Independent Component Analysis (ICA) of functional MRI (fMRI) data

Category: Life sciences
Name: Seul Lee
SUNetID: seul 05809185

Introduction

Functional Magnetic Resonance Imaging (fMRI) is an MRI method for looking at brain functional connectivity by detecting blood flow changes. If some regions of the brain are activated, blood flow of the region increases. Therefore, fMRI looks for correlations between signal time series in different brain regions and identifies brain networks that are working in synchrony. However, desired activated signals are difficult to be detected due to unwanted noises. MR signal includes not only activated signals, but includes noise such as physiological noise (e.g. cardiac, respiration and motion) and thermal noise (e.g. signal drift due to scanner instability). Furthermore, signal variations caused by activation is low so that it is hard to statistically identify BOLD signal from MR signal with those noises.

Resting-state fMRI (RS-fMRI) has been widely used to map brain functional connectivity, but it is unclear how to probe network connectivity within and around lesions. The purpose of this study was to compare functional connectivity of brain regions affected by hemorrhagic traumatic axonal injury (hTAI) lesions to the functional connectivity of a healthy control using different analysis methods.

Related works

There are several methods for analyzing fMRI data: seed-based analysis(Biswal et al., 1995; Di Martino et al., 2008; Fox et al., 2005; van de Ven et al., 2004) and Independent Component Analysis (ICA)(Chen et al., 2008; McKeown et al., 2003; van de Ven et al., 2004). In seed-based analysis, region of interest (ROI) is identified, then in order to identify brain regions that are working in synchrony with this seed region, we look for correlations between the signal time-courses in the seeded-region and the signal time-courses from the voxels in the rest of the brain. It provides information about the connectivity in specific brain area of interest, however, it requires a priori selection of ROIs. ICA separates a signal into a set of spatially independent maps and corresponding time series and it can be used to identify spatial patterns of activation and noise in fMRI. ICA will be applied to 4D fMRI dataset and all the components correspond to either signal or noise. Usually, signal components show low frequency power and cover gray matter, however, noise components show a lot of high frequency power. In ICA, no prior knowledge of brain systems necessary and physiological noise can be automatically removed to a certain extent. However, it must manually select the recognizable components and distinguish noise from physiologic signals.

Theory

PCA

Before applying ICA on fMRI data, PCA is applied to fMRI data to reduce number of components. PCA is a linear transformation that transforms the data to a new coordinate system such that variance becomes maximized. The first component

has the greatest variance and the second component has the second greatest variance, and so on. In order to maximize the variance of the projection, sum of the distance of the projection should be maximized (Eqn. 1). With fMRI data matrix X (each column has zero-mean), we can think that we project the data to the number of components axis. PCA decomposition is found using SVD and U and V are an orthogonal matrix and Σ^2 is the squares of the singular values.

$$XX^T = \sum_{i=1}^{m} x^{(i)} x^{(i)^T} = U\Sigma V^T V \Sigma U^T = U\Sigma^2 U^T \quad \dots (1)$$

By thresholding the power, the components that has significant amounts of signal can be found. The result components have reduced number of components and they can be ready to be applied ICA.

ICA

For ICA, since we assume that the sources are independent, we can think of x = As = $W^{-1}s$, where x is observed signal and W^{-1} is mixing matrix.

$$l(W) = \sum_{i=1}^{m} (\sum_{i=1}^{m} log\, g'({w_i}^T x^{(i)}) + log|W|) \text{ , where } g(x) = \frac{1}{1+e^{-x}} \quad (2)$$

In order to maximize log likelihood of given data set (Eqn. 2), we should maximize this in terms of the mixing matrix W. By taking derivatives in terms of W and set to zero, the update rule can be calculated by Eqn.3.

$$W := W + \alpha \left(\begin{bmatrix} 1 - 2g({w_1}^T x^{(i)}) \\ 2 - 2g({w_2}^T x^{(i)}) \\ \vdots \\ 2 - 2g({w_m}^T x^{(i)}) \end{bmatrix} x^{(i)^T} + (W^T)^{-1} \right) \text{ , where } \alpha \text{ is a step size.}$$

(3)

Until the gradient converges (< 10^{-3}), updating steps are repeated.

For the ICA results, connectivity maps are shown, which indicates how well the timeframes correlates with the timeframes of the region of interest (ROI). As assumption for ICA is that fMRI data consists of linear sum of spatially independent patterns of activity, each activation in brain can be modeled spatially independent components and the sum of them creates brain activity map.

Seed-based analysis

Once a seed is placed in a region of interest (ROI), we are able to compare the timeframe data to those from the rest of the brain. The correlation/connectivity map can be displayed with a threshold value and the regions where are functionally connected can be recognized.

Dataset and Features

Data acquisition: MR data were acquired using a 7T Siemens whole-body MRI scanner (Siemens Healthcare, Erlangen, Germany). All study procedures were approved by local Institutional Review Board and written informed consent was provided by all subjects. **Preprocessing**: The pre-processing pipeline included creating a mask of the brain using the Brain Extraction Tool (BET), detrending (subtracting mean of the data), regression of nuisance variables (WM and CSF: only ventricular CSF is included, movement and global signal), motion correction using MCFLIRT (Jenkinson et al., 2002) and spatial smoothing with a 3mm Gaussian kernel. The nuisance regressor masks were eroded to avoid contamination from the gray matter (GM) and only ventricular CSF was included in the CSF regressor mask (no peripheral CSF). After testing connectivity maps with and without smoothing (2mm and 3mm Gaussian kernel), a 3mm Gaussian

kernel was selected because it provided increased sensitivity but no change in the connectivity patterns observed. The first 6 volumes were discarded to avoid data acquired during the approach to steady-state magnetization. All the preprocessing steps are done using MATLAB.

Methods

Data analysis: FMRI data consists of 4D data: 128 × 128 voxels × 60 slices × 180 timeframes. (Figure 1) ICA decomposes 4D data sets into different spatial components. In order to apply PCA and ICA, 4D data were converted to (128 × 128 × 60=983040) × 180 matrix. The converted data were subtracted by mean of the each timeframe data so that each spatial map had zero mean.

1) PCA: To reduce the number of components, initial number of components was set to 100 and eigenvalues of each component were calculated using SVD (Figure 2). From the descending ordered eigenvalues, the number of component was set to 30, 60 and 100 and compared their ICA results.

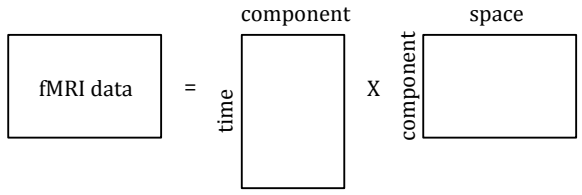

Figure 1. FMRI data set

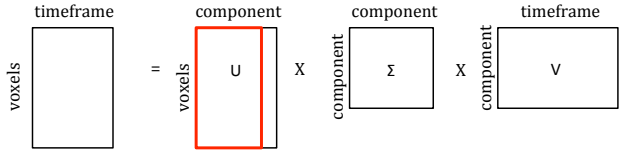

Figure 2. PCA: SVD of fMRI data. Each column of U corresponds to component. Red box represents the largest 30/60/100 components.

2) ICA

The update rule that was explained in theory was applied to the reduced component. W is initialized from random matrix. From the reduced components, I chose well-known resting-state networks: Default Mode Network (DMN) area, Right Executive Control Network (ECN) and Primary visual network from both a patient and healthy control.

3) Seed-based analysis

With fMRI data from a patient and healthy control, an ROI was placed in Default Mode Network (DMN), Right Executive Control Network (ECN) and Primary visual network and correlation maps were found using the Analysis of Functional NeuroImages (AFNI) (Cox, 1996; Cox and Hyde, 1997; Gold et al., 1998).

Results

PCA: The number of components started from 100 and the eigenvalues are shown in Figure 3 (only patient data are shown) and ratios of variance explained of each component are shown in Figure 4 (only patient data are shown). From the graph of eigenvalues/variance, the number of components sets to 30, 60 and 100. I compared the recognized networks with 30, 60 and 100 components and 30 components still showed higher sensitivity from the ICA results and only 30 component results are shown in this study. More than 60 components showed over-fitted results. In practice, 30 components are normally used for fMRI ICA. The timeframes of the first 5 components of the patient and the control are shown in Figure 5.

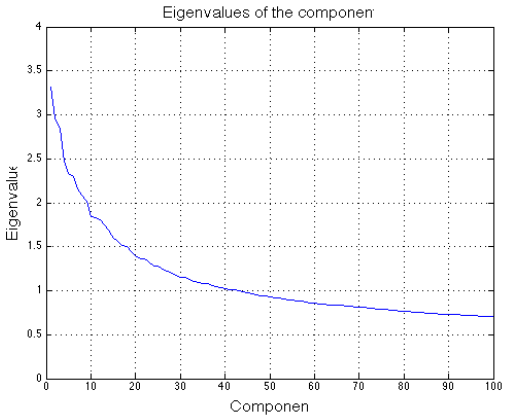

Figure 3. Descending ordered eigenvalues of the components

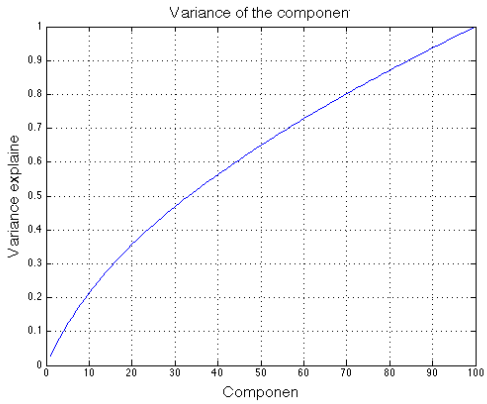

Figure 4. Variance explained with respect to the number of components

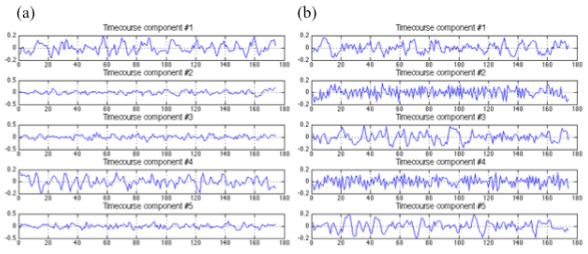

Figure 5. The timeframes of the first 5 components from PCA. (a) patient, (b) control

ICA: Well-known resting-state networks were identified from ICA in both patient and control. I compared those to the networks from the FIND Lab atlas (Shirer et al., 2012). To compare with seed-based analysis, Default Mode Network (DMN), Right Executive Control Network (ECN), primary visual network were chosen (Figure 6).

Seed-based analysis: I placed an ROI in DMN area, and looked for correlations between the signal timeframes in the ROI and the signal timeframes from the voxels in the rest of the brain. The connectivity map is shown in Figure 6 (significance threshold of $P < 0.05$).

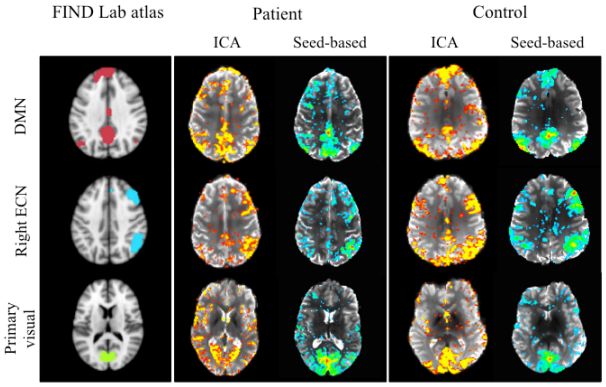

Figure 6. Connectivity maps of the patient and control from ICA and seed-based analysis.

Discussion

In this study, I compared different analysis methods for fMRI: ICA and seed-based analysis. If one has information about the connectivity in specific brain area of interest, seed-based analysis can be applied. Otherwise, ICA can be used to identify resting-state networks. ICA algorithm follows update rule that is shown in lecture note. I applied both ICA and seed-based analysis to a patient who has lesions and a control and showed that both methods recognized resting-state networks successfully. One of the challenges of ICA is to decide number of components, however, PCA can be used to decide the number of components. Using SVD, eigenvalues and

variance are calculated in terms of number of components and based on that, proper number of components for ICA is calculated in order to avoid over-fitting and under-fitting. How lesions in brain alter functional connectivity can be a next step, however, seed-based analysis could be easier compared to ICA because seed can be placed on lesion where we are interested.

Acknowledgements

The fMRI data were acquired at Harvard Medical School, Massachusetts General Hospital and I would like to thank Jonathan R. Polimeni, Brian L. Edlow for fMRI data and also thank Bo Wang, Jane Bae and Jingyuan Chen for helpful discussions. Support for this work was provided by the NIH including: NINDS R25-NS065743, NINDS K23NS094538, NIBIB K01-EB011498, NIBIB R01-EB019437, NIBIB P41-EB015891, NCRR S10-RR026351 and by the Athinoula A. Martinos Center for Biomedical Imaging. Funding was also provided by Center for Integration of Medicine and Innovative Technology (Boston, MA), American Academy of Neurology/American Brain Foundation, a Li Ka Shing-Oxford-Stanford Big Data in Human Health Seed Grant, Stanford Radiology Angel Funds and GE Healthcare.

References

Biswal, B., Yetkin, F.Z., Haughton, V.M., Hyde, J.S., 1995. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magn Reson Med 34, 537-541.

Chen, S., Ross, T.J., Zhan, W., Myers, C.S., Chuang, K.S., Heishman, S.J., Stein, E.A., Yang, Y., 2008. Group independent component analysis reveals consistent resting-state networks across multiple sessions. Brain Res 1239, 141-151.

Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29, 162-173.

Cox, R.W., Hyde, J.S., 1997. Software tools for analysis and visualization of fMRI data. NMR Biomed 10, 171-178.

Di Martino, A., Scheres, A., Margulies, D.S., Kelly, A.M., Uddin, L.Q., Shehzad, Z., Biswal, B., Walters, J.R., Castellanos, F.X., Milham, M.P., 2008. Functional connectivity of human striatum: a resting state FMRI study. Cereb Cortex 18, 2735-2747.

Fox, M.D., Snyder, A.Z., Vincent, J.L., Corbetta, M., Van Essen, D.C., Raichle, M.E., 2005. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc Natl Acad Sci U S A 102, 9673-9678.

Gold, S., Christian, B., Arndt, S., Zeien, G., Cizadlo, T., Johnson, D.L., Flaum, M., Andreasen, N.C., 1998. Functional MRI statistical software packages: a comparative analysis. Hum Brain Mapp 6, 73-84.

Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17, 825-841.

McKeown, M.J., Hansen, L.K., Sejnowsk, T.J., 2003. Independent component analysis of functional MRI: what is signal and what is noise? Curr Opin Neurobiol 13, 620-629.

Shirer, W.R., Ryali, S., Rykhlevskaia, E., Menon, V., Greicius, M.D., 2012. Decoding subject-driven cognitive states with whole-brain connectivity patterns. Cereb Cortex 22, 158-165.

van de Ven, V.G., Formisano, E., Prvulovic, D., Roeder, C.H., Linden, D.E., 2004. Functional connectivity as revealed by spatial independent component analysis

of fMRI measurements during rest. Hum Brain Mapp 22, 165-178.