Functional subtypes for prediction in schizophrenia

Christian Dansereau, Pierre Orban, Sebastian Urchs, Pierre Bellec

Abstract—In this paper we explore the classification of two clinical population namely a control population and subject with Schizophrenia. We want to explore the difficulty of the classification task and how it would apply in the context of functional imaging. During the project we have explored normalization procedure and the proper pipeline to identify the right model and parameters. We propose two new pipelines on top of the more standard ones. A multiscale bagging method to take in account the functional information at different level of abstraction and finally we explore if feature selection based on a margin maximization criteria would improve the performance of a classification model applied on neuro-imaging data.

Keywords—Schizophrenia, classification, multiscale, feature selection, SVM, margin optimization.

1 Introduction

Schizophrenia has been defined as a disconnection syndrome, which is characterized by aberrant functional brain connectivity. Schizophrenia is also a very heterogeneous clinical entity, with marked variations in both the symptoms experienced by patients and their response to treatments. This variability is most likely associated with heterogeneous neurobiological aberrations. We hypothesized that this neurobiological heterogeneity would be reflected in patterns of full brain functional connectivity (FC) in fMRI, and that specific subtypes of FC would be much more amenable than others to predict clinical symptoms.

1.1 Objectives

Our goals for this study were therefore 1) to identify subtypes of functional connectivity, in a sample mixing patients diagnosed with schizophrenia and healthy controls, 2) to show that subtypes can predict some easy cases, i.e. a subgroup of subjects that can be accurately diagnosed based on FC.

• *Mr. Dansereau is with the Department of Computer science, University of Montreal, Montreal, CA.*
E-mail: christiandansereau@gmail.com

Manuscript received March 17, 2016

1.2 Public code and data

The code used in this experiment is available on a GitHub repository at the following URL[1]. A IPython Notebook is also provided with all of the figure generation scripts. The functional imaging data is publicly available on the NITRC web site[2].

2 Method

2.1 Dataset

The dataset used in this paper is the COBRE (The Center for Biomedical Research Excellence) dataset part of the INDI initiative. It consist of raw anatomical and functional MR data from 72 patients with Schizophrenia and 75 healthy controls (ages ranging from 18 to 65 in each group). All subjects were screened and excluded if they had; history of neurological disorder, history of mental retardation, history of severe head trauma with more than 5 minutes loss of consciousness, history of substance abuse or dependence within the last 12 months. Diagnostic information was collected using the Structured Clinical Interview used for DSM Disorders (SCID).

A multi-echo MPRAGE (MEMPR) sequence was used with the following parameters:

1. https://github.com/cdansereau/vision_or/tree/master/code_project2

2. http://fcon_1000.projects.nitrc.org/indi/retro/cobre.html

TR/TE/TI = 2530/[1.64, 3.5, 5.36, 7.22, 9.08]/900 ms, flip angle = 7, FOV = 256x256 mm, Slab thickness = 176 mm, Matrix = 256x256x176, Voxel size = 1x1x1 mm, Number of echos = 5, Pixel bandwidth = 650 Hz, Total scan time = 6 min. With 5 echoes, the TR, TI and time to encode partitions for the MEMPR are similar to that of a conventional MPRAGE, resulting in similar GM/WM/CSF contrast.

Resting state data was collected with single-shot full k-space echo-planar imaging (EPI) with ramp sampling correction using the inter-comissural line (AC-PC) as a reference (TR: 2 s, TE: 29 ms, matrix size: 64x64, 32 slices, voxel size: 3x3x4 mm3).

We selected subjects matched for age, gender and motion for a total of 187 subjects (mean age 32.5, SD 10.3). 101 subjects were drawn from the COBRE study (resting-state fMRI, [1]), while 86 subjects were scanned in Montreal (emotional memory task fMRI, [2]). Individual fMRI scans were preprocessed using the NIAK pipeline [3]. We identified 12 functional brain networks using a Bootstrap Analysis of Stable Clusters (BASC [4]) on a separate dataset [5]. These networks were used as seeds to generate FC maps for each subject.

2.2 Preprocessing and feature extraction

Each fMRI dataset was corrected for slice timing; a rigid-body motion was then estimated for each time frame, both within and between runs, as well as between one fMRI run and the T1 scan for each subject [1]. The T1 scan was itself non-linearly co-registered to the Montreal Neurological Institute (MNI) ICBM152 stereotaxic symmetric template [2], using the CIVET pipeline [3]. The rigid-body, fMRI-to-T1 and T1-to-stereotaxic transformations were all combined to resample the fMRI in MNI space at a 3 mm isotropic resolution. To minimize artifacts due to excessive motion, all time frames showing a displacement greater than 0.5 mm were removed [4]. A minimum of 50 unscrubbed volumes per run was required for further analysis (13 subjects were rejected). The following nuisance covariates were regressed out from fMRI time series: slow time drifts (basis of discrete cosines with a 0.01 Hz highpass cut-off), average signals in conservative masks of the white matter and the lateral ventricles as well as the first principal components (accounting for 95% variance) of the six rigid-body motion parameters and their squares [5], [6]. The fMRI volumes were finally spatially smoothed with a 6 mm isotropic Gaussian blurring kernel. A more detailed description of the pipeline can be found on the NIAK website[3] and Github[4].

Functional connectivity matrices were obtained from 9 scales using a functional template based on an independent dataset of $\sim$ 200 subjects form the 1000 functional connectome project (Cambridge dataset). The 9 scales were obtain using the BASC pipeline [7] which is a unsupervised bootstrap clustering procedure for automatic detection of functional scales based on a stability criteria. resulting in 9 partitions of the brain in 7, 12, 20, 36, 64, 122, 197, 325, 444 networks see Figure **??** for an example of the the resulting connectome of one subject.

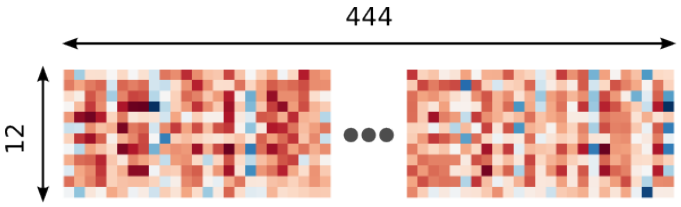

Fig. 1. Caption

2.3 Subtypes weights

Instead of performing our classification on a correlation matrix we introduce a new metric called subtype weights that is computed as follow: For each network, we identified 10 subgroups by grouping subjects with similar FC maps through hierarchical agglomerative clustering resulting in subtypes template. We then derived subtype weights for each individual using the spatial correlation of the individual FC map and each subtype template. This dimensionality reduction show the degree of association for an individual connectivity map with the corresponding subtypes templates.

3. http://niak.simexp-lab.org/pipe_preprocessing.html
4. https://github.com/SIMEXP/multisite

2.4 Subtypes associated with the pathology

To identify the subtype weights significantly associated with the pathology, we used a GLM including age, sex and frame displacement (FD) as covariates (corrected to have a zero mean across subjects). A p-value was generated for each subtype to quantify the probability that the difference between the control subjects and the patient with Schizophrenia was significant [8]. The number of false discovery was also controlled ($\alpha = 0.05$) using a BenjaminiHochberg false discovery rate (FDR) procedure [9].

2.5 Highly classifiable subjects (HCS)

In order to evaluate the predictive value of the subtypes we used the following framework. First a nested cross-validation was performed for accuracy estimation and parameters optimization using leave-one out cross validation in order to estimate the generalizability of the model.From that training set we regressed out the confounds was estimated on each connection pair from the training set.

Using the training set Regression confounds was estimated on each connection pair from the training set.

gender, age, average motion (frame displacement)

We used these subtype weights to train a logistic regression (LR) model [10] for the classification of our clinical labels (LRC). Finally, we trained a logistic regression algorithm to predict the hit (correct classification) and miss (incorrect classification) labels from the LRC (LRHE), using subtype weights as an input. Using the output of the LRHE, we were able to separate the discovery set into a subgroup of easy cases (where the logistic regression predicts a hit) and a complementary subgroup of difficult cases (where the logistic regression predicts a miss). he remaining subjects from the test set were used as a validation set to obtain unbiased accuracy scores for putative easy and difficult cases, as predicted by the logistic regression. Specifically, for each individual in the validation set, a subtype weight was estimated from subtype templates generated in the discovery set. These weights were entered in the LRHE (using the parameters learned in the discovery set), in order to predict the easy and difficult cases. Finally, the LRC (with parameters learned on the discovery set) was applied to predict clinical status for all subjects in the validation set, and an accuracy score was derived separately for the easy and difficult cases.

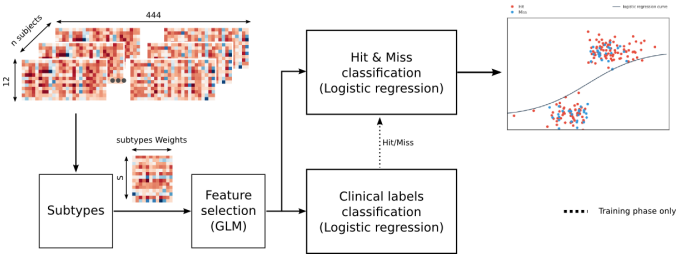

Fig. 2. Caption

2.6 Calibrating the problem

As a first step we need to assess the difficulty of the classification task at hand and to do so we are throwing our problem to a simple off the shelf model with the default parameters. This will define the complexity of the problem. In our case we have used scales 64 which will give us a good reference point in term of the problem complexity. The initial classifier model used was a linear SVM with parameter fixed to $C = 1$ (the default value). In all the subsequent analysis we are performing a stratified 10-fold cross-validation and include a parameter in the SVM classifier to account for the unbalance dataset by automatically re-weight each class by the inverse of its frequency.

The next step is to find the most adapted hyper-parameters (C and Gamma if using Gaussian kernel) for our classification problem we choose to use a grid search approach. The grid search was parametrized as follow for C (10^-2 to 10^3) and for Gamma (10^-5 to 10^1).

2.7 Preprocessing and confounds regression

SVM are not scale invariant it is therefore very important to apply normalization on the data. We therefore normalize the connectivity features of each subject to zero mean and unit

variance. Since some bias can be introduce by confounding factors we account for them by regressing the age and gender contribution based on the training set.

2.8 Optimal functional scale

The fact that we are looking at various level of data abstraction due to the clustering process of BASC in functional scales, we may be more sensitive at some particular scale than other for a given pathology. We have therefore tested the model for each of the 9 scales.

2.9 Multiscale bagging predictions

The idea in this case was to combine the votes of the classifier at each scale using a bagging approach. This ensemble approach combine the predictor of each scale in a bagging procedure as illustrated in Figure **??**. We simply take a majority vote from all the classifiers. Contrary to most ensemble methods this variant may be sensitive to the choice of the scales, particularly scales that do not have sufficient information to yield a good prediction. We therefore took a subset of the scales based on the individual performance obtain in Figure **??**.

3 Results

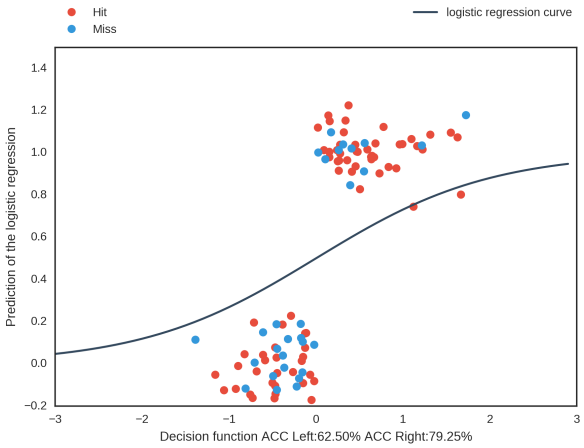

Fig. 3. Caption

If we combine the confound regression with the normalization (unit variance and zero mean for the connectivity of each subject) and we

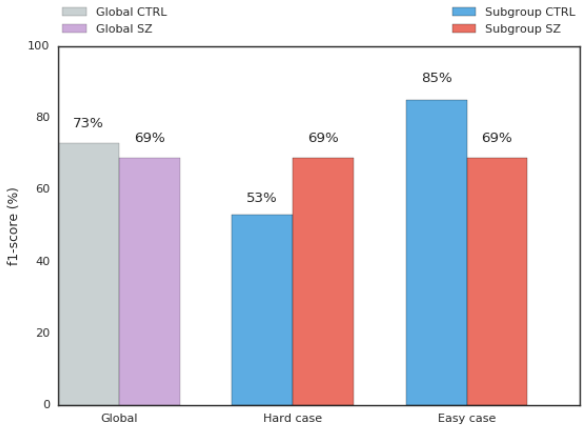

Fig. 4. Caption

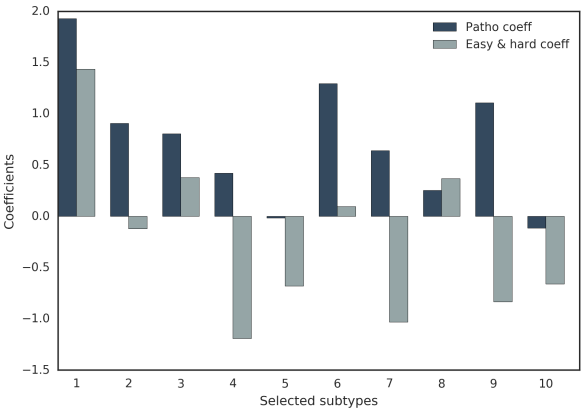

Fig. 5. Caption

search for the optimal parameter C for the linear classifier we obtain 69.89% and an AUC of 0.80 see Figure **??**.

4 Discussion and Conclusion

The use of subtypes weight is a better way to reduce the dimension of the problem and for coding for complex structure in a compact representation.

Difficult cases: factors, Noise in the data, cognitive state (sleep etc...) Easy case: factors subgroup of subject were we have enough examples to reach a conclusion physiologically.

we found subtypes associated with the pathology and we have been able to learn from the prediction to identify a subgroup of subject that can be accurately classified.

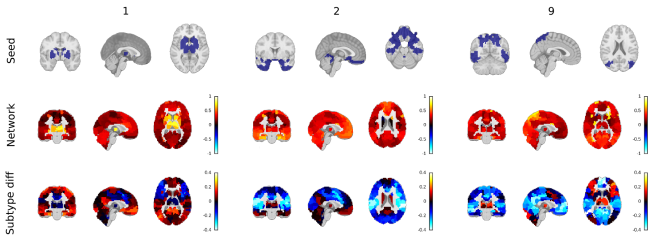

Fig. 6. Caption

TABLE 1
Summary results with no split. The accuracy is 71.3%

Class	*precision*	*recall*	*f1-score*	*support*
0	78%	69%	73%	58
1	64%	74%	69%	43
avg/total	72%	71%	71%	101

TABLE 2
Summary results with split on the right. The accuracy is 79.2%

Class	*precision*	*recall*	*f1-score*	*support*
0	81%	88%	**85%**	34
1	75%	63%	69%	19
avg/total	79%	79%	79%	53

TABLE 3
Summary results with split on the left. The accuracy is 62.5%

Class	*precision*	*recall*	*f1-score*	*support*
0	71%	42%	53%	24
1	59%	83%	69%	24
avg/total	65%	60%	61%	48

By leveraging subtypes of FC, we were able to identify a subgroup of easy cases where it is possible to accurately (80%) predict clinical diagnosis based on individual FC maps. The prediction model was trained completely independently of our validation set, and the accuracy estimate are unbiased. The 80% accuracy is very satisfactory considering that both our training and validation sets comprised subjects scanned at different sites, using different imaging and behavioral protocols. Our findings suggest that the clinical labels may be ill-defined in terms of brain connectivity patterns, and that a subgroup population exists who presents a more definitive functional signature of clinical symptoms. We plan to investigate next the spatial organization of brain connectivity subtypes associated with good vs poor prediction accuracy.

A common mistake made when trying to tackle this kind of problem is to do the feature selection before cross-validation using the labels. This may lead to an overestimate of our accuracy, it may give very good result that will not generalize well. As an example we have performed the same experiment with the feature selection procedure (I-Relief) in and out of the cross-validation loop resulting in improved performance when selecting the features outside the cross-validation loop.

REFERENCES

[1] D. L. Collins, P. Neelin, T. M. Peters, and A. C. Evans, "Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space." *Journal of computer assisted tomography*, vol. 18, no. 2, pp. 192–205, 1994. [Online]. Available: http://view.ncbi.nlm.nih.gov/pubmed/8126267

[2] V. Fonov, A. C. Evans, K. Botteron, C. R. Almli, R. C. McKinstry, D. L. Collins, and Brain Development Cooperative Group, "Unbiased average age-appropriate atlases for pediatric studies." *NeuroImage*, vol. 54, no. 1, pp. 313–327, Jan. 2011. [Online]. Available: http://dx.doi.org/10.1016/j.neuroimage.2010.07.033

[3] Y. Ad-Dab'bagh, D. Einarson, O. Lyttelton, J. S. Muehlboeck, K. Mok, O. Ivanov, R. D. Vincent, C. Lepage, J. Lerch, E. Fombonne, and A. C. Evans, "The CIVET Image-Processing environment: A fully automated comprehensive pipeline for anatomical neuroimaging research," in *Proceedings of the 12th Annual Meeting of the Human Brain Mapping Organization*, M. Corbetta, Ed. Florence, Italy: Neuroimage, 2006.

[4] J. D. Power, K. A. Barnes, A. Z. Snyder, B. L. Schlaggar, and S. E. Petersen, "Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion." *NeuroImage*, vol. 59, no. 3, pp. 2142–2154, Feb. 2012. [Online]. Available: http://dx.doi.org/10.1016/j.neuroimage.2011.10.018

[5] F. Giove, T. Gili, V. Iacovella, E. Macaluso, and B. Maraviglia, "Images-based suppression of unwanted global signals in resting-state functional connectivity studies." *Magnetic resonance imaging*, vol. 27, no. 8, pp. 1058–1064, Oct. 2009. [Online]. Available: http://dx.doi.org/10.1016/j.mri.2009.06.004

[6] T. E. Lund, K. H. Madsen, K. Sidaros, W.-L. Luo, and T. E. Nichols, "Non-white noise in fMRI: does modelling have an impact?" *NeuroImage*, vol. 29, no. 1, pp. 54–66, Jan. 2006. [Online]. Available: http://dx.doi.org/10.1016/j.neuroimage.2005.07.005

[7] P. Bellec and A. C. Evans, "A pipeline system for Octave and Matlab," 2010.

[8] K. J. Worsley and K. J. Friston, "Analysis of fMRI Time-Series Revisited—Again," *NeuroImage*, vol. 2, no. 3, pp. 173–181, Sep. 1995. [Online]. Available: http://dx.doi.org/10.1006/nimg.1995.1023

[9] Y. Benjamini and Y. Hochberg, "Controlling the false-discovery rate: a practical and powerful approach to multiple testing." *J. Roy. Statist. Soc. Ser. B*, vol. 57, pp. 289–300, 1995.

[10] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "Liblinear: A library for large linear classification," *The Journal of Machine Learning Research*, vol. 9, pp. 1871–1874, 2008.