NeuroImage 108 (2015) 1–16

Contents lists available at ScienceDirect

NeuroImage

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

Finding imaging patterns of structural covariance via Non-Negative Matrix Factorization

Aristeidis Sotiras [a,*], Susan M. Resnick [b], Christos Davatzikos [a]

[a] Section for Biomedical Image Analysis, Center for Biomedical Image Computing and Analytics, University of Pennsylvania, Philadelphia, PA 19104, USA
[b] Laboratory of Behavioral Neuroscience, National Institute on Aging, Baltimore, MD 21224, USA

ARTICLE INFO

Article history:
Accepted 18 November 2014
Available online 12 December 2014

Keywords:
Data analysis
Structural covariance
Non-Negative Matrix Factorization
Principal Component Analysis
Independent Component Analysis
Diffusion Tensor Imaging
Fractional anisotropy
Structural Magnetic Resonance Imaging
Gray matter
RAVENS

ABSTRACT

In this paper, we investigate the use of Non-Negative Matrix Factorization (NNMF) for the analysis of structural neuroimaging data. The goal is to identify the brain regions that co-vary across individuals in a consistent way, hence potentially being part of underlying brain networks or otherwise influenced by underlying common mechanisms such as genetics and pathologies. NNMF offers a directly data-driven way of extracting relatively localized co-varying structural regions, thereby transcending limitations of Principal Component Analysis (PCA), Independent Component Analysis (ICA) and other related methods that tend to produce dispersed components of positive and negative loadings. In particular, leveraging upon the well known ability of NNMF to produce parts-based representations of image data, we derive decompositions that partition the brain into regions that vary in consistent ways across individuals. Importantly, these decompositions achieve dimensionality reduction via highly interpretable ways and generalize well to new data as shown via split-sample experiments. We empirically validate NNMF in two data sets: i) a Diffusion Tensor (DT) mouse brain development study, and ii) a structural Magnetic Resonance (sMR) study of human brain aging. We demonstrate the ability of NNMF to produce sparse parts-based representations of the data at various resolutions. These representations seem to follow what we know about the underlying functional organization of the brain and also capture some pathological processes. Moreover, we show that these low dimensional representations favorably compare to descriptions obtained with more commonly used matrix factorization methods like PCA and ICA.

Introduction

The very high structural and functional complexity of the brain, captured at some macroscopic level by various structural and functional imaging methods, has prompted the development of various analytical tools aiming to extract representations of structure and function from complex imaging data, which can help us better understand brain organization. Conventional methods based on pre-defined partitioning of the brain into anatomical regions of interest, such as lobes, gyri, and fiber tracts have been complemented in the past 15 years by data-driven methods, which aim to tease out of the data anatomical and functional entities that describe brain structure and function in an unbiased, hypothesis-free way.

A widely-used family of methods that are free from regional hypotheses falls under the umbrella of Voxel-Based Analysis (VBA) (Goldszal et al., 1998; Ashburner et al., 1998; Ashburner and Friston, 2000; Ashburner, 2009; Thompson et al., 2000; Fox et al., 2001; Davatzikos et al., 2001; Shen and Davatzikos, 2003; Studholme et al., 2004; Bernasconi et al., 2004). These voxel-wise methods often lack statistical power because they repeat the same test at each voxel thus leading to multiple comparison problems. Moreover, they do not fully exploit the available data since they are myopic to the correlations between different brain regions, which often reflect characteristics of underlying brain networks.

Conversely, Multivariate Analysis (MVA) (Norman et al., 2006; Ashburner and Klöppel, 2011; McIntosh and Mišić, 2013) takes advantage of dependencies among image elements and thus, enjoys increased sensitivity. MVA methods can be separated into two sub-classes: i) *confirmatory* MVA techniques, such as structural equation modeling (McIntosh and Gonzalez-Lima, 1994) and dynamic causal modeling (Friston et al., 2003), that aim to assess the fitness of an explicitly formulated model of interactions between brain regions; and ii) *exploratory* techniques, such as Principal Component Analysis (PCA) (Friston et al., 1993; Strother et al., 1995; Hansen et al., 1999) and Independent Component Analysis (ICA) (McKeown et al., 1998; Calhoun et al., 2001; Beckmann and Smith, 2004) that aim to recover linear or non-linear relationships across brain regions and characterize patterns of common behavior. One may additionally aim to relate the extracted components to demographic, cognitive and/or clinical variables by either employing

* Corresponding author at: Section for Biomedical Image Analysis, Center for Biomedical Image Computing and Analytics, University of Pennsylvania, 3600 Market Street, Suite 380, Philadelphia, PA 19106, USA. Fax: +1 215 614 0266.
E-mail address: Aristeidis.Sotiras@uphs.upenn.edu (A. Sotiras).

http://dx.doi.org/10.1016/j.neuroimage.2014.11.045
1053-8119/

techniques like Partial Least Squares (McIntosh et al., 1996; McIntosh and Lobaugh, 2004; Krishnan et al., 2011) and Canonical Correlation Analysis (Hotelling, 1936; Friman et al., 2001; Witten et al., 2009; Avants et al., 2014), or by using the PCA and ICA components as features in supervised discriminative settings towards identifying abnormal brain regions (Duchesne et al., 2008) or patterns of brain activity (Mourão Miranda et al., 2005, 2007).

However, standard MVA methods suffer from limitations related to the interpretability of their results. PCA and ICA, which are commonly applied in neuroimaging studies, estimate components and expansion coefficients that take both negative and positive values, thus modeling the data through complex mutual cancelation between component regions of opposite sign. The complex modeling of the data along with the often global spatial support of the components, which tend to highly overlap, result in representations that lack specificity. While it may be possible to interpret opposite phenomena that are encoded by the same component through the use of opposite signs, it is difficult to associate a specific brain region to a specific effect. Finally, ICA, and particularly PCA, aim to fit the training data well, resulting in components that capture in detail the variability of the training set, but often do not generalize as well in unseen data sets.

Non-Negative Matrix Factorization (NNMF) (Paatero and Tapper, 1994; Lee and Seung, 2000) is an unsupervised MVA method that enjoys increased interpretability and specificity compared to standard MVA techniques. NNMF estimates a predefined number of components along with associated expansion coefficients under the constraint that the elements of the factorization take non-negative values. This non-negativity constraint is the core difference between NNMF and standard MVA methods and the reason for its advantageous properties. It has been shown to lead to a parts-based representation of the data, where parts are combined in additive way to form a whole. Because of this advantageous data representation, NNMF has been applied in facial recognition (Zafeiriou et al., 2006), music transcription (Smaragdis and Brown, 2003), document clustering (Xu et al., 2003), machine learning (Hoyer, 2004; Cai et al., 2011), computer vision (Shashua and Hazan, 2005) and computational biology (Brunet et al., 2004; Devarajan, 2008).

However, the application of NNMF in medical imaging has been less investigated. In the case of structural imaging, a supervised approach for sparse non-negative decomposition was introduced by Batmanghelich et al. (2009, 2012). This approach targets feature extraction in the context of a generative–discriminative framework for high-dimensional image classification. The derived overcomplete representation exploits supervised knowledge to preserve discriminative signal in a clinically interpretable way through the adoption of sparsity and non-negativity constraints. The use of NNMF for feature extraction was also briefly suggested by Ashburner and Klöppel (2011). Lastly, unsupervised matrix-factorization approaches have been tailored for connectivity analysis to take into account the nature of the data (Ghanbari et al., 2012); (Eavani et al., 2013).

Contrary to previous approaches, we investigate in this paper the use of Non-Negative Matrix Factorization in an unsupervised setting as an analytical and interpretive tool in structural neuroimaging. The goal is to use NNMF to derive a distributed representation that will allow us to identify the brain regions that co-vary across individuals in a consistent way, hence potentially being part of underlying brain networks or otherwise influenced by underlying common mechanisms such as genetics and pathologies. We argue that NNMF is well-adapted for the analysis of neuroimaging data for four reasons: i) it provides a parts-based representation that facilitates the interpretability of the results in the context of brain networks; ii) it naturally produces sparse components that are localized and align well with anatomical structures; iii) it generalizes well to new data; and iv) it allows us to analyze the data at different resolutions by varying the number of estimated components. The second reason is of significant interest because in many scenarios, one expects co-varying networks to be formed by biologically related anatomical regions.

We apply the proposed approach to recover fractional anisotropy change relationships in mouse brain development, and to recover gray matter volume change relationships in human brain aging. Moreover, we contribute a useful and comprehensive comparison with PCA and ICA. We qualitatively and quantitatively evaluate the qualities of the respective representations and we demonstrate the superiority of the use of the proposed NNMF framework in detecting highly interpretable components that contain strongly co-varying regions. The high specificity of the obtained representation allows us to elucidate the distinct role of anatomical structures. As a consequence, we advocate the use of NNMF as an alternative tool for the study of structural co-variance (Seeley et al., 2009; Zielinski et al., 2010; Alexander-Bloch et al., 2013) that is now typically performed in a hypothesis-driven way assisted by Regions of Interest (ROI) or seed-based analysis.

Materials and methods

Methods

We consider a data set consisting of non-negative values that measure, through a medical image acquisition technique such as Magnetic Resonance Imaging, the expression of local biological properties of organ tissue for N samples. These samples typically represent different subjects, but they can also represent the same subject at different time points. For neuroimaging studies, the dimension D of the samples (i.e., images) is typically in the hundred of thousands, while the number of samples is typically in the hundreds. Thus, the data are represented by a tall matrix $\mathbf{X}$ that is organized by arraying each data sample per column ($\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$, $\mathbf{x}_i \in \mathbb{R}^D$, and a data sample refers to a vectorized image).

Our aim is to extract a relatively small set of components that capture multivariate relations between the variables and reflect the inherent variability of the data. Moreover, we would like to extract these components in a purely data-driven (unsupervised) fashion without prior regional hypotheses.

Regularized matrix factorization is a broad framework encompassing diverse techniques that factorize the data matrix $\mathbf{X}$ into two matrices satisfying constraints related to modeling assumptions:

$$\mathbf{X} \approx \mathbf{CL} \tag{1}$$

where $\mathbf{C}$ denotes the matrix that contains a component in each column ($\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_K]$, K is the number of the estimated components, $\mathbf{c}_i \in \mathbb{R}^D$ and is assumed to be a unit vector $\|\mathbf{c}_i\|^2 = 1$). $\mathbf{L}$ contains the loading coefficients that, when used together with $\mathbf{C}$, approximate the data matrix. Both $\mathbf{C}$ and $\mathbf{L}$ are necessary towards a comprehensive understanding of the data. $\mathbf{C}$ conveys information regarding the spatial properties of the variability effect, while the entries of $\mathbf{L}$ specify the strength of the effect in each data sample. Depending on the implemented modeling assumptions, $\mathbf{C}$ and $\mathbf{L}$ exhibit different properties. NNMF, PCA and ICA make different assumptions regarding the components whose linear combination generates the data. In this section, we study the three methods in terms of the assumptions they employ, giving particular emphasis on the NNMF framework that is the main focus of this work.

Non-Negative Matrix Factorization

Non-Negative Matrix Factorization produces a factorization that constrains the elements of both the components and the expansion coefficients to be non-negative. This factorization is typically achieved by solving the following energy minimization problem:

$$\begin{aligned} &\min_{\mathbf{C},\mathbf{L}} \quad \|\mathbf{X}-\mathbf{CL}\|_F^2 \\ &\text{subject to } \mathbf{C} \geq 0, \mathbf{L} \geq 0 \end{aligned} \tag{2}$$

where $\|\cdot\|_F^2$ is the squared Frobenius norm ($\|\mathbf{X}\|_F^2 = \mathrm{trace}(\mathbf{X}^T\mathbf{X})$). In other words, the optimal non-negative matrices are the ones that best