NeuroImage xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

NeuroImage

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

Subject-specific functional parcellation via Prior Based Eigenanatomy

Paramveer S. Dhillon [a,b,*,1], David A. Wolk [c], Sandhitsu R. Das [a], Lyle H. Ungar [b], James C. Gee [a], Brian B. Avants [a]

[a] Penn Image Computing and Science Laboratory (PICSL), Department of Radiology, University of Pennsylvania, Philadelphia, PA, USA
[b] Department of Computer & Information Science, University of Pennsylvania, Philadelphia, PA, USA
[c] Department of Neurology and Penn Memory Center, University of Pennsylvania, Philadelphia, PA, USA

ARTICLE INFO

Article history:
Accepted 10 May 2014
Available online xxxx

Keywords:
fMRI
ROI
Data-driven parcellations
PCA
MCI
Delayed recall
Default mode network

ABSTRACT

We present a new framework for prior-constrained sparse decomposition of matrices derived from the neuroimaging data and apply this method to functional network analysis of a clinically relevant population. Matrix decomposition methods are powerful dimensionality reduction tools that have found widespread use in neuroimaging. However, the unconstrained nature of these totally data-driven techniques makes it difficult to interpret the results in a domain where network-specific hypotheses may exist.

We propose a novel approach, Prior Based Eigenanatomy (p-Eigen), which seeks to identify a data-driven matrix decomposition but at the same time constrains the individual components by spatial anatomical priors (probabilistic ROIs). We formulate our novel solution in terms of prior-constrained ℓ_1 penalized (sparse) principal component analysis. p-Eigen starts with a common functional parcellation for all the subjects and refines it with subject-specific information. This enables modeling of the inter-subject variability in the functional parcel boundaries and allows us to construct subject-specific networks with reduced sensitivity to ROI placement.

We show that while still maintaining correspondence across subjects, p-Eigen extracts biologically-relevant and patient-specific functional parcels that facilitate hypothesis-driven network analysis. We construct default mode network (DMN) connectivity graphs using p-Eigen refined ROIs and use them in a classification paradigm. Our results show that the functional connectivity graphs derived from p-Eigen significantly aid classification of mild cognitive impairment (MCI) as well as the prediction of scores in a Delayed Recall memory task when compared to graph metrics derived from 1) standard registration-based seed ROI definitions, 2) totally data-driven ROIs, 3) a model based on standard demographics plus hippocampal volume as covariates, and 4) Ward Clustering based data-driven ROIs. In summary, p-Eigen incarnates a new class of prior-constrained dimensionality reduction tools that may improve our understanding of the relationship between MCI and functional connectivity.

Published by Elsevier Inc.

Introduction & related work

The presence of large and diverse neuroimaging datasets has brought the importance of data analysis techniques into focus. These issues are particularly salient in blood oxygen level dependent (BOLD) fMRI where recent papers have highlighted the sensitivity of this modality to specific analysis choices (Carp, 2012; Haller and Bartsch, 2009). Network analysis of functional connectivity within the brain from BOLD data has received a significant amount of attention and is notorious for being sensitive to analysis decisions (Dawson et al., 2012; Eke et al., 2012).

Functional connectivity is defined as the temporal co-activation of neuronal activation patterns between anatomically separated regions of the brain (Aertsen et al., 1989) and is thought to be an indicator of functional communication between these different regions. Typically, functional connectivity studies measure the level of correlation between the time-series of the resting state BOLD signal of the different brain regions (Biswal et al., 1997; Damoiseaux et al., 2006; Salvador et al., 2005). Studying the brain as an integrative network of functionally interacting brain regions can shed new light on large scale neuronal communication in the brain and how this communication is impaired in neurological diseases (Bullmore and Sporns, 2009; Mohammadi et al., 2009; Seeley et al., 2009).

There are two predominant approaches for the analysis of functional connectivity:

- Seed (ROI) based approaches: These are straightforward and operate in the traditional confirmatory network paradigm (Tukey, 1977). They involve computing the correlation between the time series of a given (preselected) *protect seed* brain region (ROI)[2] against all the other brain regions, resulting in a set of functional connectivity

* Corresponding author at: Penn Image Computing and Science Laboratory (PICSL), Department of Radiology, University of Pennsylvania, Philadelphia, PA, USA.
E-mail address: dhillon@cis.upenn.edu (P.S. Dhillon).
[1] Tel.: +1 215 588 9636.

[2] One can compute these correlations either voxelwise or by averaging over the voxels in an entire ROI.

http://dx.doi.org/10.1016/j.neuroimage.2014.05.026
1053-8119/Published by Elsevier Inc.

maps of the given brain regions (Biswal et al., 1997; Cordes et al., 2000). These functional connectivity maps can then be used to construct *resting-state-networks* of functionally correlated regions in the brain (Fox et al., 2005). The *ct seed* region can either be selected based on prior clinical knowledge or it can be selected from the activation map of a separate task dependent fMRI scan.

- Learning based approaches: These approaches use statistical techniques to explore functional connectivity in the brain, obviating the need to define a *seed* region. Typical methods employed are Principal Component Analysis (PCA) (Friston, 1998), Independent Component Analysis (ICA) or its variants e.g. Group ICA (Beckmann and Smith, 2004; Beckmann et al., 2005; Damoiseaux et al., 2006; Petrella et al., 2011; Varoquaux et al., 2010b) or hierarchical methods (Blumensath et al., 2013; Cordes et al., 2002; Salvador et al., 2005). These methods strive to find a set of orthogonal or independent signals in the time series that can explain the resting state activity patterns. ICA based methods are the popular methods in this setting as they can find a set of independent signals from whole brain voxelwise data and also due to the public availability of tools like MELODIC in FSL (Jenkinson et al., 2012) for ICA and Group ICA of fMRI Toolbox (GIFT) (Calhoun et al., 2001). Subsequently, one can create brain connectivity networks from the outputs of these approaches by computing correlations between the different (independent/orthogonal) signals they find.

The brain networks found by the above approaches are represented as a set of vertices (brain regions) connected by edges which represent the strength of correlation between those two regions (He and Evans, 2010; Stam et al., 2007). Various independent studies (surveyed here (van den Heuvel and Hulshoff Pol, 2010)) have consistently found a set of eight functional connectivity networks in the brain. One can use a set of key properties of the network graph e.g. clustering coefficient, centrality and modularity to get further insights into the flow of neuronal signals within a network (He and Evans, 2010; Stam et al., 2007).

The above-mentioned approaches for analyzing functional connectivity and constructing brain networks suffer from a variety of problems. The Group ICA based approaches do a group decomposition of the time series' images of the entire cohort; they have an averaging effect and erode away any subject specific characteristics of the network. So, the Group ICA analysis is usually followed by a back reconstruction step to generate subject-specific functional connectivity maps (Smith et al., 2011). However, it is unclear how to choose a statistically justified threshold to binarize these maps.

The seed based approaches also suffer from the problem of averaging the signal and may be sensitive to ROI placement (Zhang et al., 2012), co-registration errors and the specific ROI boundaries. These approaches assume that the signal lies totally within a predefined region. However, the important signal may have slightly different boundaries than the scientist's conception. The data representation (or spatially varying noise) may also lead to strong or weak signal within different parts of the ROI. Such dataset specific information is not taken into account by a traditional seed based approach. When effects are localized to the selected region, and that region is well-defined, a seed based analysis may provide the most sensitive testing method. However, some conditions involve a network of regions that may not be fully identified.

Furthermore, it has been shown that decreased/impaired functional connectivity in certain brain networks, for instance, the default mode network (DMN) has association with neurodegenerative disorders e.g. Alzheimer's disease (AD) (Greicius et al., 2004; Sheline et al., 2010), schizophrenia (Liu et al., 2008; Whitfield-Gabrieli et al., 2009), multiple sclerosis (MS) (Lowe et al., 2008), mild cognitive impairment (MCI) (Agosta et al., 2012; Bai et al., 2009; Hedden et al., 2009; Petrella et al., 2011). So, it has become even more imperative to improve statistical analysis methods to efficiently leverage the scarce patient BOLD fMRI data that is typically available.

In this paper, we propose a method that integrates ideas from both the seed based and learning based approaches. Our contributions in this paper are threefold.

(1) We contribute a general method for prior constrained eigen decomposition (p-Eigen) of high-dimensional matrices and a novel algorithm for its optimization.
(2) Publicly available implementation of our approach in C++.
(3) Application of p-Eigen for deriving subject specific functional parcellations from BOLD data (which are later used to derive functional networks) and an evaluation of these novel measurements in the context of MCI and prediction of delayed recall in a memory task.

Our approach provides a principled way of incorporating priors in an otherwise totally data-driven approach based on Sparse Principal Component Analysis (SPCA) (d'Aspremont et al., 2007; Shen and Huang, 2008; Witten et al., 2009; Zou et al., 2006).

p-Eigen allows an initial binary or probabilistic ROI to adapt to the underlying subject-specific covariation within the data. At the same time, p-Eigen maintains proximity to (and the locality of) the original region and thus retains the advantages of the standard seed based approach. p-Eigen also maintains non-negativity in the estimated anatomically-constrained eigenvector, thereby keeping ROI interpretability. This allows us to modify the definitions of labels to capture the variation in dataset (a given subject's time series) while still staying close to the initial ROI definitions. p-Eigen therefore produces labelings with "soft" weighted averages and as we show in the experimental section, are more sensitive to the underlying brain data than a standard ROI.

Given an ROI set, p-Eigen has only one key parameter to tune the weight of the prior term guiding the decomposition. Therefore, our optimization objective provides a tradeoff between 1) staying close to the initial ROI definitions and 2) allowing data to lead the exploratory analysis by explaining variance through PCA. A good way to think about this is as ROI definitions forcing us to be *conservative* and staying close to the initial brain parcellation; on the other hand the SPCA component gives us *liberty* to be either more exploratory or more focused on the content of the given dataset. The tradeoff between the two competing paradigms is defined by user tunable (prior strength) parameter, which is chosen via cross validation.

p-Eigen does a prior constrained sparse decomposition of each subject's time series image separately to create subject-specific functional networks, so it does not suffer from the problem of averaging as Group ICA does. Moreover, the priors help us maintain a direct correspondence between the anatomy of the same regions across different subjects hence leading to better clinical interpretability. Our proposed approach is shown in Fig. 1.

We have drawn a clear contrast between our approach and the two related approaches namely seed based approaches (no influence of data) and Group ICA/PCA based approaches (only data driven). That said, there has also been substantial work on incorporating prior information across subjects to build subject-specific functional networks as proposed by this paper.

Some early work that performed PCA on fMRI signal within ROIs (Nieto-Castanon et al., 2003) clearly foreshadowed p-Eigen. Thirion et al. (2006) also proposed a spectral learning based technique for parcellation that delineates homogeneous and connected regions across subjects, providing subject-specific functional networks.

The research that is perhaps closest to ours is Ng et al. (2009a), Deligianni et al. (2011) and Blumensath et al. (2013). Ng et al. (2009a) used group replicator dynamics (GRD) for finding sparse functional networks that are common across subjects but have subject-specific weightings of the brain regions. Langs et al. (2010) performed functional alignment across subjects to achieve improved functional correspondences across subjects. Deligianni et al. (2011) used brain anatomical connectivity to constrain the conditional

Please cite this article as: Dhillon, P.S., et al., Subject-specific functional parcellation via Prior Based Eigenanatomy, NeuroImage (2014), http://dx.doi.org/10.1016/j.neuroimage.2014.05.026