Published in final edited form as:

Conf Proc IEEE Eng Med Biol Soc. 2018 July ; 2018: 275–278. doi:10.1109/EMBC.2018.8512200.

Biophysically interpretable recurrent neural network for functional magnetic resonance imaging analysis and sparsity based causal architecture discovery

Yuan Wang,

Department of Electrical and Computer Engineering, New York University, Brooklyn, NY 11201 USA

Yao Wang [Member, IEEE],

Department of Electrical and Computer Engineering, New York University, Brooklyn, NY 11201 USA

Yvonne W. Lui

Department of Radiology, New York University, New York, NY 10003 USA

Abstract

Recent efforts use state-of-the-art Recurrent Neural Networks (RNN) to gain insight into neuroscience. A limitation of these works is that the used generic RNNs lack biophysical meaning, making the interpretation of the results in a neuroscience context difficult. In this paper, we propose a biophysically interpretable RNN built on the Dynamic Causal Modelling (DCM). DCM is an advanced nonlinear generative model typically used to test hypotheses of brain causal architectures and associated effective connectivities. We show that DCM can be cast faithfully as a special form of a new generalized RNN. In the resulting DCM-RNN, the hidden states are neural activity, blood flow, blood volume, and deoxyhemoglobin content and the parameters are biological quantities such as effective connectivity, oxygen extraction fraction and vessel wall elasticity. DCM-RNN is a versatile tool for neuroscience with great potential especially when combined with deep learning networks. In this study, we explore sparsity-based causal architecture discovery with DCM-RNN. In the experiments, we demonstrate that DCM-RNN equipped with l^1 connectivity regulation is more robust to noise and more powerful at discovering sparse architectures than classic DCM with l^2 connectivity regulation.

I. Introduction

Deep learning, which has shown considerable potential in the last decades as a powerful tool for data analysis, can help explore new possibilities of understanding of brain behavior [1]. Recurrent Neural Network (RNN) is particularly suitable for temporal signals, because it models temporal correlation among data explicitly with the recurrent structure. Many efforts have been made to analyze functional Magnetic Resonance Imaging (fMRI) signals with

state-of-the-art RNNs [1]. However, the used generic RNNs lack biophysical meaning, making the interpretation of results in a neuroscience context difficult.

Dynamic Causal Modelling (DCM) [2] is a highly nonlinear generative model and is considered by many to be the most biologically plausible as well as the most technically advanced fMRI modeling method [3]. A causal architecture, along with associated effective connectivity, indicates how the neural activities interact with each other between distributed brain regions and how input stimuli may alter the pattern. Traditionally, DCM is used to compare pre-designed hypotheses of the causal architecture. If no pre-designed hypotheses are available, DCM searches the hypothesis space based on a centralized scheme [4], which heavily relies on the accuracy of the central anchor.

In this study, we propose a biophysically interpretable RNN based on DCM and explore an alternative data-driven way of causal architecture discovery based on sparsity.

II. Background

DCM can be expressed compactly as

$$\dot{x}(t) = f(x(t), u(t); \theta) + w(t; \theta) \quad (1)$$

$$y(t) = g(x(t), u(t); \theta) + z(t; \theta) \quad (2)$$

where $u(t) \in \mathbb{R}^M$ is the experimental stimuli or input. $x(t) \in \mathbb{R}^{N_b}$ is the neural activity and $y(t) \in \mathbb{R}^{N_b}$ is the modeled fMRI signal. $w(t) \in \mathbb{R}^{N_b}$ is random fluctuation in the neural activity space and $z(t) \in \mathbb{R}^{N_b}$ is observation noise. θ is the whole set of DCM parameters. The superscript M and N_b indicate the number of stimuli and the number of brain regions involved in a DCM study respectively. Temporal derivative is denoted by a dot above a variable. f models the evolution of the neural activity and g maps the neural activity to the fMRI signals.

In the original DCM [2], w is absent, and the neural evolution function takes a bilinear form as the majority of DCMs do:

$$\dot{x}(t) = Ax(t) + \sum_{m=1}^M u_m(t)B_m x(t) + Cu(t) \quad (3)$$

where $A \in \mathbb{R}^{N_b \times N_b}$, $\{B_m \in \mathbb{R}^{N_b \times N_b} | m \in \{1, 2, \dots, M\}\}$, $C \in \mathbb{R}^{N_b \times M}$ are the effective connectivity to be estimated. Although the neural activity is coupled cross brain regions, the hemodynamics of one region does not depend on other regions, given the neural activity in that region. We drop the region index n of the DCM states and parameters in (4)-(9) for

simplicity and clarity. First, x evokes vasodilatory signal s , which modulates the blood inflow f to a brain region

$$\dot{s}(t) = x(t) - \kappa s(t) - \gamma(f(t) - 1) \quad (4)$$

$$\dot{f}(t) = s(t) \quad (5)$$

where κ is signal decay and γ is feedback regulation. The blood volume change in a brain region is determined by blood inflow and outflow

$$\dot{v}(t) = \frac{1}{\tau} f(t) - \frac{1}{\tau} f_{out}(t) \quad (6)$$

where v is the volume of blood in a brain region, f_{out} is blood outflow, the volume of blood leaving the brain region, and τ is mean transit time of blood. The blood outflow is a function of the blood volume based on the balloon model for vasculature:

$$f_{out}(t) = v(t)^{\frac{1}{\alpha}} \quad (7)$$

where α is vessel stiffness. Under the assumption that oxygen extraction is tightly coupled to blood flow

$$\dot{q}(t) = \frac{1}{\tau} \left[\frac{1 - (1 - E_0)^{\frac{1}{f(t)}}}{E_0} f(t) - \frac{q(t)}{v(t)} v(t)^{\frac{1}{\alpha}} \right] \quad (8)$$

where q is deoxyhemoglobin content, E_0 is the oxygen extraction fraction at rest. The modeled BOLD signal is a function of v and q

$$y(t) \approx V_0 \left[k_1 (1 - q(t)) + k_2 \left(1 - \frac{q(t)}{v(t)} \right) + k_3 (1 - v(t)) \right] + z(t) \quad (9)$$

$$k_1 = 4.3 \vartheta_0 E_0 TE, k_2 = \varepsilon r_0 E_0 TE, k_3 = 1 - \varepsilon$$

where V_0 is the resting venous blood volume fraction, ϑ_0 is the frequency offset at the outer surface of the magnetized vessel for fully deoxygenated blood, E_0 is the oxygen extraction fraction at rest, TE is the Echo Time, r_0 is the slope of the relation between the intravascular relaxation rate and oxygen saturation, and ε is the ratio of intra- and extravascular signal. z is zero mean Gaussian noise. A large covariance matrix Σ is used to characterize noise

correlation cross time and brain regions, which takes a restrict form $\Sigma = \Sigma_i (\exp(\lambda_i))^{-1} Q_i$ where Q_i are predefined covariance patterns. As a summary, the whole DCM parameter set $\theta = \{\theta^c, \theta^h, \theta^l\}$ includes connectivity parameters $\theta^c = \{A, B_m, C \mid m \in \{1, 2, \dots, M\}\}$, hemodynamic parameters $\theta^h = \{\alpha, \varepsilon, \tau, \kappa, \gamma, E_0, V_0, \theta_0, r_0\}$ and hyper parameters $\theta^l = \{\lambda\}$.

In DCM, the causal architecture, or the supports of θ^c , is a key different between different hypotheses, or models. Since parameter inference for all possible supports can be prohibitively expensive, DCM searches the model space by trimming out connections from a fully supported model and evaluating evidence of reduced models with a simple operation without direct interaction with data [4]. This method heavily relies on the estimation accuracy of the fully supported model and assumes the posterior Gaussian distribution of the parameters in the fully supported model can capture all the information in the data.

III. Method

This study is performed under institutional review board (IRB) compliance for human subjects research.

A. Convert DCM into DCM-RNN

The classic RNN models the relationship between its input and output as

$$h_t = f^h(W^{hx}x_t + W^{hh}h_{t-1} + b^h) \quad (10)$$

$$y_t = f^y(W^{yh}h_t + b^y) \quad (11)$$

where x is the input, h is the hidden state, and y_t is the output. W and b are weighting matrix and bias, the tunable parameters. Subscript t indicates time and superscripts are used to differentiate parameters. f^h and f^y are simple nonlinear functions. We generalize the classic RNN with more nonlinearities to accommodate the complexity of DCM:

$$h_t = f^h(W^h\phi^h(x_t, h_{t-1}; \xi^h) + b^h) \quad (12)$$

$$y_t = f^y(W^y\phi^y(h_t; \xi^y) + b^y) \quad (13)$$

where ϕ^h and ϕ^y are the added nonlinear functions, which are parameterized by ξ^h and ξ^y .

We use Euler's method to adapt DCM for discrete fMRI signals and the discrete nature of RNN:

$$\dot{a}_t \approx \frac{a_{t+1} - a_t}{\Delta t} \text{ for } a \in \{x, s, f, v, q\} \quad (14)$$

where Δt is the time interval between adjacent time points. Taking $\dot{x}_t$ as an example.

Substituting (14) into the neural evolution equation, equation (3) becomes

$$x_{t+1} \approx (A\Delta t + I)x_t + \sum_{m=1}^M \Delta t u_{mt} B_m x_t + \Delta t C u_t \quad (15)$$

It can be organized and rewritten in the form of the generalized RNN

$$\begin{aligned} x_{t+1} & \quad (16) \\ & \approx [(\Delta t A + I) \quad \Delta t B \quad \Delta t C] \phi^x \begin{pmatrix} x_t \\ u_t \end{pmatrix} \\ & \equiv [W^{xx} W^{xu} W^{xu}] \phi^x \begin{pmatrix} x_t \\ u_t \end{pmatrix} \end{aligned}$$

where $B = [B_1, B_2, \dots, B_M]$ is a concatenation of the B_m matrices and

$$\phi^x \begin{pmatrix} x_t \\ u_t \end{pmatrix} = \begin{bmatrix} x_t^T & (u_t \otimes x_t)^T & u_t^T \end{bmatrix}^T \quad (17)$$

where $\otimes$ is Kronecker product. Applying similar approximations to other parts of DCM, one obtains the whole DCM-RNN as in Fig. 4. Equation (16) is at the bottom of it.

B. MAP Estimation of DCM-RNN parameters

Given the observed fMRI signals over several brain regions, we can estimate the parameters θ of DCM-RNN by maximizing the *a posteriori* probability $p(\theta/y) \propto p(y/\theta) p(\theta)$. We make following assumptions about the distributions: 1) the observation noise z is zero mean Gaussian, *i.e.* $p(y|\theta) = N(g_{1:T}(\theta^c, \theta^h), \Sigma(\theta^\lambda))$, where $g_{1:T}$ is the modeled fMRI for all time points and regions, and T is the number of time point. 2) θ^c , θ^h , and θ^λ are independent. 3) θ^c follows Laplace distribution, $p(\theta^c) = \prod_{\theta \in \theta^c} \frac{1}{2b} \exp\left(-\frac{|\theta|}{b}\right)$. 4) θ^h follows Gaussian distribution, $p(\theta^h) = N(\mu^h, \Sigma^h)$. 5) θ^λ follows Gaussian distribution, $p(\theta^\lambda) = N(\mu^\lambda, \Sigma^\lambda)$. The loss function is $-\ln p(\theta/y)$ after removing constant terms:

$$l(\theta) = \frac{1}{2}(\epsilon^y)^T \Sigma^{-1} \epsilon^y + \frac{1}{2} \ln(|\Sigma|) + \frac{1}{b} |\theta^c|_1 \quad (18)$$

$$+ \frac{1}{2}(\epsilon^h)^T (\Sigma^h)^{-1} \epsilon^h + \frac{1}{2}(\epsilon^\lambda)^T (\Sigma^\lambda)^{-1} \epsilon^\lambda$$

where $\epsilon^y = y - g_{1:T}$, $\epsilon^h = \theta^h - \mu^h$, and $\epsilon^\lambda = \theta^\lambda - \mu^\lambda$. A key difference between the loss functions of DCM-RNN and DCM is the assumed prior distribution for θ^c . In DCM, θ^c is assumed to be Gaussian. By assuming Laplacian distribution for θ^c , we effectively enforce limited support in the estimated connectivity parameters. We estimate $\theta = \{\theta^c, \theta^h, \theta^\lambda\}$ by minimizing $l(\theta)$ using the Truncated Backpropagation Through Time method that has been developed for training RNN.

IV. Experiments and results

A. Experiment settings

Effective connectivity estimation is conducted by DCMRNN and DCM, separately with simulated fMRI and *in vivo* fMRI. We implement DCM-RNN on Tensorflow and use SPM 12 for classic DCM, which will be referred to as DCM-SPM. We use *spm_int_J* option in DCM-SPM for discretization, because it uses an explicit Jacobian-based update rule that is infallible and preserves nonlinearities in DCM, although it runs slower than the default *spm_int*. DCM-SPM uses variational Bayes for parameter estimation.

For the experiment with simulated data, to avoid potential bias, fMRI data are generated with both models separately and estimations are done with their own simulated data. The input is randomly generated, θ^c is set as in Fig. 1, and θ^h is set to their prior mean in DCM-SPM. Following the default setting in DCM-SPM, the noise covariance is set such that the noise is temporally and spatially uncorrelated but can have different strength in different regions denoted by $\exp(\lambda_i)$. During data simulation, we use the same λ for all brain regions and λ is chosen so that the overall signal to noise ratio (SNR) is 3, where SNR is defined as the l_2 norm of the fMRI signal divided by the noise standard deviation.

For the experiment with *in vivo* fMRI data, we used the ‘Attention to Visual Motion fMRI data set’ on the SPM website. The photic stimulation experiment involves three brain regions: V1, the primary visual cortex; V5, the middle temporal visual area responsible for motion perception; and superior parietal cortex (SPC).

For parameter estimation, the simulated or *in vivo* data are first up sampled to meet the resolution requirement for parameter estimation. All parameters $\{\theta^c, \theta^h, \theta^\lambda\}$ are updated jointly during estimation. The prior parameters $\{\mu^h, \mu^\lambda, b, \Sigma_h, \Sigma_\lambda, Q_i\}$ are fixed.

Relative root mean square error (rRMSE) is used to measure estimation accuracy which is defined as the l_2 norm of the error divided by the l_2 norm of the truth. All the codes used in the experiments are available online at https://github.com/YuanWangOnward/DCM_RNN.

B. Results and discussions

First, we compare effective connectivity estimation results from simulated data under four conditions. The results are summarized in TABLE I and Fig. 2. Under the first condition, fMRI is noiseless and the supports of θ^c are the true ones. Both DCM-RNN and DCM-SPM obtain fairly good estimation. Under the second condition, A is fully supported. Both DCM-RNN and DCM-SPM are able to discover the correct causal architecture and obtain fairly good estimation accuracy. When noise is added to the fMRI data under the third condition, the θ^c estimation accuracy of DCM-SPM degrades significantly, and some non-trivial false connectivity values appear which means the correct causal architecture is not identified. While the θ^c estimation accuracy of DCM-RNN only decreases mildly and the causal architecture remains correct after thresholding out a trivial false connectivity. Under the fourth condition, both A and B_3 are fully supported. It is the least informed configuration, which challenges the estimation power of the models. In this extreme case, DCM-RNN can still get reasonable estimation accuracy, and, notably, one can still get the correct causal architecture by applying a proper thresholding. These results suggest that DCM-RNN with sparsity prior is more robust to noise and has more power at discovering parsimonious causal architecture from noisy fMRI data.

Second, we compare effective connectivity estimation results from *in vivo* fMRI data. 12 entries are supported in θ^c . DCM-RNN is able to discovery a causal architecture with 7 non-zero connectivities and achieves a reproduction rRMSE of 59.0% for the fMRI signal. DCM-SPM cannot find a sparse causal architecture achieving a reproduction rRMSE of 53.9%. It suggests that DCM-RNN with sparsity prior has a great promise in discovering parsimonious causal architectures.

Acknowledgments

This work is supported in part by grant funding from the National Institutes of Health (NIH): R01 NS039135–11 and R21 NS090349, National Institute for Neurological Disorders and Stroke (NINDS). This work is also performed under the rubric of the Center for Advanced Imaging Innovation and Research (CAI2R, www.cai2r.net), a NIBIB Biomedical Technology Resource Center (NIH P41 EB017183).

References

- [1]. Gonzalez RT, Riascos JA, and Barone DAC, “How Artificial Intelligence is Supporting Neuroscience Research: A Discussion About Foundations, Methods and Applications,” in *Computational Neuroscience*, 2017, pp. 63–77.
- [2]. Friston KJ, Harrison L, and Penny W, “Dynamic causal modelling,” *Neuroimage*, vol. 19, no. 4, pp. 1273–1302, 2003. [PubMed: 12948688]
- [3]. Smith SM, “The future of FMRI connectivity,” *Neuroimage*, vol. 62, no. 2, pp. 1257–1266, 2012. [PubMed: 22248579]
- [4]. Friston KJ, Litvak V, Oswal A, Razi A, Stephan KE, Van Wijk BCM, Ziegler G, and Zeidman P, “Bayesian model reduction and empirical Bayes for group (DCM) studies,” *Neuroimage*, vol. 128, pp. 413–431, 2016. [PubMed: 26569570]
- [5]. Friston KJ, “Bayesian Estimation of Dynamical Systems: An Application to fMRI,” *Neuroimage*, vol. 16, no. 2, pp. 513–530, 2002. [PubMed: 12030834]

Fig. 1.

Causal architecture used to simulate fMRI data. N indicates brain region/node and u is input stimulus.

Fig. 2. Effective connectivity estimation results with simulated fMRI under different conditions. (a) Noiseless fMRI with the correct supports. (b) Noiseless fMRI with extended support of A matrix. (c) Noisy fMRI with extended support of A matrix. (d) Noisy fMRI with extended supports of A and B_3 matrices.

Fig. 3.
Effective connectivity estimation results with *in vivo* fMRI.

Fig. 4.
An overview of DCM-RNN

TABLE I

Estimation Results with Simulated data

Condition	If noisy fMRI	Extended support	Reproduction rRMSE		θ rRMSE	
			<i>DCM-RNN</i>	<i>DCM-SPM</i>	<i>DCM-RNN</i>	<i>DCM-SPM</i>
1	No	None	1.80%	1.02%	2.92%	4.44%
2	No	A	1.78%	1.02%	3.81%	4.50%
3	Yes	A	16.4%	15.9%	7.56%	35.9%
4	Yes	A & B_3	16.3%	15.8%	8.43%	51.9%