Spatial Discriminant ICA for RS-fMRI characterisation

Alejandro Tabas
Faculty of Science and Technology
Bournemouth University
atabas@bournemouth.ac.uk

Emili Balaguer-Ballester
Faculty of Science and Technology
Bournemouth University, UK
Berstein Center for Comp. Neuroscience
eb-ballester@bournemouth.ac.uk

Laura Igual
Dept. of Applied Mathematics & Analysis
Universitat de Barcelona
Computer Vision Center of Barcelona
ligual@ub.edu

***Abstract*—Resting-State fMRI (RS-fMRI) is a brain imaging technique useful for exploring functional connectivity. A major point of interest in RS-fMRI analysis is to isolate connectivity patterns characterising disorders such as for instance ADHD. Such characterisation is usually performed in two steps: first, all connectivity patterns in the data are extracted by means of Independent Component Analysis (ICA); second, standard statistical tests are performed over the extracted patterns to find differences between control and clinical groups. In this work we introduce a novel, single-step, approach for this problem termed Spatial Discriminant ICA. The algorithm can efficiently isolate networks of functional connectivity characterising a clinical group by combining ICA and a new variant of the Fisher's Linear Discriminant also introduced in this work. As the characterisation is carried out in a single step, it potentially provides for a richer characterisation of inter-class differences. The algorithm is tested using synthetic and real fMRI data, showing promising results in both experiments.**

I. Introduction

Resting-State fMRI (RS-fMRI) [1] is a technique exploring functional connectivity in the brain. In contrast to classical fMRI paradigms, which are task-driven, resting-state studies are based on BOLD responses associated with background brain activity in subjects at rest [1]. Recently, RS-fMRI analyses successfully enabled to correlate activation patterns in BOLD responses with disorders like schizophrenia [2] or ADHD [3].

Resting-State responses are often represented as linear combinations of fixed spatial patterns in the 3-dimensional brain with time-dependent coefficients [4]. These spatial patterns are termed resting-state networks (RSN) [1] and summarise functional connectivity between brain areas showing simultaneous activity during rest. Connectivity patterns are not localised in any particular region of the brain, and thus analyses are commonly performed over the whole brain. Thus, samples in the analysis present a large dimensionality ($\sim 10^5$ dimensions using voxels of $4\,\mathrm{mm}^3$).

RSNs are known to be sparse on the MRI space showing super-Gaussian shapes [4]. Thus, Independent Component Analysis (ICA) can be used for extracting the RSNs from the BOLD signal [5]. The networks can be studied on its own or can be rather used to extract the corresponding linear time-dependent coefficients (or *time-courses*). The time-courses are computed using, for instance, a General Linear Model [6] fitting the linear coefficients of the RSNs that optimally reconstruct the original fMRI signal [7].

978-1-4799-4149-0/14/$31.00

In fMRI the spatial patterns (rather than their associated time-courses) are considered independent [8]. Consistently, Spatial-ICA, an architecture of ICA considering solutions with the same dimension than the samples, is used instead of the standard ICA architecture [8]. The solutions of Spatial-ICA are linear combinations of the fMRI volumes in the signal, and thus preserve the spatial information (i.e. localisation of each voxel within the brain) of the data. Thus, solutions can be clinically interpreted.

To characterise the differences between two populations of subjects, some inference procedure has to be carried over the RSNs, that are actually common to all the subjects. Several approaches have been proposed in the literature in this direction. Most of them (e.g. [3], [4]) consist on first extracting the RSNs and then performing inference over time-courses associated to each network and group of subjects. Other approaches perform the inference without explicitly computing the time-courses (e.g. [9], [10]) but they still require to extract the RSNs as a first step previous to further statistical inference.

The goal of this work is to design an efficient algorithm for the identification of spatial patterns characterising differences between two groups (typically a clinical population and a control group) in RS-fMRI. Remarkably, and in contrast with previous approaches, the RSNs are not necessary to perform the inference, potentially presenting two main advantages in comparison with the current state of the art methods. First, our approach avoids unnecessary computation steps by directly extracting important (i.e. discriminating) patterns in the data. Second, as the solutions are not constrained to be RSNs or variations of RSNs, they are potentially more accurate in representing differences between the two classes.

Towards this goal, we developed a supervised version of ICA, termed Spatial Discriminant ICA, maximising both independence and discriminating power of the solutions. A similar approach named Discriminant-ICA has been proposed in [11] in the context of feature extraction combining both, ICA and the Fisher's (Linear) Discriminant Analysis (FDA) objective functions. Discriminant-ICA successfully extracts a (small) set of mutually independent features from a dataset maximising the linear separation of the classes in the data. Nevertheless, such approach is not applicable in fMRI, since Spatial-ICA and FDA architectures are not compatible with each other. Indeed,

Spatial-ICA solutions are linear combinations of the samples (i.e. volumes) [5] while FDA solutions are linear combinations of the original features in the data (i.e. voxels) [12].

In this work, we first introduce Spatial-FLD, a variation of the FLD compatible with the architecture of Spatial-ICA. Spatial-FLD evaluates the *discriminant power* of a candidate component by measuring the linear separability, as given by the FLD, of the points of the associated time-courses belonging to each class. Then, we combine Spatial-FLD and Spatial-ICA to formulate Spatial Discriminant ICA.

II. The Method

A. Spatial-ICA

Consider the set of fMRI volumes $\mathbf{x} \in \mathcal{X}_j$ representing the different time-points of a recording of the subject j. The aggregation of all volumes in a multi-subject experiment is denoted by $\mathcal{X} \equiv \cup_j \mathcal{X}_j$. A standard procedure in Spatial-ICA is to first reduce the number of elements on $\mathcal{X}$ using Principal Component Analysis (PCA) in order to decrease the computational complexity of the algorithm.

PCA reduces $\mathcal{X}$ to a smaller representative set $\mathcal{Z}$, preserving the spatial information of the original samples $\mathbf{x} \in \mathcal{X}$ (i.e. the voxels if the instances in $\mathcal{Z}$ can still be directly mapped into brain areas). This smaller set is used to span the search space for the Independent Components $\boldsymbol{\xi}$:

$$\boldsymbol{\xi} = \sum_j w^j \mathbf{z}_j, \qquad \mathbf{z}_j \in \mathcal{Z} \tag{1}$$

where w^j are the coefficients of $\boldsymbol{\xi}$ in the basis defined by $\mathcal{Z}$. The vectors $\mathbf{w}$ are often used to characterise the ICs $\boldsymbol{\xi}$.

A common approach [8] to find the components is to maximise the negentropy of the voxel distribution in the target IC $\boldsymbol{\xi}$. The procedure is repeated several times to find several components, enforcing orthogonality between the weight vectors $\mathbf{w}$ that characterise each solution in order to guarantee that each component is extracted only once [5]. If the ICs are non-Gaussian, this procedure guarantees that the extracted components are maximally independent to each other [5].

The negentropy $J(\boldsymbol{\xi})$ of the candidates is estimated by comparing the approximate entropy of the distribution of $\boldsymbol{\xi}$ with the entropy associated with a Gaussian distribution with same mean and variance than $\boldsymbol{\xi}$. Thus, it is convenient to keep a constant variance in the search space of the algorithm. As all the samples in $\mathcal{Z}$ are whitened after the PCA, this constrain is imposed by enforcing $\|\mathbf{w}\|^2 = 1$ during the optimisation process. Further details on Spatial-ICA can be found in [5].

B. Spatial-FLD

In this work we introduce Spatial-FLD, a simple variant of the FLD adapted to be compatible with the architecture of Spatial-ICA. Spatial-FLD can be presented as a two stages measure. In the first stage, a new representation of the fMRI data is constructed by computing an estimation of the time-courses of the component $\boldsymbol{\xi}$. In the second stage, such representation is fed into a classical FLD to evaluate the linear separation of the classes under the considered component.

In the literature, the time-courses are computed after extracting all the RSNs. In the current approach, we estimate them for a single candidate component by assuming that the whole set of RSNs is orthogonal. This orthogonality assumption can be heuristically justified by inspecting the RSNs extracted by Biswal [1], where can be observed that the expected cross scalar product between any two of the RSNs is less than 0.05 times the norm of the components.

Under the hypothesis that the RSNs are orthogonal, the direct projection of one component $\boldsymbol{\xi}$ over the fMRI signal, assumed to be constructed as a linear combination of RSNs, directly yields the linear coefficients of such component in the data. Thus, if the time-course associated to the expression of a given RSN in the data maximises the linear separation between the classes, the Spatial-FLD will show a maximum in the linear separation of the data when the considered component equals such RSN.

In more formal terms, the data points considered by the FLD for the subject i are expressed as the projections $p_j(\mathbf{w})$ of the volumes $\mathbf{x}_j \in \mathcal{X}_i$ onto the considered component $\boldsymbol{\xi}$. Thus, the FLD considers as samples each of the linear coefficients associated with each of the volumes in each of the recordings.

$$p_j(\mathbf{w}) \equiv \langle \mathbf{x}_j, \xi \rangle, \qquad \mathbf{x}_j \in \mathcal{X}_i, \qquad \forall \mathcal{X}_i \in \mathcal{X} \tag{2}$$

The labels of each sample $\mathbf{x} \in \mathcal{X}_i$ are chosen according to the label of its subject i (i.e. all the time-points of the volumes of a *control* subject are considered as *control* samples).

By using Equation (1), we can rewrite the definition in Equation (2) in a more convenient way:

$$p_i(\mathbf{w}) = \sum_j w^j \langle \mathbf{x}_i, \mathbf{z}_j \rangle = \sum_j w^j Q_{ij}$$

where, in the last step, we have defined the Q-matrix $Q_{ij} = \langle \mathbf{x}_i, \mathbf{z}_j \rangle$. The Q-matrix can be precomputed to avoid the repetition of the computationally expensive projections $\langle \mathbf{x}_i, \mathbf{z}_j \rangle$ (note that each of the instances $\mathbf{z}$ and $\mathbf{x}$ have the dimension of an fMRI volume).

The classical FLD formulation is optimal if projections of samples of different classes onto the discriminant subspace are normally distributed [6]. The normality assumption is reasonable over the time-course approximations p as they are computed as a projection onto a large number of voxels that show a certain degree of statistical independence on an area by area basis. Thus, the samples p approximately verify the central-limit theorem [6]. However, we cannot expect the differences between classes to be exclusively reflected in the mean of these time-courses, but also in the variance of such distributions (i.e. in a quadratic term) [4]. To take this factor into account, instead of considering scalar samples p in the discriminant we consider vector samples $\boldsymbol{\eta}_i$ containing both, linear and quadratic terms of the projections:

$$\boldsymbol{\eta}_i(\mathbf{w}) \equiv \begin{pmatrix} p_i \\ p_i^2 \end{pmatrix} \tag{3}$$

Following the Rayleigh ration used in a classical two-class FLD (the extension to a multivariate discriminant is straightforward) we simply define the Spatial-FLD as:

$$\Phi(\mathbf{w}) \equiv \frac{|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2|^2}{\sigma_1^2 + \sigma_2^2} \tag{4}$$

where $\boldsymbol{\mu}_c$ is the mean of the vectors $\boldsymbol{\eta}$ on class (c) and σ_c represents the intraclass variance.

$$\boldsymbol{\mu}_c \equiv \frac{1}{M_c} \sum_{i \in \mathcal{D}_c} \boldsymbol{\eta}_i, \qquad \sigma_c \equiv \frac{1}{M_c} \sum_{i \in \mathcal{D}_c} \|\boldsymbol{\mu}_c - \boldsymbol{\eta}_i\|^2$$

In this last equation, $\mathcal{D}_c$ represents the set of samples and M_c the number of volumes belonging to class c.

Note that the only *free* parameters to be optimised in the Spatial-FLD are the components of $\mathbf{w}$. This restricts the search space of the Spatial-FLD to the subspace spanned by $\mathcal{Z}$, which is computed using unsupervised methods. This alleviates any potential overfitting issues, as PCA only preserves few variance-maximising components of the space, therefore filtering out noise such as inter-subject variability.

C. Spatial Discriminant ICA

Lastly, we formulate Spatial Discriminant ICA as a joint convex optimisation of both the ICA objective function $J(y)$ (see [5]) and the Spatial-FLD:

$$\mathcal{J}(\mathbf{w}) = \kappa\, \Phi(\mathbf{w}) + (1 - \kappa)\, J(\xi(\mathbf{w})) \tag{5}$$

where $0 \leq \kappa \leq 1$ is a weighting factor modulating the relative importance of each term. This dual process contributes to alleviate possible overfitting issues of the Spatial-FLD by adding an unsupervised component to the objective function.

The algorithm operates using standard gradient ascent with the following restrictions over the weight vector $\mathbf{w}$: it is forced to be unitary (i.e. $\mathbf{w} \leftarrow \mathbf{w}/\|\mathbf{w}\|^2$) and orthogonal to the weight vectors characterising any previously extracted component. This last condition is imposed using Gram-Schmidt orthogonalisation [5].

Code for Spatial Discriminant ICA, including all libraries necessary for the preprocessing and computation of the Q-matrix, is freely available in https://www.github.com/qtabs/sdica/ under the MIT license.

III. Experiments and Results

A. Experimental sets

1) Synthetic data: We tested the algorithm using a two-class synthetic dataset designed to satisfy our assumptions (i.e. BOLD signal constructed as a linear combination of quasi-orthogonal RSNs and Gaussian distribution of the time courses). The objective was to empirically verify that the algorithm was able to isolate hidden discriminant RSNs in the data. We used the RSN library by Biswal et al. [1], containing 20 common RSNs, to generate the dataset. The RSN sampled differently for each of the two classes (or *discriminant-RSN*) was chosen randomly in each run. Parameters for the time-course Gaussian distributions of all RSNs were chosen to be similar for all subjects except for the one associated with the discriminant-RSN, showing different variances for subjects belonging to different classes.

The resulting dataset was further modified by adding white noise of different moderate intensities (with ratios signal to noise between 0.001 and 0.1) and an extreme-noise intensity (with a ratio of 1) in order to evaluate the robustness of the algorithm. Additional tests were performed with two and three discriminant-RSNs in order to evaluate the capacity of the algorithm for isolating several (rather than only one) components. Additionally, we re-scaled the datasets in different factors to evaluate the efficiency of the algorithm. Specifically, we run experiments considering datasets with $10^3 - 10^5$ dimensions and $100 - 2000$ volumes.

2) Real data: We further tested the algorithm using real fMRI data with a balanced version of the Pekin University subset of the ADHD200[1]. In order to reduce the problem to a binary classification scenario, we collapsed the three classes of ADHD disorders (combined, inattentive and hyperactive) into a single class. The classes in the dataset were then balanced by removing 38 randomly chosen control subjects. This yields to a final dataset of 78 vs 78 subjects. Then we applied a Spatial-PCA independently over the recordings of each subject in order to reduce the size of the data and to filter out noise. This reduced the data of each subject to 61 volumes, resulting in a final 156×61 volumes dataset $\mathcal{X}$.

To quantitatively assess the information contained in the solutions extracted by the algorithm we used the components to perform a classification task over the two classes in the data. Following that purpose, we used our algorithm to extract 20 components. The projectors associated with those components $\boldsymbol{\eta}$s (see Equation (3)) were then used as inputs of a linear SVM. Note that the purpose of this approach is simply to devise a complementary test of the reliability of the extracted components. Supervised classification is not a primary objective of our algorithm. Thus, we selected a simple yet robust linear classifier.

Performance was computed using a 10-Fold cross-validation strategy over the whole process (i.e. the extraction of the components and the training of the SVM) across the 156×61 volumes. Predictions over the subjects can be obtained using a majority-vote strategy over the predictions of the volumes. The aim of this study, however, is to assess the class-specific information contained in the components extracted with our algorithm and therefore subject classification was not attempted.

The experimentation was carried out using different configurations to measure the impact of the different parameters of the algorithm in the information contained in the solutions. Specifically, we vary the relative weight of the two parts of the algorithm κ and the size of the reduced set $\mathcal{Z}$ (i.e. the number of degrees of freedom during the search). To compare

[1]The freely available dataset, license and documentation can be found in http://fcon_1000.projects.nitrc.org/indi/adhd200/#beijing. The dataset was fully preprocessed by The Neuro Bureau, The ADHD-200 consortium, and the Virginia Tech's ARC. Details in the preprocessed are published in http://www.nitrc.org/plugins/mwiki/index.php/neurobureau:AthenaPipeline

our method with other ICA-driven approaches we additionally performed a plain Spatial-ICA running the algorithm with $\kappa = 0$ (i.e. removing all contributions of the Spatial-FLD to the objective function).

B. Results

1) Synthetic data: Synthetic data yield very satisfactory results. The algorithm was able to correctly find the discriminant RSNs for all noise conditions with one and two discriminant RSNs. With regards to the three discriminant RSNs experiment, the algorithm was able to find the first two networks for all levels of noise, but it failed on revealing the third one in the dataset with the extreme noise level. Correlation coefficients between the extracted and the original RSNs were > 0.95 for the moderate noise intensities and oscillated between 0.6 and 0.8 for the extreme noise condition.

Moreover, the algorithm found the solutions efficiently. Experiments performed with different-sized datasets showed that the time required for each iteration of the gradient ascent scaled linearly with both, the dimension and the number of instances of the dataset.

2) Real data: Results for the real data are summarised in Figure 1. Black triangles denote the accuracy obtained using the classical Spatial-ICA based approach (i.e. $\kappa = 0$) for feature extraction with a linear SVM classifier. Those accuracies are used as a comparison baseline here. The performance in this baseline is not meant to reflect the accuracy that can be achieved using state of the art methods, but the information contained in the networks extracted using Spatial-ICA. Further process applied after the network extraction often used in state of the art works on RS-fMRI classification [4] can notably improve these results. However, we stress that such procedures can also be applied on the top of our solutions. Nevertheless, as indicated before, the target of this experiment is not to test the potential of the algorithm for classification tasks, but to quantitatively assess the value of the information contained in the solutions provided by the algorithm.

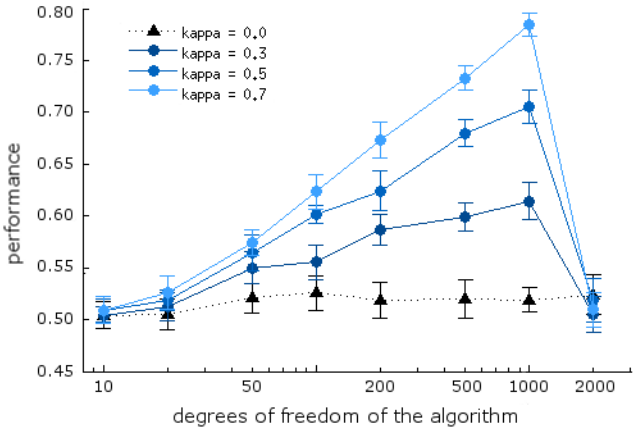

Fig. 1. 10-fold-cross-validation performances for different values of κ and degrees of freedom of the algorithm. Error bars are MSE.

The blue shapes correspond to the results of the Spatial Discriminant ICA for different parameter values, showing improved performance with respect to the baseline for runs with 50 (t-test, $p < 0.02$) and $100 - 1000$ (t-test, $p < 0.00001$) degrees of freedom (dof, i.e. the number of instances in $\mathcal{Z}$, see 1) and for all $\kappa > 0.2$. With 10 dof, the solutions found by the algorithm are equivalent representations of $\mathcal{Z}$, and thus solutions for any κ contain the same information. Solutions found for the largest subspace (2000 dof) seem to overfit the data providing poorer results.

IV. Conclusions

In this work we have introduced a novel method termed Spatial Discriminant ICA. This technique combines Independent Component Analysis and a spatial variant of the Fisher's Linear Discriminant in order to identify new spatial patterns characterising differences between two groups of recordings (typically, a set of patients and a control group) without previously extracting the RSNs characterising the dataset. To our knowledge, this has not been addressed by previous approaches.

Experiments on synthetic and real data show that the algorithm potentially presents a strong capacity to isolate discriminant spatial patterns hidden in RS-fMRI data. Moreover, we found that the algorithm scales linearly in time with the dimension and the number of volumes of the data.

In future work, we will extend these results analysing the extracted spatial patterns found by the algorithm from a clinical point of view, comparing them to with the classical RSNs and the standard solutions extracted with Spatial-ICA.

References

[1] B. B. Biswal, M. Mennes, X.-N. Zuo, S. Gohel, C. Kelly, S. M. Smith, C. F. Beckmann, J. S. Adelstein, R. L. Buckner, S. Colcombe, and Others, "Toward discovery science of human brain function," *Proceedings of the National Academy of Sciences*, vol. 107, no. 10, pp. 4734–4739, 2010.

[2] H. Liu, Z. Liu, M. Liang, Y. Hao, L. Tan, F. Kuang, Y. Yi, L. Xu, and T. Jiang, "Decreased regional homogeneity in schizophrenia: a resting state functional magnetic resonance imaging study," *Neuroreport*, vol. 17, no. 1, pp. 19–22, 2006.

[3] Z. Yu-Feng, H. Yong, Z. Chao-Zhe, C. Qing-Jiu, S. Man-Qiu, L. Meng, T. Li-Xia, J. Tian-Zi, and W. Yu-Feng, "Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI," *Brain and Development*, vol. 29, no. 2, pp. 83–91, 2007.

[4] V. D. Calhoun, J. Liu, and T. Adali, "Understanding pitch perception as a hierarchical process with top-down modulation." *PLoS computational biology*, vol. 5, no. 3, p. e1000301, Mar. 2009.

[5] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*, ser. Adaptive and Learning Systems for Signal Processing, Communications and Control Series. Wiley, 2004.

[6] C. M. Bishop *et al.*, *Pattern recognition and machine learning*. Springer New York, 2006, vol. 1.

[7] K. J. Friston, J. Ashburner, S. J. Kiebel, T. E. Nichols, and W. D. Penny, Eds., *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press, 2007.

[8] J. V. Stone, *Independent component analysis: a tutorial introduction*. The MIT Press, Cambridge, Massachusetts, 2004.

[9] Y. Fan, Y. Liu, H. Wu, Y. Hao, H. Liu, Z. Liu, and T. Jiang, "Discriminant analysis of functional connectivity patterns on Grassmann manifold," *Neuroimage*, vol. 56, no. 4, pp. 2058–2067, 2011.

[10] Y. Du and Y. Fan, "Group information guided ICA for fMRI data analysis." *NeuroImage*, vol. 69, pp. 157–197, 2013.

[11] C. S. Dhir and S.-Y. Lee, "Discriminant Independent Component Analysis," *Trans. Neur. Netw.*, vol. 22, no. 6, pp. 845–857, Jun. 2011.

[12] F. Pereira, T. Mitchell, and M. Botvinick, "Machine learning classifiers and fMRI: a tutorial overview," *Neuroimage*, vol. 45, no. 1, pp. S199—S209, 2009.