Discriminative Network Models of Schizophrenia: Supplementary Material

Guillermo A. Cecchi, Irina Rish
IBM T. J. Watson Research Center
Yorktown Heights, NY, USA

Benjamin Thyreau
Neurospin
CEA, Saclay, France

Bertrand Thirion
INRIA
Saclay, France

Marion Plaze
INSERM - CEA - Univ. Paris Sud
Research Unit U.797
Neuroimaging & Psychiatry
SHFJ & Neurospin, Orsay, France

Marie-Laure Paillere-Martinot
AP-HP, Adolescent Psychopathology
and Medicine Dept., Maison de Solenn,
Cochin Hospital, University Paris Descartes
F-75014 Paris, France

Catherine Martelli
Departement de Psychiatrie
et d'Addictologie
Centre Hospitalier Paul Brousse
Villejuif, France

Jean-Luc Martinot
INSERM - CEA - Univ. Paris Sud
Research Unit U.797
Neuroimaging & Psychiatry
SHFJ & Neurospin, Orsay, France

Jean-Baptiste Poline
Neurospin
CEA, Saclay, France

1 Computing Activations maps

Given the time-series for stimulus $s(t)$ (e.g., s=1 if the stimulus/event is present, and s=0 otherwise), and the BOLD signal intensity time-series $v_i(t)$ for voxel i, General Linear Model (GLM)[1] is simply a linear regression $v_i(t) = \beta_i * \hat{s}(t) + b_i + \epsilon$, where $\hat{s}(t) = HRF(t) \bigotimes s(t)$ is the regressor corresponding to the stimulus convolved with the *hemodynamic response function (HRF)* in order to account for delay between the voxel activation and change in the BOLD signal, ϵ is noise, b is the baseline (mean intensity) and β_i coefficient is the *amplitude* that serves as an *activation score* (note that β_i coefficient is simply the correlation between v_i and $\hat{s}(t)$ when both are normalized and centered prior to fitting the model). Given multiple trials, multiple estimates of β_i are obtained and a statistical test (e.g., t-test) is performed for the mean $\bar{\beta}_i$ against the null-hypothesis that it comes from Gaussian noise distribution with zero mean and fixed noise σ (the level of noise for BOLD signal is assumed to be known here).

In case of multiple stimuli, the GLM model uses a vector of regressors $\hat{s}(\mathbf{t})$ and obtains the vector of the corresponding coefficients β. For example, in our studies, the following stimuli/events were considered: 'FrenchNative', 'Foreign', and 'Silence', together with several additional regressors, such as some low-frequencies trends and the movement parameters (additional 1-only column is added to account for the mean of the signal, as above - a standard step in linear regression with the un-normalized data). Once the GLM is fit, we focus on the β_i coefficients obtained for the above three stimuli, and the corresponding three activation maps. Next, we compute several "contrast" maps by subtracting some maps from the others (hoping that such differences, or contrasts, may provide additional information). The following activation "contrast" maps were computed: *activation contrast 1: "FrenchNative - Silence", activation contrast 2: "FrenchNative - Foreign", activation contrast 3: "Silence - FrenchNative", activation contrast 4: Foreign - FrenchNative* (note that maps 2 and 4 are just negations of the maps 1 and 3, respectively), *activation contrast 5: "Foreign - Silence"*; also, the following three contrast maps are simply the difference of the corresponding β_i coefficient

---

[1] The GLM analysis described here is a standard component of the Statistical Parametric Mapping (SPM) toolkit (see [1] for more details).
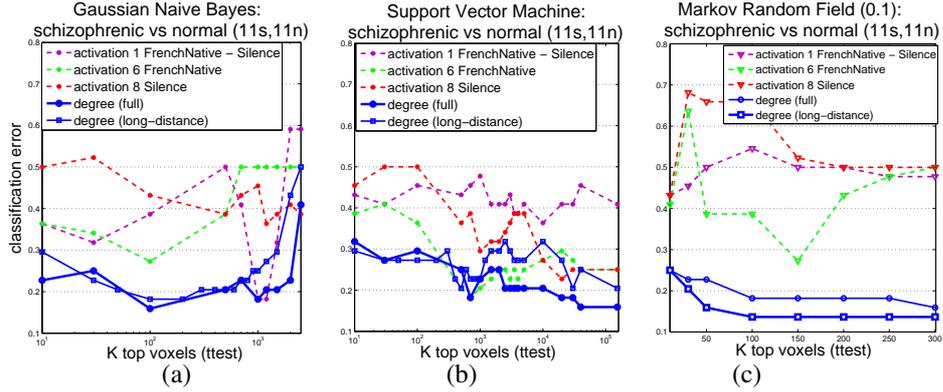
Figure 1: Classification results comparing GNB, SVM and sparse MRF classifiers on *unnormalized* (raw) activation maps vs degree maps.

(activation) and the mean (b_i): *activation contrast 6: "FrenchNative"*, *activation contrast 7: "Foreign"*, *activation contrast 8: "Silence"*. For each of those maps, t-values are computed at each voxel (with a null-hypothesis corresponding to zero-mean Gaussian). *In the analysis presented in this paper, we use the resulting t-value maps, rather than just the "raw" activation maps (i.e., β coefficient maps), and to simplify the terminology, just refer to them as "activation" or "activation contrast" maps.*

2 Classification Results with Unnormalized Activation Maps

We observed that normalization was essential for improving the performance of activation maps when using sparse MRF classifiers; while the raw activation maps perform similarly to normalized ones in case of GNB and SVM, their performance degrades dramatically (to 40-50% error) when using MRFs: (compare Figure 1c with the figure 4c in the main paper).

3 Classification Results with Subset of Subject

For completeness sake, we include here our initial results from the submitted version of this paper in Figure 2, were a subset of 9 schizophrenic subjects (and thus only 2*9+2*11 = 40 samples, and 20 cross-validation fold) was used instead of the full set of 11 subjects; after checking the motion correction results more carefully in response to the reviewers comments, we realized that the two schizophrenic subjects excluded originally were actually good-quality data, and therefore included then in the final set of results. Also, the initial submission included comparison with only one activation maps - map 8 that performed best while using SVM on the full set of voxels; afterwards, we ran experiments with all activation maps using voxel selection and included the best-performing activation maps in all regimes. Note that the results on the subset of data are very similar to the ones presented in the final version of the paper, with similarly noticeable differences in the performance of degree maps (only 12% error) and activation maps (over 30%) when using MRFs.

4 Global Topological Features

Various *global* topological features were extracted: (1) the *mean degree*, i.e. the number of links for each node (corresponding to a voxel), averaged over the entire network; (2) the *mean geodesic distance*, i.e. the minimal number of links needed to reach any to from any other node, averaged over the entire network; (3) the *mean clustering*, i.e. the fraction of triangulations formed by a node with its first neighbors relative to all possible triangulations, averaged over the entire network; (4) the *giant component*, i.e. the size (number of nodes) of the largest connected sub-graph in the network; (5) the *giant component ratio*, i.e. the ratio of the giant component to the size of the network; (6) the *total number of links* in the network.
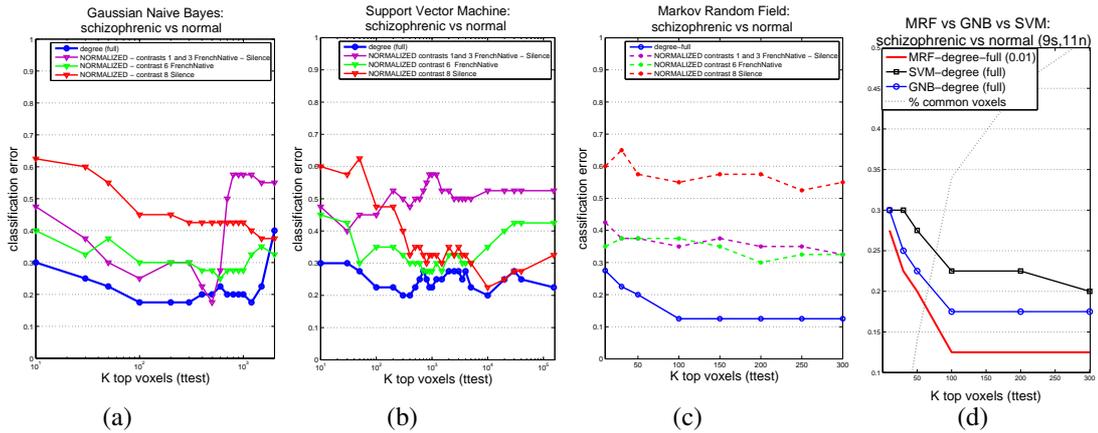
Figure 2: Initial results on a subset including 9 schizophrenic subjects. Classification results comparing (a) GNB, (b) SVM and (c) sparse MRF on degree versus activation maps; (d) all three classifiers compared on best-performing features (full degree maps); dashed line shows the stability of variable-selection as % of common voxels over 20 CV folds.

Feature	Mean normal+alcoholic	Mean schizophrenic	p-value
geodesic dist.	11.8	22.0	1.5×10^{-3}
clustering coeff.	0.14	0.07	$< 10^{-4}$
giant component ratio	0.87	0.48	$< 10^{-5}$
giant component	26220.2	15698.0	6.5×10^{-3}
nodes	28416.9	20823.6	2.9×10^{-2}
degree	28.1	16.7	5.3×10^{-2}
total links	925573	627114	0.29
mean activation	0.335	0.289	5.5×10^{-4}

Table 1: Global features.

5 Classification Results with Alcoholic Group

We also considered the task of discriminating between the schizophrenic group versus the aggregate of the normal group and a second control group formed by patients suffering from alcoholism [2]. The rational to include this additional group was to test for possible non-specific disruptions in the signal simply due to the brain being in a dysfunctional state. Moreover, a prominent feature of this group is a higher level of movement inside the scanner, prior to movement corrections; therefore, the alcoholic group also provides a test for potential movement confounds. Indeed, the classification results are such that both these hypotheses (i.e. non-specificity and movement artifacts) can be rejected.

In these experiments were focused on the comparison with activation map 8 that demonstrated best predictive performance when using SVM on the full set of voxels; while full exploration of all activation maps in the low-voxel regime was not yet performed with inclusion of alcoholic group, it is more important to see that the degree maps again demonstrate similarly high predictive power to the one observed in our main study (schizophrenic vs normal groups).

(**a**) Results of applying Gaussian Naïve Bayes classification to the degree maps, compared with the activation map, are shown in Fig. 1a. Full degree maps achieve an error rate of 12% around 300 voxels, whereas contrast maps reach 22% for 3000 voxels. Moreover, the difference in the error rate for small number of voxels is remarkable, as contrast maps perform nearly at random level, whereas the degree maps perform consistently below 20% error. (**b**) Support Vector Machine classification results are shown in Fig. 1b. In this case, the best error rate (10%) is achieved by inter-hemispheric maps with around 200 voxels, which is indeed the least for all maps. Contrast maps, on the other hand, stay above 30% error for most pre-selected voxel sizes, and reach 35%

for the size corresponding to the least degree map error of 10%. (**c**) Sparse Markov Random Field results are shown in Fig. 1c. Given the computational demands of this approach, we only show results for up to 300 pre-selected voxels. For contrast maps the error rate is consistently above 30%, although the densest model (0.0001) is significantly better than the sparser ones, which provides more evidence that, even for contrast maps, synergetic interactions are relevant. The picture is even sharper for degree maps: these maps are very sensitive to the sparsity parameter, with sparse models averaging 30% error, and denser ones reaching stable error rates of 12% for 0.001. Taking together, the comparison between contrast maps and sparser degree maps on the one hand, and on the other the denser degree maps, indicates that the functional synergy captured by the correlation networks, and included in the MRF model, provide a very significant signal for the classification task. This effect supports our hypothesis that schizophrenia implies a functional network disruption which is not reducible to an activation disruption. (**d**) The p-values for the comparison between the schizophrenic group v. normal+alcoholic group, for the various global features, are presented in Table 1. Observe that the mean activation shows a good statistical power; however, when these global features are fed into a classifier (Table 2), the corresponding results for mean activation are quite poor, whereas the topological features display much better accuracy.

Feature	(GNB	SVM	MRF(0.01)
degree (D)	19%	13.8%	15.5%
clustering coeff. (C)	19.0%	29.3%	31.0%
geodesic dist. (G)	34.5%	31.0%	31.0%
mean activation (A)	29.3%	31%	31%
D + A	13.4%	13.8%	15.5%
C + A	17.2%	31%	31.0%
G + A	39.7%	31.0%	31.0%
G +D +C	24.0%	14.0%	19.0%
G+D+C+A	17.2%	13.8%	15.5%

Table 2: Classification errors using global features schizophrenics vs. normal+alcoholics, baseline error about 31%.

References

[1] K. Friston, J. Ashburner, S. Kiebel, T. Nichols, and W. Penny. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier, New York, NY, 2009.

[2] Chanraud-Guillermo S, Andoh J, Martelli C, Artiges E, Pallier C, Aubin HJ, Martinot JL, and Reynaud M. Imaging of language-related brain regions in detoxified alcoholics. *Alcohol Clin Exp Res.*, 33(6):977–84, June 2009.
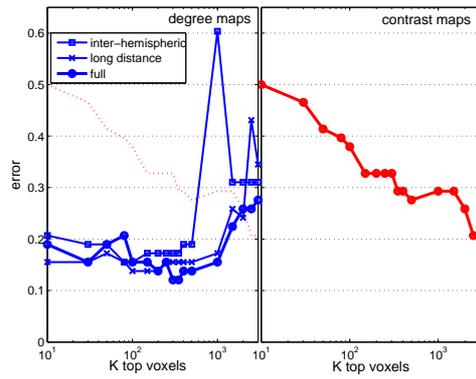
(a) GNB results
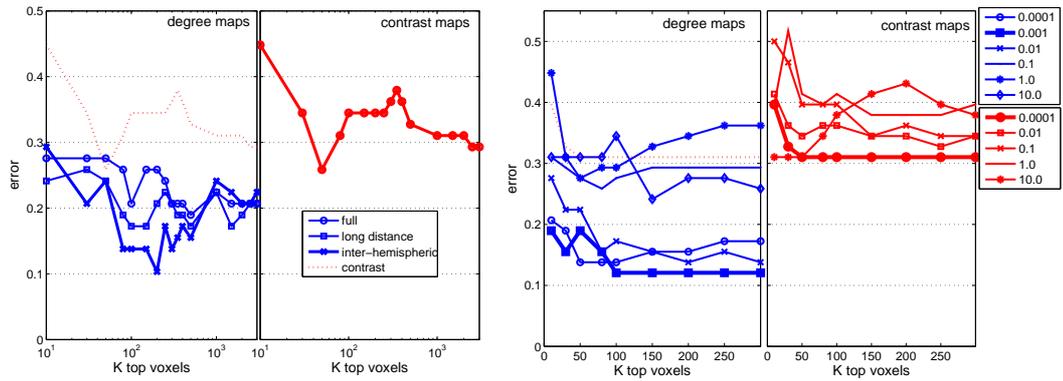
(b) SVM results (c) MRF results

Figure 3: Results for schizophrenic vs (normal+alchoholic) classification.