BRAIN AS A SELF-PREDICTOR: SPARSE FULL-BRAIN AUTO-REGRESSIVE MODELING IN FMRI

Rahul Garg, Guillermo A. Cecchi, A. Ravishankar Rao

IBM T.J. Watson Resarch Center, Yorktown Heights, NY, USA

ABSTRACT

We demonstrate a method to build an autoregressive model for the whole brain without carrying out any aggregation of the fMRI data. The model gives biologically meaningful results and has several desirable properties. We show that the model gives significantly improved prediction on unseen data as compared to baseline methods. The voxels with better prediction are distributed throughout the brain, including the task positive and task negative regions. In addition to the active regions identified by the general linear model (GLM), our analysis also uncovers complex interactions among the regions involved in the default mode networks.

Index Terms— Autoregressive modeling, fMRI, Granger causality, prediction, functional connectivity.

1. INTRODUCTION

The resolution of temporal interactions between brain areas in functional magnetic resonance imaging (fMRI) has gained increased attention in recent years [10, 11], resulting in a proliferation of methods. This impetus stems from the assumption that the input to the brain is, for the most part, the brain itself [18, 6, 2], and made evident by the recent identification of the system known as default mode network [22, 9]. Most functional dynamics methods in the recent literature are constrained by the need to aggregate voxels in pre-defined regions of interests [16], or restricted to voxel pairwise zero-lag covariance [8]. Voxel aggregation potentially implies discarding useful voxel-level information, and inviting artifacts unless very carefully performed, while disregarding lagged temporal interactions might imply missing potentially richer dynamics. Here, we introduce an approach to solve whole-brain voxel-wise auto-regressive models, utilizing the concept of regularized or sparse regression to avoid the problem of under-constrained number of samples. We present results demonstrating that our approach is computationally feasible, and validate it by showing that, on real data, the multi-variate auto-regressive model significantly outperforms uni- and bi-variate modeling at predicting future brain states up to 4 fMRI temporal resolutions (TRs).

2. METHODS

We first describe the sparse full-brain auto-regressive modeling followed by two baseline methods, namely univariate autoregressive modeling and bivariate autoregressive modeling, which are used for comparisons.

Full-brain Multivariate Autoregressive Modeling (FARM): The FARM model is designed to discover the complex spatio-temporal interactions among the voxels in the brain. It models the future activity of a voxel as a linear combination of the past activities of all the other voxels. Formally, $x_i(t) = \sum_{\tau=1}^{k} \sum_{j=1}^{n} a_{ij}(\tau) x_j(t-\tau)$, where $x_i(t)$ represents the activity of voxel i at time t and $a_{ij}(\tau)$ represent the model parameters to be estimated from the data. The maximum likelihood estimate of model parameters (under the standard assumption of iid Gaussian probability distributions) is given by solving the following program: Minimize: $\sum_{t=k+1}^{T} (x_i(t) - \sum_{\tau=1}^{k} \sum_{j=1}^{n} a_{ij}(\tau) x_j(t-\tau))^2$.

A reliable solution the above system of equations requires a large number of temporal observations ($T \gg nk$). Unfortunately, for fMRI data, T is in the range of a few hundreds, while n is in the range of tens of thousands. Thus, it is impossible to solve the above program even for model order 1.

In reality, the neuronal activity in the brain propagates through the synaptic connections. These connections are sparse (i.e., every neuron is not directly connected to every other neuron) and remain mostly unchanged over short periods of time. Therefore, it is reasonable to assume that the spatio-temporal interactions are sparse and only a small number of the coefficients $a_{ij}(\tau)$ are non-zero. With the above assumption, it becomes possible to solve the model identification problem using sparse regression techniques. In particular, the "ℓ_1 regularized" lasso regression [21] has been particularly found useful in finding sparse solutions. We add an ℓ_1 regularization term to the model identification and solve the following: Minimize: $\sum_{t=k+1}^{T} (x_i(t) - \sum_{\tau=1}^{k} \sum_{j=1}^{n} a_{ij}(\tau) x_j(t-\tau))^2 + \lambda \sum_{\tau=1}^{T} \sum_{i=1}^{n} |a_{ij}(\tau)|$, where $|v|$ represents the absolute value of v. Our earlier work [13] demonstrated the computational feasibility of solving the above system on the BlueGene supercomputer using a LARS-based implementation of lasso [7]. Using realistic simulations, we also demonstrated that our approach recovers the true model satisfactorily [5]. An earlier work [23] also

studied sparse autoregressive modeling of fMRI but resorted to data aggregation for evaluations, leading to results that were fundamentally different from our results.

Univariate Autoregressive Modeling: The brain activity in a voxel is modeled as a linear combination its past values. For an order k univariate autoregressive model, the brain activity at the i^{th} voxel is modeled as $x_i(t) = \sum_{\tau=1}^{k} a_i(\tau)x_i(t-\tau)$. The maximum likelihood estimate (under the standard assumptions) of the model parameters $a_i(\tau)$ are determined by solving the following least squares program for all i: Minimize: $\sum_{t=k+1}^{T}(x_i(t) - \sum_{\tau=1}^{k} a_i(\tau)x_i(t-\tau))^2$. Standard closed form solutions [4] are available to solve the problem.

Bivariate Autoregressive Modeling: The bivariate model considers the experimental condition, in addition to the past activity at a voxel in order to predict its future activity. According to an order k bivariate autoregressive model the brain activity at voxel i is modeled as $x_i(t) = \sum_{\tau=1}^{k}(a_i(\tau)x_i(t-\tau) + b_i(\tau)s(t-\tau))$. As before, the model parameters $a_i(\tau)$ and $b_i(\tau)$ may be determined by solving the following least squares problem: Minimize: $\sum_{t=k+1}^{T}(x_i(t) - (\sum_{\tau=1}^{k} a_i(\tau)x_i(t-\tau) + b_i(\tau)s(t-\tau)))^2$.

2.1. Model metrics

We evaluate the above models using the metrics of *prediction accuracy* and *prediction power* as defined below.

Prediction accuracy. The models learned may be used to predict the future activity of voxels. As a first step we restrict our attention to autoregressive models of order one. Let $x'_i(t)$ represent the activity of voxel i at time t in the test data.

For the univariate autoregressive model, the one step prediction of voxel i's activity is given by $\hat{x}_i^1(t) = a_i x'_i(t-1)$. The k-step prediction of voxel i is given by $\hat{x}_i^k(t) = a_i \hat{x}_i^{k-1}(t-1)$. For the bivariate autoregressive model the one step prediction is given by $\hat{x}_i^1(t) = a_i x'_i(t-1) + b_i s'(t-1)$. The k-step prediction of voxel i is given by $\hat{x}_i^k(t) = a_i \hat{x}_i^{k-1}(t-1) + b_i s'(t-1)$. Similarly, the one step prediction for multivariate autoregressive model is given by $\hat{x}_i^1(t) = \sum_{j=1}^{n} a_{ij}(1)x'_j(t-1)$. The k-step prediction of voxel i is given by $\hat{x}_i^k(t) = \sum_{j=1}^{n} a_{ij}(1)\hat{x}_j^{k-1}(t-1)$.

The k-step *prediction accuracy* of voxel i is defined as $\rho_i(k) = 1 - [\sum_{t=k+1}^{T'}(\hat{x}_i^k(t) - x'_i(t))^2]/[\sum_{t=k+1}^{T'} x'_i(t)^2]$, where $\hat{x}_i(t)$ represents the predicted activity in voxel i at time t. For convenience, we label the prediction accuracy in the training data as the 0-step prediction accuracy.

Prediction power. The prediction power of a voxel j (π_j) in the FARM model, represents its ability to predict the future activity of other voxels. Formally, $\pi_j = \sum_{\tau=1}^{k}\sum_{i=1}^{n}|a_{ij}(\tau)|$. Intuitively, the prediction power of a voxel represents the total influence of the voxel in predicting the future activity of other voxels. The properties of the prediction power maps are reported in our earlier work [14]. In particular, we find that the prediction power maps, which follow a heavy-tailed distribution, are consistent within as well as across subjects and provide improved functional localization as compared to the GLM activation maps.

2.2. fMRI data acquisition and processing

The data used in this paper (and its pre-processing) is identical to [8]. Six right-handed subjects were scanned three times (a total of 18 sessions) in a block-based self-paced finger tapping paradigm. Each session comprised 20 blocks, each block with 10 volumes of finger tapping followed by 10 volumes of rest (with a TR of 2.5 seconds).

The first 350 volumes were used for training; the remainder 50 volumes were used for testing. For choosing the parameter λ in the FARM model, a suitable range was obtained using simulations on data of comparable sizes (see [5]). Six equally spaced values of $1/\lambda$ were selected from this range to build a model of one of the sessions. The prediction accuracies of all the voxels in the test data were computed for this session. The value of λ giving the best average 1-step full-brain prediction accuracy was used to build the FARM model for the remainder of the sessions. A separate cross-validation for each session could not be carried out due to the large computational needs of the method.

Now, for all the sessions, the k-step prediction accuracy ($k \in [0 \ldots 4]$) was computed for all the voxels in the brain using each of the three methods. The prediction accuracy maps thus obtained were aligned to the standard MNI brain atlas using FSL [1].

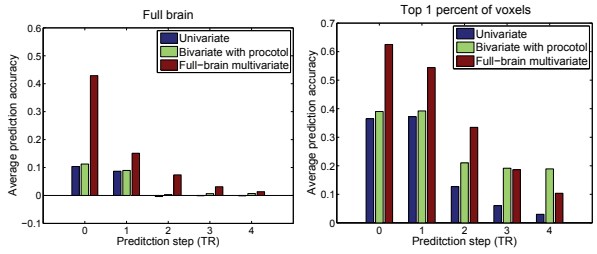

Fig. 1. Mean prediction accuracy as a function of future TRs (1-4); the first point (0) corresponds to the training accuracy. **Left**: for all the voxels in the brain. **Right**: for the top 1% of voxels.

3. RESULTS

The full-brain auto-regressive model (FARM), shows a significantly higher prediction accuracy than the uni-variate and the bi-variate models. Fig. 1 (left panel) shows the mean prediction accuracy of the entire brain (averaged over the 18 sessions), for the training set (labeled as TR=0) and for the testing set at TRs from 1 to 4. Fig. 1 (right panel), shows the mean prediction accuracy for the top 1% of most predictable voxels,

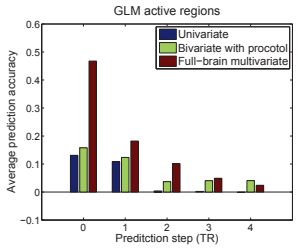

Fig. 2. Mean prediction accuracy over the GLM-identified regions.

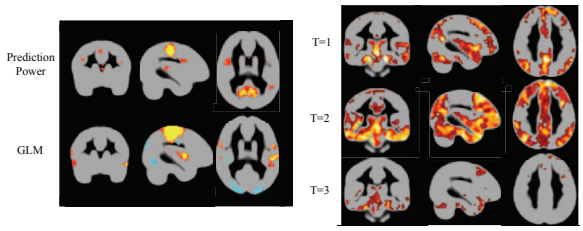

Fig. 3. **Left**: Clusters of high prediction power and GLM active regions. **Right**: t-maps of regions with significantly higher predictability in the FARM model over the bi-variate model (red-yellow) and vice versa (blue-light blue).

which include those directly regulated by the task. Even for these voxels, the FARM model, which does not make use of the task information, consistently outperforms the bivariate model which uses the task information for prediction. The comparison between FARM model and uni/bivariate model in relation to the task response can be further analyzed by concentrating specifically on the areas identified as active by the GLM analysis. Fig. 2 shows that the mean prediction accuracy of the FARM model is better than the univariate and bivariate models even in the active regions.

One may wonder if the FARM model gives better prediction accuracy throughout the brain or in specific brain regions. Are there regions in the brain where bivariate model gives better prediction accuracy than the FARM model? To address these questions, we carried out normalized t-test as outlined below.

For a comparison of two models, the prediction accuracy map obtained using one model was subtracted from the corresponding map using the other model. The resulting map was transformed to a normal space using the Fisher transform. One sided t-tests using all the 18 sessions, were carried out for each voxel in the brain. The p-values thus obtained were corrected for the multiple comparison problem using the false discovery rate (FDR) correction method [15].

Figure 3 (right) shows the regions where FARM gives better prediction accuracy than the bivariate model. Though, the voxels where FARM model gives better prediction are distributed throughout the brain, the voxels in the default networks [22] have particularly better prediction accuracy. Con-

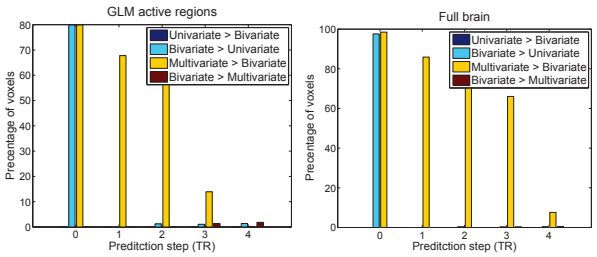

Fig. 4. Fraction of voxels that can be predicted more accurately using FARM; x-axis as in Fig. 1. **Left**: GLM-identified regions. **Right**: Full-brain.

versely, there are very few voxels where bivariate model gives better prediction than the FARM model.

Fig. 4 shows a pairwise comparison of the three models using the number of voxels in which one method gives better prediction accuracy than the other, at a FDR corrected q-value of 0.05. The left panel shows a comparison for the entire brain and the right panel shows the comparison for GLM active regions. The FARM model gives a better prediction accuracy than the bivariate model in $80-95\%$ of the voxels in the entire brain and in $60-80\%$ of voxels in GLM active regions.

The above results demonstrate the importance of studying the interactions among the brain regions. Even in the active regions, complex interactions among the voxels are present which are discovered by the FARM model leading to a better prediction accuracy even without using any task information. This is probably the reason why FARM model gives much localized maps as compared the GLM activity maps (see [14] and Figure 3 (left)).

The GLM identifies all the voxels in a region as active whereas FARM selects a small subset of voxels as "drivers of activity" (i.e. voxels with high prediction power) and models the activities in other voxels as being "driven" by these voxels. Figure 3 (left) shows a comparison between the regions of consistently high prediction power (i.e., the drivers) and the consistent GLM activations, using 5% FDR corrected binomial tests on the eighteen sessions. The voxels of high prediction power are present in motor regions as well as the regions (precuneous/PCC) corresponding to the default mode networks [22]. The GLM activation maps show a larger "blob of activity" in the motor region, but no activity in the default mode networks.

4. DISCUSSION AND CONCLUSIONS

In our earlier works we showed that it feasible to build a full-brain autoregressive model (FARM) of the brain from fMRI data [5, 13]. The model gives new biological insights [14] and has several desirable properties such as improved functional resolution and good inter-subject and intra-subject consistency. It is possible to visualize the interactions among the

brain regions in new ways [12]. In this paper, we show that even without any task information, the FARM model gives significantly improved prediction performance on unseen data as compared to the univariate and the bivariate models which makes use of the task information for model building as well as prediction. The voxels with better prediction are distributed throughout the brain, including the regions identified as active by the GLM. We hypothesize that the brain activity as measured by fMRI, comprises, in addition to the response to the task conditions, complex spatio-temporal interactions among brain regions. A part of these interactions are uncovered by the FARM model, leading to better predictions of future brain activity up to several TRs.

In order to verify the above hypothesis, a better understanding of the mechanisms involved in the neuro-vascular coupling is needed. The interactions between neuronal activity and the blood flow that gives rise to the fMRI (BOLD) signal are not very well understood till date [19]. It is well-known that the "hemodynamic response function" which maps the experimental condition to fMRI responses is highly variable across subjects, across brain regions of the same subject, across different experiments and even across different trials of the same experiment [3, 17]. The intra-subject variability is much less than the inter-subject variability. The inter-subject variability in time-to-peak of hemodynamic response function was reported to have a standard deviation of 1.1 seconds [3]. The impact of such a variability on an experiment with TR 2.5 seconds, though expected to be small, needs to be quantified. Moreover, it is not known if the observed variability in the hemodynamic response functions (across different brain regions of the same subject) originate due to differences in the neuro-vascular coupling or due to neuronal activity differences [20]. More work is needed to ascertain the truth of the above hypothesis.

5. REFERENCES

[1] *FSL Release 3.3*, http://www.fmrib.ox.ac.uk/fsl (2006).

[2] A. Arieli, A. Sterkin, A. Grinvald and A. Aertsen. Dynamics of ongoing activity: Explanation of the large variability in evoked cortical responses. *Science*, 273(5283):1868–1871, 1996.

[3] G. K. Aguirre, E. Zarahn, and M. D'Esposito. The variability of human, BOLD hemodynamic responses. *NeuroImage*, 8(4):360 – 369, 1998.

[4] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel. *Time Series Analysis: Forecasting and Control (Wiley Series in Probability and Statistics)*. Wiley, 4^{th} edition, June 2008.

[5] G. A. Cecchi, R. Garg, and A. R. Rao. Inferring brain dynamics using Granger causality on fMRI data. In *Proceedings of ISBI'08*, pages 604–607, 2008.

[6] G. Edelman and G. Tononi, editors. *A Universe of Consciousness How Matter Becomes Imagination*. Basic Books, 2001.

[7] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *Annals of Statistics*, 32(1):407–499, 2004.

[8] V. Eguiluz et al. Scale-free functional brain networks. *Physical Review Letters*, 94(018102), 2005.

[9] M. Fox, A. Snyder, J. Vincent, M. Corbetta, D. C. V. Essen, and M. Raichle. The human brain is intrinsically organized into dynamic, anti-correlated functional networks. *Proc. Natl. Acad. Sci. USA*, 102:9673–9678, 2005.

[10] K. Friston and C. Buchel. Attentional modulation of effective connectivity from V2 to V5/MT in humans. *Proc. Natl. Acad. Sci. USA*, 97(13):7591–7596, 2000.

[11] K. Friston, L. Harrison, and W. Penny. Dynamic causal modelling. *NeuroImage*, 19:1273–1302, 2003.

[12] R. Garg, G. A. Cecchi, and A. R. Rao. Full-brain autoregressive modeling (FARM) using fMRI. Preprint.

[13] R. Garg, G. A. Cecchi, and A. R. Rao. Applications of high-performance computing to functional magnetic resonance imaging (fMRI) data. In A. R. Rao and G. A. Cecchi, editors, *High-throughput Image Reconstruction and Analysis*, chapter 12, pages 263–282. Artech House Publishers, 2009.

[14] R. Garg, G. A. Cecchi, and A. R. Rao. Characteristics of voxel prediction power in full-brain Granger causality analysis of fMRI data. In *Proceedings of SPIE Medical Imaging*, 2011.

[15] C. R. Genovese, N. A. Lazar, and T. Nichols. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage*, 15(4):870 – 878, 2002.

[16] R. Goebel, A. Roebroeck, D.-S. Kim, and E. Formisano. Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. *Magnetic Resonance Imaging*, 21:1251–1261, 2003.

[17] D. A. Handwerker, J. M. Ollinger, and M. D'Esposito. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *NeuroImage*, 21(4):1639 – 1651, 2004.

[18] C. Koch and J. Davis, editors. *Large-scale neuronal theories of the brain*, chapter 6: Perception as an oneiric-like state modulated by the senses. MIT Press, 1995.

[19] M. T. Lippert, T. Steudel, F. Ohl, N. K. Logothetis, and C. Kayser. Coupling of neural activity and fMRI-BOLD in the motion area MT. *Magnetic resonance imaging*, 28(8):1087–1094, Oct. 2010.

[20] Y. Lu, A. P. Bagshaw, C. Grova, E. Kobayashi, F. Dubeau, and J. Gotman. Using voxel-specific hemodynamic response function in EEG-fMRI data analysis. *NeuroImage*, 32(1):238–247, 2006.

[21] R. Tibshirani. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, 58(1):267–288, 1996.

[22] M. Raichle, A. MacLeod, A. Snyder, W. Powers, D. Gusnard, and G. Shulman. A default mode of brain function. *Proc. Natl. Acad. Sci. USA*, 98(2):676682, 2001.

[23] P. A. Valdes-Sosa, J. M. Sanchez-Bornot, A. Lage-Castellanos, M. Vega-Hernandez, J. Bosch-Bayard, L. Melie-Garcia, and E. Canales-Rodriguez. Estimating brain functional connectivity with sparse multivariate autoregression. *Philos Trans R Soc Lond B Biol Sci.*, 360(1457):969–81, 2005.