ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

Dimension Reduction for Individual ICA to Decompose FMRI during Real-World Experiences: Principal Component Analysis vs. Canonical Correlation Analysis

Valeri Tsatsishvili[1], Fengyu Cong[1], Tuomas Puoliväli[1], Vinoo Alluri[1,2], Petri Toiviainen[2], Asoke K Nandi[1,3,4], Elvira Brattico[4], Tapani Ristaniemi[1*]

1. -University of Jyväskylä, Department of Mathematical Information Technology, PO Box 35, FI-40014, Finland
2. - Finnish Centre of Excellence in Interdisciplinary Music Research, University of Jyväskylä, Department of Music, PO Box 35, FI-40014,Finland
3. -Brunel University, Department of Electronic and Computer Engineering, Uxbridge, Middlesex UB8 3PH, UK
4. -University of Helsinki, Institute of Behavioral Sciences, Cognitive Brain Research Unit, PL 9, 466, Finland

Abstract. Group independent component analysis (ICA) with special assumptions is often used for analyzing functional magnetic resonance imaging (fMRI) data. Before ICA, dimension reduction is applied to separate signal and noise subspaces. For analyzing noisy fMRI data of individual participants in free-listening to naturalistic and long music, we applied individual ICA and therefore avoided the assumptions of Group ICA. We also compared principal component analysis (PCA) and canonical correlation analysis (CCA) for dimension reduction of such fMRI data. We found interesting brain activity associated with music across majority of participants, and found that PCA and CCA were comparable for dimension reduction.

1. Introduction

Study of brain activations elicited by natural continuous auditory and visual stimuli is relatively new and a promising domain in the field of fMRI research[1-3].Generated brain responses by such stimuli are of much more complex nature than in commonly utilized controlled design (block or event-related) experiments. This yields to adopting more data-driven approaches rather than holding on more traditional methods following to the hypothesis-driven models[4]. Group ICA has been used for analyzing fMRI during real-world experiences [4,5]. Assumptions for Group ICA require at least the number of sources and their order to be invariant for different subjects [6]. However, it is unknown whether these assumptions are met in real life. Therefore, in this study we apply individual ICA to each participant’s fMRI dataset elicited by naturalistic, continuous and long piece of music.

[*] This work was financially supported by TEKES (Finland) under grant40334/10 “Machine Learning for Future Music and Learning Technologies”. A.K. Nandi would like to thank TEKES for their award of the Finland Distinguished Professorship.

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

Before subjecting fMRI to ICA decomposition, dimension reduction using PCA with model order selection is a common pre-processing routine that helps in identifying and separating signal and noise subspaces.

Although fairly old method, CCA [7] has only been recently employed for pre-processing [8] or post-processing [5] fMRI data. It finds correlated and uncorrelated subspaces from two datasets using second order statistics [8,9]. In an experiment to collect brain data, it is often expected to find the common information across different participants belonging to the same group. Therefore, CCA theoretically matches this goal, and its strength in the dimension reduction for ICA has been shown through the analysis of simulated and real fMRI data obtained during the controlled design experiment[8]. However, it is unknown whether CCA can also work well as the pre-processing step for ICA to decompose very noisy fMRI data elicited during real-world experiences. Present study compares performances of CCA, implemented according to [8], and more widely used PCA for dimension reduction for ICA.

2. Method

2.1. Data description

Dataset here consists of continuous fMRI scans (time resolution was 2 seconds) obtained from eleven healthy musicians (mean age: 23.2 ± 3.7 SD; 5 females) while they listened to the tango 'Adios Nonino' by Astor Piazzolla with duration of 8 minutes and 32 seconds. Six high-level musical features including Fullness, Brightness, Timbral Complexity, Key Clarity, Pulse Clarity, and Activity were extracted from the stimulus. Detailed information about the fMRI data can be found in [1].

2.2. Dimension Reduction

2.2.1 PCA

If we denote the matrix of observed centered (zero mean) signals by $\boldsymbol{X} \in \mathbb{R}^{n\times l},\ l \gg n$, then the goal of PCA is to find orthogonal transform diagonalizing the covariance matrix of $\boldsymbol{X}, \boldsymbol{C_{xx}} = \frac{\boldsymbol{1}}{\boldsymbol{n}}\boldsymbol{X}\boldsymbol{X^T}$. This is achieved e.g. by eigenvalue decomposition:

$$\boldsymbol{C_{xx}V} = \boldsymbol{DV},$$

where$\boldsymbol{V} \in \mathbb{R}^{n\times n}$ is a matrix whose each column contains eigenvectors and $\boldsymbol{D}$ is a diagonal matrix of eigenvalues ranked decreasingly.

2.2.2 CCA

While PCA analyses one dataset at a time, CCA analyses two datasets to measure linear relationships between them. It finds two bases $\boldsymbol{W_1}$and $\boldsymbol{W_2} \in \mathbb{R}^{n\times n}$for two centered data matrices $\boldsymbol{Y_1}$ and $\boldsymbol{Y_2} \in \mathbb{R}^{n\times l}$, such that correlations between the projections $\boldsymbol{Z_1} = \boldsymbol{Y_1}^{\boldsymbol{T}}\boldsymbol{W_1}$ and $\boldsymbol{Z_2} = \boldsymbol{Y_2}^{\boldsymbol{T}}\boldsymbol{W_2}$ are mutually maximized. According to [8] CCA can be calculated by singular value decomposition of cross-covariance matrix of two whitened and normalized datasets:

$$\boldsymbol{C_{y_1y_2}} = \mathbf{U\Sigma Y}^{\boldsymbol{T}},$$

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

where $\boldsymbol{C}_{\boldsymbol{y_1 y_2}} = \frac{\mathbf{1}}{\boldsymbol{n}}\boldsymbol{Y_1}\boldsymbol{Y_2}^{\boldsymbol{T}}$ is the cross-covariance matrix, $\mathbf{U}$ and $\boldsymbol{\Upsilon}$ are two orthogonal bases, one for each input dataset, and $\boldsymbol{\Sigma}$ contains singular values representing the canonical correlations.

2.2.3 Dimension reduction

In the both methods described above, dimension reduction is achieved in similar fashion. For PCA first *k* eigenvectors are selected from $\boldsymbol{V} \in \mathbb{R}^{n\times n}$ basis, such that it becomes $\boldsymbol{\mathcal{V}} \in \mathbb{R}^{n\times k}$, and then input dataset $\boldsymbol{X}$ is projected onto it: $\boldsymbol{\mathfrak{X}} = \boldsymbol{\mathcal{V}}^{\boldsymbol{T}}\boldsymbol{X}$. The procedure is similar for CCA where $\boldsymbol{Y_1}$and $\boldsymbol{Y_2}$ are projected onto $\mathbf{U}$ and $\boldsymbol{\Upsilon}$.

Neither of the presented methods estimates target dimensionality of the input data automatically. In fact, evaluating the number of sources (i.e. target dimensions) is one of the challenges in fMRI analysis, which is frequently solved by empirical approaches [10]. Nevertheless, several methods for estimation of number of sources from the data have been proposed [11,12]. We employed model order selection method Gap proposed in [13] and previously employed for EEG data due to its computing efficiency [14]. With different numbers of sources experimented, the strength of Gap was examined for dimension reduction.

2.3. ICA decomposition

In this study we decompose each participant's fMRI dataset separately using spatial ICA, as opposed to the group-level approach where the data is concatenated first. The model of spatial ICA is $\boldsymbol{x} = \boldsymbol{\mathcal{A}s}$, where $\boldsymbol{x} \in \mathbb{R}^{n\times l}$ is a matrix of fMRI scans (*n* denotes time points and *l* - voxels), $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n\times k}$is the mixing matrix containing respective time courses of the sources in $\boldsymbol{s}$, and $\boldsymbol{s} \in \mathbb{R}^{k\times l}$ is the source matrix containing spatial activation patterns. If we denote dataset after dimension reduction by $\mathbf{x}$, then the above model will become determined by $\mathbf{x} = \boldsymbol{\mathcal{V}}^{\boldsymbol{T}}\boldsymbol{\mathcal{A}s} = \boldsymbol{As}$, where $\boldsymbol{\mathcal{V}}^{\boldsymbol{T}} \in \mathbb{R}^{k\times n}$ is dimension reduction matrix obtained from the dimension reduction method, and $\boldsymbol{A} = \boldsymbol{\mathcal{V}}^{\boldsymbol{T}}\boldsymbol{\mathcal{A}}$, $\boldsymbol{A} \in \mathbb{R}^{k\times k}$ becomes the mixing matrix of the determined ICA model. The goal is to learn unmixing matrix $\mathbf{W}$ such that: $\mathbf{y} = \mathbf{Wx}$. After the decomposition, original time courses of extracted sources are reconstructed by projecting extracted sources back to the scan field [15] via $\mathbf{U} = \mathbf{V}\mathbf{W}^{-\mathbf{1}}$.

As a stochastic algorithm ICA is not intrinsically stable and therefore, it can provide different results if run several times. A software package Icasso [16] analyzes the stability and robustness of ICA decomposition. The idea of Icasso is to run ICA repeatedly N times (N=100 in this study), each time with randomly initialized unmixing matrix and to cluster extracted independent components into the predefined number of clusters. In this study, FastICA algorithm with the nonlinear function *tanh* was selected as the separation algorithm. For the clustering, the agglomerative hierarchical clustering with average-linkage criterion was used. The number of clusters was the same to the number of components extracted by ICA. For characterizing decomposition stability, cluster quality index $\boldsymbol{I_q}$ was calculated, which is a parameter estimating compactness of each cluster and degree of separation from others [16]. It is calculated by:

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

$$\boldsymbol{I_q} = \frac{\mathbf{1}}{|\boldsymbol{C_m}|^2} \sum_{i,j \in C_m} \boldsymbol{\sigma_{ij}} - \frac{\mathbf{1}}{|\boldsymbol{C_m}||\boldsymbol{C_{-m}}|} \sum_{i \in C_m} \sum_{j \in C_{-m}} \boldsymbol{\sigma_{ij}}$$

where $\boldsymbol{C_m}$ denotes the set of estimated independent components in the cluster $\boldsymbol{m}$, $|\boldsymbol{C_m}|$ is the size of the cluster, $\boldsymbol{C_{-m}}$ is the set of indices outside the cluster $\boldsymbol{m}$, and $\boldsymbol{\sigma_{ij}}$ is an absolute value of mutual correlations between estimated independent components. It is a good measure for estimating stability of the extracted component as well as detecting possible overfitting. Therefore, $\boldsymbol{I_q}$ is a suitable parameter for performance comparison of employed dimension reduction algorithms.

2.4. Individual-level data processing

Obtained fMRI images went through the pre-processing procedure described in [1]. Next, temporal course of each voxel in the dataset was filtered using digital filter based on Fourier transform. The cut-off frequencies of the band-pass filter were set to 0.008Hz and 0.05Hz, determined by power spectrum of stimulus feature time series.

Dimension of the filtered data was reduced using two different methods. First, PCA and Gap were employed, where Gap estimated 46 sources. Next, CCA was performed on six pairs of subjects. We implemented CCA according to the algorithm proposed in [8]. However, for dimension reduction authors in [8] rejected CCA components with correlations below an arbitrary threshold of 0.5. Here we employed Gap method again that determined different number of sources for different pairs of datasets, varying between 43 and 45. To test if Gap performance was optimal we also experimented with different numbers of sources (k=20 and k=30).

Resulted six datasets (three for each dimension reduction method) were separately decomposed using Icasso [16].

2.5. Group-level data analysis

Obtained time courses of independent components were correlated with time courses of stimulus features. Significance thresholds of the correlations were set using Monte-Carlo simulation presented in [1] and only significant correlations at the significance level $p<0.01$ were considered for further analysis. Finally, spatial maps with significant correlations were visually examined to find common stimulus-related brain activations. We considered common activation map only if it was shared between more than five (half of all) participants.

3. Results

For compactness of representation we denote CCA and PCA-based ICA results as PCA+ICA and CCA+ICA. Experiments showed that ICA decomposition stability is affected little by employed dimension reduction method. In the Fig.1 quality indexes for CCA and PCA are provided. Indeed, for all numbers of components the difference between two methods for mean ICA decomposition stability is subtle.

Visual examination of activation maps significantly correlated with one or more musical features ($p<0.01$) revealed one common map showing activation in Auditory cortex, shared between more than five participants. Table 1 summarizes the observed

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

common map for PCA and CCA. In overall, the spatial map was detected in activations of nine subjects for PCA+ICA and seven subjects for CCA+ICA. Due to the space limitation, spatial maps are not shown.

Manually reducing dimensionality to 20 and 30 resulted in less stable ICA decomposition for CCA as well as PCA. However, desired common map was still observed for both methods: for PCA+ICA among seven and six participants respectively. For CCA+ICA the common map was found in 7 participants' activations regardless of the number of sources.

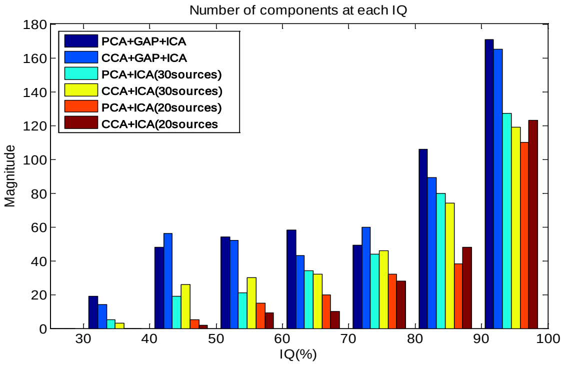

Fig 1: Cluster quality indexes for CCA and PCA for GAP,30, and 20 components.

Feature	Subject numbers	
	CCA+GAP+ICA	**PCA+GAP+ICA**
Fulness	3,4,5,10	3,7,10
Brightness	3,4,5,7,9,11	1,3,4,5,7,8,9,10,11
Timb. Complexity	0	8,11
Key Clarity	0	3
Pulse Clarity	0	0
Activity	3,4,5,9,10,11	3,4,7,10

Table 1: Summary of common spatial map. Numbers represent numbering of subjects and zero denotes absence of participant for which observed map was significantly correlated with acoustic features.

4. Conclusions

In order to study fMRI during real-world experiences, we proposed an individual-level data processing and group-level analysis approach mainly based on ICA and correlation. Meanwhile, two different methods for dimensionality reduction were tested for ICA in processing such challenging data.

We found similar spatial maps with corresponding temporal courses significantly correlated with musical features among individual participants. For dimension reduction in processing fMRI during real-world experiences, we found both PCA and CCA performed reasonably well.

In addition, we repeated the process with two different numbers of sources to check whether the employed model order selection was optimal in estimating number of target dimensions. We found that the number of sources suggested by model order selection was optimal for ICA decomposition stability for both methods. Interestingly, in production of stimulus-related spatial maps CCA was less sensitive to lower dimensions than PCA in our experiment.

It should be noted that CCA was implemented according to [8], which does not precisely follow the conventional CCA definition [7]. In the future we will investigate the conventional CCA and Partial least squares [17] for dimension reduction of fMRI during real-world experiences.

References

[1] V. Alluri, P. Toiviainen, I.P. Jaaskelainen, E. Glerean, M. Sams, E. Brattico, Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm, Neuroimage. 59 (2012) 3677-3689.

[2] J.P. Kauppi, I.P. Jääskeläinen, M. Sams, J. Tohka, Inter-subject correlation of brain hemodynamic responses during watching a movie: localization in space and frequency, Frontiers in neuroinformatics. 4 (2010) DOI: 10.3389/fninf.2010.00005.

[3] K.N. Carvalho, G.D. Pearlson, R.S. Astur, V.D. Calhoun, Simulated Driving and Brain Imaging, CNS Spectr. 11 (2006) 52-62.

[4] S. Malinen, Y. Hlushchuk, R. Hari, Towards natural stimulation in fMRI—Issues of data analysis, Neuroimage. 35 (2007) 131-139.

[5] J. Ylipaavalniemi, E. Savia, S. Malinen, R. Hari, R. Vigário, S. Kaski, Dependencies between stimuli and spatially independent fMRI sources: Towards brain correlates of natural stimuli, Neuroimage. 48 (2009) 176-185.

[6] F. Cong, Z. He, J. Hämäläinen, P.H.T. Leppänen, H. Lyytinen, A. Cichocki, T. Ristaniemi, Validating rationale of group-level component analysis based on estimating number of sources in EEG through model order selection, J. Neurosci. Methods. 212 (2013) 165-172.

[7] H. Hotelling, Relations between two sets of variates, Biometrika. 28 (1936) 321-377.

[8] J. Karhunen, T. Hao, J. Ylipaavalniemi, A canonical correlation analysis based method for improving BSS of two related data sets, Latent Variable Analysis and Signal Separation. Lecture Notes in Computer Science. 7191 (2012) 91-98.

[9] M. Koskinen, J. Viinikanoja, M. Kurimo, A. Klami, S. Kaski, R. Hari, Identifying fragments of natural speech from the listener's MEG signals, Hum. Brain Mapp. DOI: 10.1002/hbm.22004 (2012).

[10] V. van de Ven, C. Bledowski, D. Prvulovic, R. Goebel, E. Formisano, F. Di Salle, D.E.J. Linden, F. Esposito, Visual target modulation of functional connectivity networks revealed by self-organizing group ICA, Hum. Brain Mapp. 29 (2007) 1450-1461.

[11] D. Cordes, R.R. Nandy, Estimation of the intrinsic dimensionality of fMRI data, Neuroimage. 29 (2006) 145-154.

[12] Y.O. Li, T. Adalı, V.D. Calhoun, Estimating the number of independent components for functional magnetic resonance imaging data, Hum. Brain Mapp. 28 (2007) 1251-1266.

[13] Zhaoshui He, A. Cichocki, Shengli Xie, Kyuwan Choi, Detecting the Number of Clusters in n-Way Probabilistic Clustering, Pattern Analysis and Machine Intelligence, IEEE Transactions on. 32 (2010) 2006-2021.

[14] F. Cong, Z. He, J. Hamalainen, A. Cichocki, T. Ristaniemi, Determining the number of sources in high-density EEG recordings of event-related potentials by model order selection, (IEEE MLSP2011) 1-6.

[15] F. Cong, I. Kalyakin, Z. Chang, T. Ristaniemi, Analysis on subtracting projection of extracted independent components from EEG recordings, Biomed. Tech. (Berl). 56 (2011) 223-234.

[16] J. Himberg, A. Hyvarinen, F. Esposito, Validating the independent components of neuroimaging time series via clustering and visualization, Neuroimage. 22 (2004) 1214-1222.

[17] H. Wold, Soft modelling : The Basic Design and Some Extensions, Systems Under Indirect Observation, Part II. (1982) 36-37.