

PREDICTING INDIVIDUAL SCORES FROM RESTING STATE FMRI USING PARTIAL LEAST SQUARES REGRESSION

Djalel-E. Meskaldji^{*†‡} Maria Giulia Preti^{*†} Thomas Bolton^{*†} Marie-Louise Montandon[∨]
Cristelle Rodriguez[^] Stephan Morgenthaler[‡] Panteleimon Giannakopoulos[^]
Sven Haller[∨] Dimitri Van De Ville^{*†}

^{*} Medical Image Processing Lab (MIPLAB), Institute of Bioengineering, EPFL, CH-1015 Lausanne, Switzerland

[†] Department of Radiology and Medical Informatics, University of Geneva, CH-1211 Geneva, Switzerland

[∨] Divisions of Diagnostic and Interventional Neuroradiology, Geneva University Hospitals, Geneva, Switzerland

[^] Department of Mental Health and Psychiatry, University Hospitals of Geneva, Switzerland

[‡] Applied statistics, Institute of mathematics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

ABSTRACT

An important question in neuroscience is to reveal the relationship between individual performance and brain activity. This could be achieved by applying model regression techniques, in which functional connectivity derived from resting-state functional magnetic resonance imaging (fMRI), is used as a predictor. However, due to the large number of parameters, prediction becomes problematic and regression models cannot be found using the traditional least squares method. We study the ability of fMRI data to predict long-term-memory scores in mild cognitive impairment subjects, using partial least squares regression, which is an adapted method for high-dimensional regression problems. We also study the influence of the sample size on the performance, the stability and the reproducibility of the prediction.

Index Terms— Functional connectivity, sample size, bootstrap, anti-correlation, fMRI.

1. INTRODUCTION

Functional connectivity (FC) measures the interaction between spontaneous fluctuations of brain activity at different spatial locations. Conventionally, FC is estimated between each pair of brain regions of interest (ROIs) using the Pearson correlation between the corresponding functional magnetic resonance imaging (fMRI) time courses. However, other measures that better reflect spontaneous co-activation-deactivation of ROIs, namely accordance and discordance, were introduced in [1]. Aiming to highlight the relationship between brain activity and individual memory performance, regression models are fitted to study the importance of connections on individual scores. However, in typical brain connectivity studies, the number of ROIs is $\mathcal{O}(10^2)$ nodes, resulting in $\mathcal{O}(10^3 - 10^4)$ connections, which is much larger than the number of observations (subjects). It is then necessary to consider regression methods that overcome this limitation. In this paper, using cross validated partial least

squares regression (PLSR), we study the ability of predicting individual long-term-memory (LTM) scores from resting state fMRI data of subjects with mild cognitive impairment (MCI), expressed in brain connectivity measured by either Pearson correlation, accordance or discordance measures. This is much more challenging than a binary classification task. We also study the influence of the sample size on the stability of the learned models, which is of high importance to guarantee a certain level of multi site reproducibility and also in meta-analysis studies.

2. METHODS

2.1. Description of the data

This study involves 57 subjects (mean age 74.5 ± 6.15 , 34 males and 23 females) diagnosed with MCI. For each subject, we have resting-state fMRI data (TR=2300 ms) [2] as well as CERAD 10-word List Delayed Recall Test (Consortium to Establish a Registry for Alzheimer's Disease; [3]) scores. The memory test consists in recalling ten words. The score is the number of recalled words and hence, it ranges in $\{0, \dots, 10\}$.

2.2. FC estimation

Anti-correlation in brain activity could be a key information to assess brain disfunction and disorder. However, this information is hidden by positive correlations since the Pearson correlation is an averaging operator. Meskaldji et al. [1] proposed to separate anti-correlation from positive correlation and introduced two new FC measures, namely; accordance and discordance. After soft thresholding the normalized fMRI signal to keep only significant activations and deactivations of the brain ROIs, accordance and discordance measure the proportion of time points in which two time courses have the same sign or opposite sign, respectively. More specifically, fixe a positive and a negative threshold u and l , respectively. Let $\mathbf{x}^u = \{x_t^u, t = 1, \dots, T, x_t^u = 1 \text{ if } x_t > u \text{ and } 0 \text{ otherwise}\}$ and $\mathbf{x}^l = \{x_t^l, t = 1, \dots, T, x_t^l = -1 \text{ if } x_t < l \text{ and } 0 \text{ otherwise}\}$ be the upper and lower soft thresholded vectors of the nor-

This work was supported in part by the Swiss National Science Foundation (grant number PP00P2-146318), and in part by the Center for Biomedical Imaging (CIBM) of the Geneva-Lausanne Universities and the EPFL, as well as the Leenaards and Louis-Jeantet foundations.

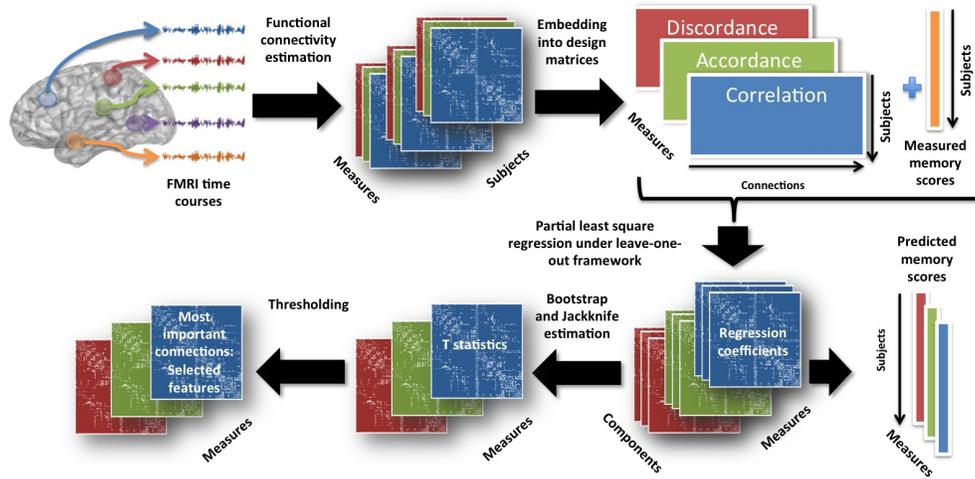


Fig. 1. Processing pipeline to derive predictive models and to select most important features.

malized time course \mathbf{x} of length T . Set

$$\sigma = \sqrt{(\mathbf{x}^u * \mathbf{x}^u) + (\mathbf{x}^l * \mathbf{x}^l)}, \quad (1)$$

where $*$ designate the inner product operator. Then, the accordance and the discordance values between two ROIs i and j , with corresponding time courses x_i and x_j , are given by

$$a_{i,j} = (\mathbf{x}^u_i * \mathbf{x}^u_j + \mathbf{x}^l_i * \mathbf{x}^l_j) / (\sigma_i \sigma_j) \quad (2)$$

$$d_{i,j} = (\mathbf{x}^u_i * \mathbf{x}^l_j + \mathbf{x}^l_i * \mathbf{x}^u_j) / (\sigma_i \sigma_j). \quad (3)$$

In this study, the ROIs are defined according to the AAL atlas of 88 ROIs. Hence, for each FC measure (Pearson, accordance and discordance) and for each subject, we end up with $\frac{1}{2}88 \times 87 = 3828$ connectivity values.

2.3. Regression for high dimension problems

We assume the following model for our data

$$Y = \mathbf{X}\beta + \mathcal{E} \quad \mathcal{E} \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \mathbf{I}_n) \quad (4)$$

where \mathbf{X} is an $n \times p$ matrix that contain the $p = 3828$ connectivity values for the $n = 57$ subjects, and Y is an $n \times 1$ response vector representing the 57 individual LTM scores. The least squares minimizes the quadratic loss

$$L(\hat{\mathbf{Y}}) = \|\mathbf{Y} - \hat{\mathbf{Y}}\|_2^2. \quad (5)$$

and its solution for the model is given by

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \quad (6)$$

This solution demands the computation of the inverse of the $(\mathbf{X}^T \mathbf{X})^{-1}$ matrix. This is possible only in full rank regression problems, that is, when the number of samples n is larger than the number of parameters p . In modern applications, p is often large with $p \gg n$ as in our case, making the prediction and feature selection more difficult.

2.4. Prediction

Partial least squares regression (PLSR) is one of the multivariate high-dimensional regression methods that has been widely used in various application fields such as social sciences, bioinformatics and neuroscience [4, 5, 6]. Instead of projecting the response vector into the space spanned by the co-variables, PLSR projects both the response and the co-variables into a new space formed by latent variables by decomposing both \mathbf{X} and \mathbf{Y} into orthogonal scores and loadings.

The general formulation is given by

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathcal{E}_X \quad (7)$$

$$\mathbf{Y} = \mathbf{U}\mathbf{Q}^T + \mathcal{E}_Y, \quad (8)$$

where \mathbf{T} and \mathbf{U} are scores, while \mathbf{P} and \mathbf{Q} are the loading matrices for \mathbf{X} and \mathbf{Y} , respectively. \mathcal{E}_X and \mathcal{E}_Y are independent errors. The scores and the loadings are chosen to explain the maximum variance between \mathbf{X} and \mathbf{Y} . PLSR tries to find the directions in \mathbf{X} that better explain the maximum variance in \mathbf{Y} . There are many algorithms to solve the PLSR problem. In our application, we used the "PLS" [7] R-package (<http://cran.r-project.org>).

It is important to quantify the ability of a model to predict response values for samples that have not been included in the training data. This technique is called cross validation (CV), in which a subset of individuals is excluded (testing data), a regression model is built on the basis of the remaining data (training data), then a response value is predicted using testing data. In our study, we used leave one out (LOO), in which the testing data consists of one observation in each iteration. The quality of the model is quantified by the distance between real and predicted response values. In this paper, we quantify this distance by the correlation between predicted and measured LTM scores.

2.5. Feature selection

In typical high dimension applications, most of the elements of β are 0. Hence, it is worthwhile to measure importance

of the regressors in the model, and to identify a small subset of predictive features. This could have different purposes namely, model interpretation and guiding future data collection and extraction. Model selection then solves

$$\min L(\mathbf{X}\hat{\beta}) \quad \text{subject to} \quad \|\hat{\beta}\|_{l_0} = \sum_{i=1}^p I_{\{\beta_i \neq 0\}} \leq K, \quad (9)$$

where the number of nonzero features, K , is the desired sparsity. For example, the least absolute shrinkage and selection operator (LASSO) [8] imposes an l_1 regularization on the regression coefficients. However, the number of regressors could not exceed the sample size under LASSO regularization. For this reason, [9] introduced the elastic net in which an l_2 constraint (Ridge) is added to the Lasso regularization. However, it adds more complexity in tuning the sparsity parameters.

The PLSR benefits also from dimensionality reduction by using only few loading components. The optimal number of components is based on a cross-validated root mean squared error of prediction (RMSEP) [7]. In addition, to assess the importance of the variables in the regression model, we use the JackKnife bootstrap estimator implemented in the PLS R-package [7], which ends up with a significance p-value for each variable in the model, that is, each connection. Feature selection is done by thresholding those p-values (Fig. 1).

2.6. Sample size influence

In order to study the influence of the sample size on the performance and the stability of the prediction, we randomly selected a sub-sample of size $n < 57$ (20, 30, 40 and 50) and we applied the prediction process described in Fig. 1. We repeated this operation 1000 times and we estimated some prediction performance summary statistics, namely, mean, median and standard deviation.

3. RESULTS AND DISCUSSION

The main objective of this work was to study the ability of predicting individual scores from fMRI data. With respect to this first question, it was found that both FC measures, that is, Pearson correlation, accordance and discordance, contain a considerable amount of information which could be related to individual memory performance. The prediction performance in the case of the whole sample is 0.555, 0.567 and 0.645 for Pearson correlation, accordance and discordance, respectively. Interestingly, the discordance is the FC measure that contains the most relevant information. This result might be explained by the fact that intuitively, brain regions and in particular, those implicated with memory tasks are working in a cooperative way, but some other regions are working in a competitive way. There are, however, other possible explanations. These findings suggest that anti-correlation between fMRI time courses should be considered carefully in brain FC studies. After thresholding the p-values corresponding to the connections at 0.005, we obtain a certain number of connections that implicates default mode network regions (DMN) regions, such as superior frontal and temporal lobes as well as hippocampus (Fig. 2), which supports previous findings on relationship between LTM scores and DMN brain regions [10, 11]. The results obtained in this work raise intriguing

questions regarding the nature and the extent of information contained in the discordance measure.

n	Correlation	Accordance	Discordance
20	0.434 (0.147)	0.417 (0.148)	0.444 (0.159)
30	0.451 (0.118)	0.448 (0.118)	0.479 (0.118)
40	0.481 (0.104)	0.494 (0.094)	0.545 (0.080)
50	0.520 (0.070)	0.540 (0.060)	0.607 (0.040)
57	0.555 (0)	0.567 (0)	0.646 (0)

Table 1. Average (and standard deviation) of PLSR prediction performance in terms of correlation between predicted and measured LTM scores. FC is derived either by Pearson correlation, accordance or discordance. The sub-sample size is either 20, 30, 40 or 50. The estimations are obtained after 1000 iterations for each case.

The present study was also designed to determine the effect of the sample size on two important factors; prediction and feature selection. The most obvious finding to emerge from the analysis is that removing subjects could have a dramatic influence on the learned predictive models. For all FC measures, the mean (Tab. 1) and the median (Fig. 3) of the performance increases with the sample size. However, this performance has larger variance for smaller sample size and can reach very high and very small values as well. This shows that the sample size is an important factor to produce prediction models that have certain stability features, which is extremely important for reproducibility and meta-analysis studies. But surprisingly, discordance model was also found to be the most stable compared to the other measures. The optimal number of components chosen by the RMSEP is also fluctuating more with small samples. The median number of components increases, however, with the sample size and seems to converge to the optimal number of components for the whole sample which is 6, 5 and 6 for Pearson correlation, accordance and discordance, respectively. The fact that the number of components needed for the prediction increases with the sample size reflects the heterogeneity of the sample and indicates that it is important to chose a sample that reflects the same heterogeneity of the population being studied.

4. CONCLUSION

The major finding of this paper is the ability of fMRI data to predict individual scores under LOO-CV, which is promising and encouraging for further investigations on the relationship that exists between brain behavior and function using FC. We hope that this work will serve as a basis for future brain predictive studies. We showed also that the new measures of FC, in particular, the discordance, contains key information concerning brain function. In addition, our study confirms the role of DMN regions in individual memory performance.

We also studied the influence of the sample size on the performance of PLSR to predict individual scores from fMRI data under LOO-CV framework. We showed the importance of having a reasonable training sample size, to learn a good and robust prediction model. This guarantees more stability and lead to better reproducibility of the results, which is very important for meta-analysis studies. Finally, PLSR is a good

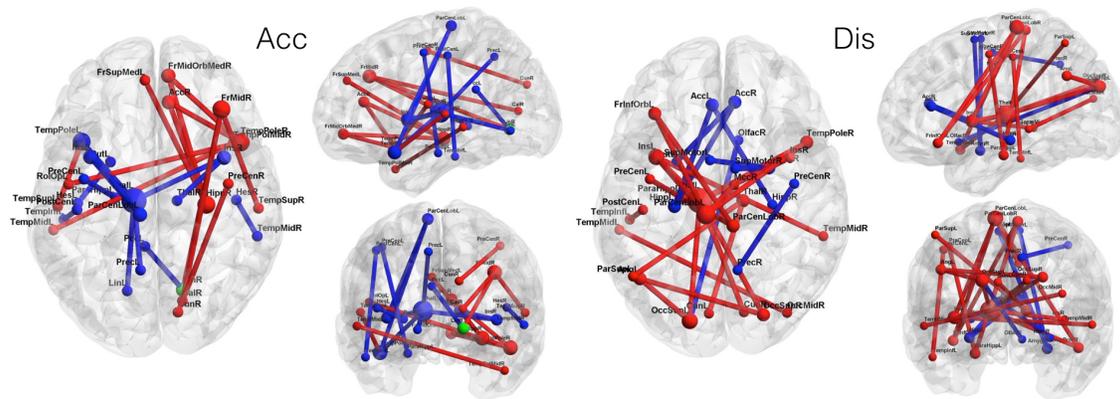


Fig. 2. Significant connections obtained by thresholding p-values (at level 0.005) corresponding to the coefficients of the accordance and the discordance based PLSR models. The connection color represents the sign of the coefficient corresponding to each connection in the PLSR model (red for positive coefficients and blue for negative coefficients).

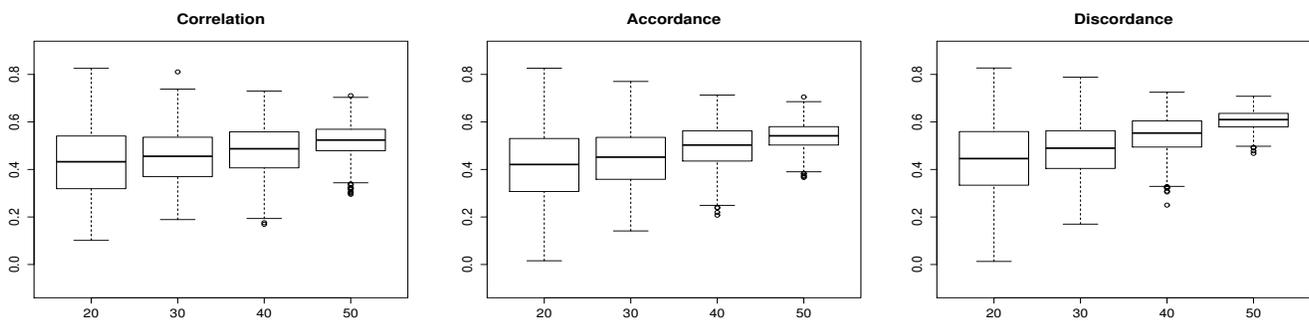


Fig. 3. The influence of sample size on PLSR prediction performance in terms of correlation between predicted and measured memory scores. FC is derived either by Pearson correlation, accordance or discordance. The sub-sample size is either 20, 30, 40 or 50. The box plots are obtained after 1000 iterations.

solution for high-dimensional regression methods. However, how to assess the importance of variables in the model is still an open question in machine learning and signal processing fields.

5. REFERENCES

- [1] D.-E. Meskaldji, S. Morgenthaler, and D. V. D. Ville, "New measures of brain functional connectivity by temporal analysis of extreme events," *Proceedings of the Twelfth IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, vol. (ISBI'15), pp. 26–29, 2015.
- [2] J. Richiardi, A. U. Monsch, T. Haas, F. Barkhof, D. Van de Ville, E. W. Radü, R. W. Kressig, and S. Haller, "Altered cerebrovascular reactivity velocity in mild cognitive impairment and alzheimer's disease," *Neurobiology of aging*, vol. 36, no. 1, pp. 33–41, 2015.
- [3] K. A. Welsh, N. Butters, R. C. Mohs, D. Beekly, S. Edland, G. Fillenbaum, and A. Heyman, "The consortium to establish a registry for alzheimer's disease (cerad). part v. a normative study of the neuropsychological battery," *Neurology*, vol. 44, no. 4, pp. 609–609, 1994.
- [4] R. D. Cramer, D. E. Patterson, and J. D. Bunce, "Comparative molecular field analysis (comfa). I. effect of shape on binding of steroids to carrier proteins," *Journal of the American Chemical Society*, vol. 110, no. 18, pp. 5959–5967, 1988.
- [5] D. V. Nguyen and D. M. Rocke, "Tumor classification by partial least squares using microarray gene expression data," *Bioinformatics*, vol. 18, no. 1, pp. 39–50, 2002.
- [6] A. McIntosh, F. Bookstein, J. V. Haxby, and C. Grady, "Spatial pattern analysis of functional brain images using partial least squares," *Neuroimage*, vol. 3, no. 3, pp. 143–157, 1996.
- [7] B.-H. Mevik and R. Wehrens, "The pls package: principal component and partial least squares regression in r," *Journal of Statistical Software*, vol. 18, no. 2, pp. 1–24, 2007.
- [8] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.
- [9] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.
- [10] B. T. Yeo, F. M. Krienen, J. Sepulcre, M. R. Sabuncu, D. Lashkari, M. Hollinshead, J. L. Roffman, J. W. Smoller, L. Zöllei, J. R. Polimeni, et al., "The organization of the human cerebral cortex estimated by intrinsic functional connectivity," *Journal of neurophysiology*, vol. 106, no. 3, pp. 1125–1165, 2011.
- [11] J. R. Andrews-Hanna, "The brain's default network and its adaptive role in internal mentation," *The Neuroscientist*, vol. 18, no. 3, pp. 251–270, 2012.