our linear system is underdetermined and there will be many possible $\mathbf{x}$ which solve the system. One way to address this problem is to impose prior knowledge about the biological mechanism, such as a preference for sparser $\mathbf{x}$, and select the solution which best fulfills this preference. We will review this approach in a later section. Another approach is to restrict our analysis to components of the solution which may be more easily estimated, such as via dimensionality reduction; an intuitive example of this approach is to group voxels into low-resolution regions. These alternatives are depicted in Fig. 1(b), as extremes on a continuum of possible methods, where the goal of this paper is to find intermediate information which utilizes the benefits of both extremes.