

Comparative study of several multivariate fMRI processing methods: PCA, Factor Analysis, Infomax, FASTICA, MELODIC

R. Arya¹, V. Calhoun², S. Roys¹, T. Adali³, J. Greenspan⁴, R. Gullapalli¹

¹Dept. of Radiology, Univ of Maryland Medical Center, Baltimore, MD, United States, ²Instit of Living/Dept of Psychiatry, Yale Univ, New Haven, CT, United States,

³Dept of Electrical Engg., Univ of Maryland Baltimore County, Baltimore, MD, United States, ⁴Dept of Oral & Cranial Biol Sciences, Univ of Maryland Medical Center, Baltimore, MD, United States

Abstract: Various blind source separation techniques have been used for distinguishing different signal sources from noisy data. Independent component analysis (ICA) algorithms which use higher order statistics to separate signal sources appear to be a promising tool for fMRI data analysis. In this study, we have compared several ICA algorithms and implementations: Infomax, FASTICA, FSL (MELODIC), and also other multivariate methods: Factor analysis and principle component analysis.

Introduction: Several blind source analysis techniques have been used in last few years to analyze fMRI data. Principle component analysis (PCA) projects the data onto a lower dimensional space minimizing the mean square error between the original data set and the reduced data set. Factor Analysis starts with PCA, then selects a reduced number of components that are rotated to counteract the effects of random error (or bias). Hence, the components are no longer pair-wise orthogonal. ICA approaches include Infomax, FASTICA and MELODIC. The Infomax algorithm minimizes the mutual information between the estimated sources.¹ The FASTICA implementation maximizes the negentropy (approximated by kurtosis) of the output.² Melodic uses modified FASTICA algorithm which normalizes the variance of the data and uses pow3 as the non-linearity (compared to the default tanh nonlinearity used in FASTICA).⁴ Since these algorithms take different paths to arrive at a functional map, the performance of these algorithms was determined at different signal to noise levels using simulated data and fMRI data.

Methods: A digital phantom containing four signal sources similar to those observed in fMRI data was generated as shown in fig1. Spatial sources were randomly chosen from rayleigh, lognormal, chi-square, and laplacian such that their distribution was skewed or kurtotic.³ The samples were then sorted and arranged in a fashion such that the higher intensity pixels were centered at each corner of the image representing cardiac, respiratory, transient task related, and the task waveform convolved with hemodynamic response function. These signals were then overlaid on a uniform image with an additive white gaussian noise of zero mean and variance one. The standard deviation of the noise was increased in small steps to study the performance of Infomax, FASTICA, MELODIC, Factor Analysis and PCA. We defined our signal to noise measure as the sum of the standard deviation of the sources divided by the standard deviation of overall noise sources. A performance index was obtained by correlating with the reference waveform.⁵ Each stage in these simulations was repeated twenty times to take into account the variability associated with the initial guess as well as random noise. The corresponding correlation coefficients were then averaged to get the mean and standard deviation. Motor data was obtained on six subjects using a Philips Eclipse 1.5T scanner equipped with high-speed gradients. An EPI pulse sequence was used at a TR of 320 ms and TE of 35 ms. The paradigm used was 20 seconds of finger thumb opposition task followed by 20 seconds of rest for 8 cycles. Each algorithm was applied 20 times for all the fMRI datasets to account for randomness due to initial guess. Mean and standard deviation of the correlation values were calculated.

Results: Fig2 shows the correlation values with increasing noise levels for task waveform when applied to simulated data. At higher SNRs, Infomax, FASTICA and MELODIC performed equally well but as the SNR decreased, the performance of Infomax was superior to others. Table1 shows the mean and standard deviation of correlation coefficient for these algorithms when applied on Motor data.

Discussion: Infomax performed better in terms of estimating functional maps and their associated time series, likely because infomax uses mutual information, thus incorporating higher order statistics, which are an important measure of independence. Whereas, FASTICA and MELODIC utilize only fourth order statistics. At higher SNRs, MELODIC performed slightly better than FASTICA, but as SNR decreased, the opposite results were noted. As both of FASTICA and Melodic use fourth order statistics, they performed better than Factor Analysis and PCA, as the latter two use second order statistics to separate the sources. The performance of Factor Analysis and PCA was very close, the fact that factor analysis performed slightly better could be attributed to bias adjustment in Factor analysis algorithm.. To further differentiate between FASTICA and MELODIC, the choice of non-linearity is important. The choice of pow3, a non-symmetric non-linearity, makes an a priori decision about the amount of asymmetry between activation and deactivations whereas tanh assume similar amounts of activation and deactivation. It is not yet clear which choice is correct for real fMRI data. Our finding here suggests that all ICA algorithm performed similarly well on the motor data which has reasonable high signal to noise. Data that suffers from low SNR or functional activation that tend to be weak may benefit through the use of Infomax which seems to be effective at low SNRs.

References:

1. Bell AJ, Sejnowski TJ. *Neural Comput.* 7:1129-1159; 1995
2. A. Hyvarinen and E. Oja. *Neural Comput.* 9:1483-1492; 1997
3. Stone JV, *NeuroImage*, 15(2), p407, 2002.
4. Beckmann, C. et al. *Sixth Int. Conf. on func. mapp. Of the Human Brain*, p643, 2000b.
5. Calhoun VD et al *Journal of VLSI Sig. Proc. Sys.*, 2002