BRAIN DECODING OF FMRI CONNECTIVITY GRAPHS USING DECISION TREE ENSEMBLES

Jonas Richiardi[1,2]*, Hamdi Eryilmaz*[3]*, Sophie Schwartz*[3]*, Patrik Vuilleumier*[3]*, Dimitri Van De Ville*[1,2]

[1]Medical Image Processing Lab, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne

[2]Medical Image Processing Lab, Université de Genève, CH-1211 Genève

[3]Laboratory of Neurology and Imaging of Cognition, Université de Genève, CH-1211 Genève

ABSTRACT

Functional connectivity analysis of fMRI data can reveal synchronized activity between anatomically distinct brain regions. Here, we exploit the characteristic connectivity graphs of task and resting epochs to perform classification between these conditions. Our approach is based on ensembles of decision trees, which combine powerful discriminative ability with interpretability of results. This makes it possible to extract discriminative graphs that represent a subset of the connections that distinguish best between the experimental conditions. Our experimental results also show that the method can be applied for group-level brain decoding.

Index Terms— fMRI, brain decoding, functional connectivity, graphs, decision tree

1. INTRODUCTION

Functional magnetic resonance imaging (fMRI) has opened an intriguing window on brain function. Traditionally, univariate statistical analysis is applied to assess the activity of each voxel induced by the stimulation paradigm. Recently, multivariate pattern analysis—commonly termed "brain decoding"—has allowed to exploit more subtle interactions between voxels' intensities [1, 2]. Basically, a classifier is learned (at the individual subject level) for selected voxels and then tested on the unseen data.

There is also an increasing interest of the neuroscience community for resting state; i.e., brain activity when the subject is relaxing and letting his thoughts wander around. Spatially coherent spontaneous fluctuations of the blood-oxygen-level dependent (BOLD) signal allow the identification of large-scale cortical networks, among which the default-mode network [3]. The study of resting-state networks mostly relies on seed-voxel correlation [4, 5] and source separation such as independent component analysis [6, 7]. More recently, Achard *et al.* studied the network properties of the undirected graph obtained from temporal correlation matrices between different brain regions [8].

Here, we bring together brain decoding techniques and graph representations based on functional connectivity measures. In particular, we compare functional connectivity from multiple subjects and between different brain states using pattern recognition. We adapt decision-tree classification to unveal a discriminative network. This approach also allows to give relevant visual feedback to the neuroscientist about which connections are most discriminative. As a proof of concept, we consider fMRI data from a block-based paradigm with long resting periods. We show that the graph representations of stimulation versus rest reflect the distinct nature of both states at the group level.

The paper is organized as follows. In Sect. 2, we describe the complete data processing pipeline that leads to the classification of connectivity graphs. In Sect. 3, we detail the experimental setup and discuss the results.

2. CLASSIFYING CONNECTIVITY GRAPHS

2.1. Timecourse preprocessing

After realignment of the functional volumes using SPM5[1], we use the IBASPM toolbox [9, 10] to build an individual brain atlas (based on the structural MRI) that contained $M = 90$ anatomical regions. We then obtain spatially-averaged timecourses corresponding to these regions in the functional space. For N repetitions (which can be intra- or inter-subject, in our case N is the number of subjects) and C conditions, we obtain the matrix

$$\mathbf{X} : M \times T \times N \times C,$$

that contains $M \cdot N \cdot C$ timecourses of length T. We denote the submatrix $\mathbf{X}_{n,c} : M \times T$ for the M timecourses of subject n and condition c.

The timecourses are then decomposed using the (redundant) discrete wavelet transform (DWT) along the

This work was supported in part by the Swiss National Science Foundation (grant PP00P2-123438), in part by the Brain & Behavior Laboratory (BBL), and in part by the Center for Biomedical Imaging (CIBM) of the Geneva and Lausanne Universities, EPFL, and the Leenaards and Louis-Jeantet foundations.

[1]Available at `http://www.fil.ion.ucl.ac.uk/spm/`.

temporal dimension. This results into J matrices $\mathbf{X}^{(i)}$, $i = 1, \ldots, J$ that reflect the regional brain activities at different temporal scales. We used orthogonal Battle-Lemarié wavelets (fourth order).

2.2. Functional connectivity graphs

We use pairwise Pearson correlation coefficients to form the correlation matrix $\mathbf{R}^{(i)}_{n,c} = \mathbb{E}[\mathbf{X}^{(i)}_{n,c}\mathbf{X}^{(i)\,T}_{n,c}] : M \times M$, where the temporal detrending is provided by the vanishing moments of the wavelet decomposition.

We now consider the brain regions as a set of vertices V and the correlation coefficients as weights on the set of edges E, leading to an undirected complete weighted graph $\mathcal{G} = (V, E)$. The weighted graph adjacency matrix $\mathbf{A}^{(i)}_{n,c}$ can be defined as $\mathbf{A}^{(i)}_{n,c} = \mathbf{R}^{(i)}_{n,c} - \mathbf{I}$.

The intra-condition variability of these graphs is usually quite high. Therefore, we follow the statistical thresholding procedure proposed by Achard *et al.* [8]; i.e., we construct a mask based on the false discovery rate (FDR) procedure at significance level α of the adjacency matrices of multiple subjects. Since our goal is to perform classification, it is important to exclude the graph to classify for the construction of this mask. We propose to intersect the masks that preserve edges significant for each condition. This procedure leads to thresholded adjacency matrices $\mathbf{A}'^{(i)}_{n,c}$ where significant edges have been selected.

It is now of interest to be able to compare the graphs quantitatively. To this end, we propose to adopt a classification framework, where conditions are considered as classes.

2.3. Embedding functional connectivity graphs in vector space

In representing functional connectivity graphs as points in a vector space, one important property is to preserve the "similarity": graphs that are similar should correspond to close points in vector space.

We use a simple embedding procedure: since $\mathbf{A}'^{(i)}_{n,c}$ is symmetric and its main diagonal is $\mathbf{0} : M \times 1$, it is fully characterised by its upper triangular part above the main diagonal. For each graph, we thus generate a feature vector $\mathbf{F} : \binom{M}{2} \times 1$ from the edge weights of all the edges in the upper triangular part of $\mathbf{A}'^{(i)}_{n,c}$. All edges that were statistically thresholded because of their insignificance (see Section 2.2) are mapped to 0 in feature space. It would also be possible to omit these coordinates in the embedding.

Thus, the statistically motivated significance testing procedure on graph edges corresponds to a FDR-controlled feature selection procedure focused on minimising intra-class variance.

2.4. Classification

Decision trees are discriminative classifiers performing recursive partitioning of a feature space to yield a potentially non-linear decision boundary. In particular, the C4.5 algorithm [11] and its variants seek to find cutpoints in continuous variables f that minimize the conditional entropy on class labels $\mathcal{C} = \{1, \ldots, C\}$ attached to points in the corresponding subdomains of the discretised variable f'. The choice of the feature and its cutpoint are critical. More precisely, we can express the entropy of the dataset partitioned by feature f' as:

$$H(\mathcal{C}|f') \triangleq -\sum_{j=1}^{2} P_j \sum_{c=1}^{C} P_{j,c} \log_2 P_{j,c}, \qquad (1)$$

where P_j is the relative frequency of points in the subset that have value j for feature f', and $P_{j,c}$ is the relative frequency of points that belong to class c and have value j for feature f'.

The goal of decision tree growing is then to minimize (1), which is equivalent to maximizing the mutual information between $I(\mathcal{C};f')$, and involves recursively selecting features and computing the result of applying different cutpoints to them.

In our case, the features correspond to dimensions in the feature vector, which in turn correspond to edge weights in the functional connectivity graph. Recalling that several dimensions have been set to zero by the FDR thresholding procedure, it is clear that the decision tree will never select these dimensions because for a constant random variable K we have $H(\mathcal{C}|K) = H(\mathcal{C})$, and thus we have no decrease in conditional entropy.

It is well known that decision tree classifiers can be unstable, in the sense that the decision boundary can be sensitive to even small changes in the training set. By using bagging [12], which creates multiple "bootstrapped" sets of data by repeatedly sampling with replacement from the training set, we can train diverse classifiers on each of the bootstrap samples, and we can average their prediction by voting. This reduces classifier variance and in a wide variety of classification problems leads to improved accuracy.

In order to avoid overfitting, we never train and test the classifier on data of the same subject. We derive the bootstrap samples and grow the decision trees using a leave-one-subject-out cross-validation procedure, by which the data of $N-1$ subjects are used for training, and the data of 1 subject is used for testing. The training and testing partition is then rotated N times. This well-motivated evaluation procedure implies that inter-subject decoding takes place in our experimental setting.

2.5. Discriminative graph computation

Based on the N sets of ensembles of trained decision tre classifiers, we propose to compute statistics about th connections that were most often used in discriminatio across cross-validation folds and within bootstrappe samples. Every time a unique connection (between di ferent brain regions) appears in one of the ensemble decision trees within a fold, the corresponding count is incremented.

This procedure allows us to build a discriminative graph, which represents functional connectivity that is markedly different between the conditions of interest, and is useful for analysis. The vertices of the discriminative graph are the brain regions linked by the connections used by the decision trees. The weights of the edges connecting vertices can be set to be proportional to the number of times the connection has been used in decision tree growing.

3. EXPERIMENTAL RESULTS

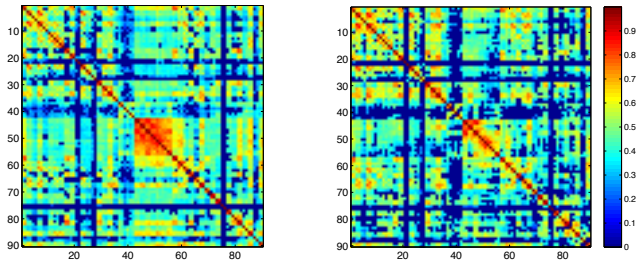
3.1. Data acquisition and preprocessing

The experimental design is a blocked design with alternating epochs of movie excerpts (50 seconds) and resting periods (90 seconds). During rest, subjects are instructed to close their eyes, relax, let their mind wander and avoid thinking of something in particular.

The subjects (1 male, 9 female) are aged between 18 and 36 years old, without history of neurological disorders. Scanning was performed on a Siemens 3T Tim Trio. Functional imaging data were acquired in two sessions using gradient-echo echo-planar imaging (TR/TE/FA = 1.1 s/27 ms/90°, matrix = 64×64, voxel size = 3.75×3.75×4.2 mm^3, 21 contiguous transverse slices, 1.05mm gap, 2598 volumes). Structural imaging data was acquired using a three-dimensional MPRAGE sequence (192 slices, TR/TE/FA = 1.9 s/2.32ms/9°, matrix = 256 × 256, voxel size = 0.90 × 0.90 × 0.90 mm^3).

Because of the experimental paradigm, task and non-task blocks alternate. Within each subject, the blocks containing the 90 time-courses corresponding to the same condition (movie, resting) are concatenated after being linearly detrended to compensate for potential drift, resulting in a single matrix of time courses for the resting condition, and one for the movie condition.

We show in Fig. 1 two averaged thresholded adjacency matrices, one for the resting condition and one for the movies condition.



(a) Resting condition (b) Movies condition

Fig. 1. Adjacency matrices for 90-regions connectivity graphs in subband 3, thresholded at $\alpha = 0.1\%$. Dark blue entries are not significant.

3.2. Classification and discriminative graph

Each subject provides one connectivity graph of the resting class ($c = 1$) and one connectivity graph of the movie class ($c = 2$). Classifier training and testing is performed using the procedure outlined in Section 2.4, with 51 bootstrapped samples each 1.5 times the size of the base training set. Note that the intersection significance masks (see Section 2.2) are computed in-fold using only the training data. They are applied to both the training and testing adjacency matrices.

Classification results are given in Table 1.

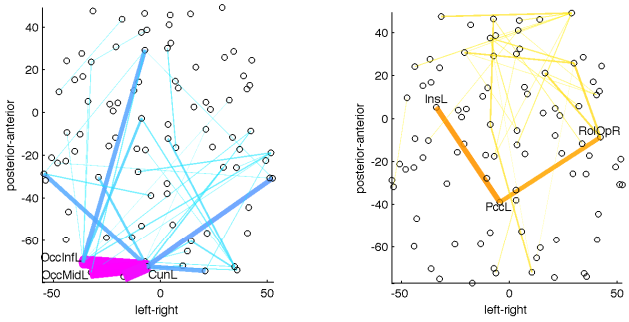
Subband	$\alpha = 5\%$	$\alpha = 0.1\%$	$\alpha = 0.01\%$
1 (0.23-0.45 Hz)	55%	—	—
2 (0.11-0.23 Hz)	65%	75%	—
3 (0.06-0.11 Hz)	90%	90%	90%
4 (0.03-0.06 Hz)	70%	70%	75%

Table 1. Leave-one-subject-out accuracy. Results marked — are missing since no significant connections.

The number of connections passing the significance threshold is significantly higher for low frequency subbands. This reflects that resting-state networks are found to be quite consistent across subjects [13, 5], which in turn will yield relatively low standard deviation on inter-subject graph edge weights.

Across frequency bands, it can be seen that feature selection via FDR-thresholding of connectivity graphs generally brings slight improvements in accuracy. While the connections selected in growing decision trees are very similar across significance thresholds, the slightly improved performance can be attributed to the fact that the significance mask is also applied to the test samples, thereby reducing intra-class variability and contributing to counteract the overfitting tendency of decision trees.

A discriminative graph $\mathcal{H}$ can be extracted for each subband from the distribution of classifier parameters as



(a) $\mathcal{H}_+$ (rest > movies) (b) $\mathcal{H}_-$ (movies > rest)

Fig. 2. Axial view of subgraphs extracted from the discriminative graph (fourth subband). Darker colour and thicker line means more discriminative connection. Circles indicate atlas region centroids. Weighting is the same on both subgraphs.

explained in Section 2.5. To ease visualisation and interpretation of the fourth subband, we split the associated $\mathcal{H}$ in two discriminative subgraphs; i.e., from the sign of the contrast "rest > movies", obtained from the N adjacency matrices, we derive the subgraphs $\mathcal{H}_+$ and $\mathcal{H}_-$ for positive and negative values, respectively. It should be noted that both subgraphs carry discriminative power to distinguish rest versus movies. However, $\mathcal{H}_+$ reflects those connections that are stronger in rest than movies condition, and vice versa. We thus have $\mathcal{H} = \mathcal{H}_+ \cup \mathcal{H}_-$.

The axial views of the subgraphs in anatomical space are shown in Figure 2. In (a), we observe a strong connectivity pattern for the resting condition in the occipital cortex with notable bilateral coactivation, also with the auditory cortices. This type of coherent BOLD fluctuations are reminiscent of the low-frequency subband [4, 5]. These connections are also the most discriminative. In (b), the movies condition has important connections between limbic and central areas, including prefrontal areas. They are, however, less discriminative. These preliminary results demonstrate the feasibility of the proposed methodology.

4. CONCLUSION

The proposed approach is a promising avenue for analysis of functional networks in fMRI data, as shown by the high classification rate attained in an inter-subject brain decoding problem. The choice of decision trees and the method proposed to extract discriminative connections combines high-end classification performance of ensemble methods with interpretability of results. The approach might also be useful in investigating changes in network activity related to pathologies. In the case of Alzheimer's disease, for example, changes in the default-mode network have been reported [3].

In future work, we will evaluate the use of different classifiers as well as the influence of the dependency measure that is used to characterize the connectivity strength.

5. REFERENCES

[1] J. V. Haxby, I. M. Gobbini, M. L. Furey, A. Ishai, J. L. Schouten, and P. Pietrini, "Distributed and overlapping representations of faces and objects in ventral temporal cortex," *Science*, vol. 293, no. 5539, pp. 2425–2430, September 2001.

[2] J.-D. Haynes and G. Rees, "Decoding mental states from brain activity in humans, 7(7), 523-534.," *Nature Reviews Neuroscience*, vol. 7, no. 7, pp. 523–534, 2006.

[3] R. L. Buckner, J. R. Andrews-Hanna, and D. L. Schacter, "The brain's default network: Anatomy, function and relevance to disease," *Ann. N. Y. Acad. Sci.*, vol. 1124, pp. 1–38, 2008.

[4] B. Biswal, F. Yetkin, V. Haughton, and J. Hyde, "Functional connectivity in the motor cortex of resting human brain using echo-planar MRI," *Magnetic Resonance in Medicine*, vol. 34, no. 4, pp. 537–541, 1995.

[5] M. J. Lowe, B. J. Mock, and J. A. Sorenson, "Functional connectivity in single and multislice echoplanar imaging using resting-state fluctuations„" *NeuroImage*, vol. 7, no. 2, pp. 119–132, Feb. 1998.

[6] M. McKeown, S. Makeig, G. Brown, T. Jung, S. Kindermann, A. Bell, and T. Sejnowski, "Analysis of fMRI data by blind separation into independent spatial components," *Human Brain Mapping*, vol. 6, pp. 160–188, 1998.

[7] V. D. Calhoun, T. Adali, G. D. Pearlson, P. C. M. van Zijl, and J. J. Pekar, "Independent component analysis of fMRI data in the complex domain," *Magnetic Resonance in Medicine*, vol. 48, no. 1, pp. 180–192, July 2002.

[8] S. Achard, R. Salvador, B. Whitcher, J. Suckling, and E. Bullmore, "A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs," *The Journal of Neuroscience*, vol. 26, no. 1, pp. 63–72, Jan. 2006.

[9] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *NeuroImage*, vol. 15, pp. 273–289, 2002.

[10] Y. Alemán-Gómez, L. Melie-García, and P. Valdés-Hernandez, "IBASPM: Toolbox for automatic parcellation of brain structures," in *12th Annual Meeting of the Organization for Human Brain Mapping*, Florence, Italy, June 2006, vol. 27 of *NeuroImage*.

[11] J. R. Quinlan, "Improved use of continuous attributes in c4.5," *Journal of Artificial Intelligence Research*, vol. 4, pp. 77–90, 1996.

[12] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, Aug. 1996.

[13] R. Buckner, J. Andrews-Hanna, and D. Schacter, "The brain's default network: Anatomy, function, and relevance to disease," *Annals of the New York Academy of Sciences*, vol. 1124, pp. 1–38, 2008.