

A Bayesian Variable Selection Model for the Clustering of Time Courses in fMRI data

Un modello Bayesiano di selezione di variabili per il clustering di serie temporali nell'analisi di dati fMRI.

Michele Guindani, Linlin Zhang, Francesco Versace and Marina Vannucci

Abstract We describe a wavelet-based Bayesian nonparametric regression model for the analysis of functional magnetic resonance imaging (fMRI) data. Our framework detects regions of the brain exhibiting neuronal activity in response to a stimulus and, simultaneously, infers the association, or clustering, of spatially remote voxels exhibiting similar temporal characteristics. We use spike and slab, Markov Random Field and Dirichlet Process priors to detect brain activations, account for the complex spatial correlation structure of the brain as well as cluster correlated time courses. The model performance is illustrated on a set of fMRI synthetic data.

Abstract Si presenta un modello di regressione Bayesiano basato su trasformazioni wavelet per l'analisi di immagini di risonanza magnetica funzionale (fMRI). Il modello identifica regioni del cervello che esibiscono attività neuronale in risposta ad uno stimolo e, allo stesso tempo, inferisce l'associazione, o clustering, di voxel spazialmente lontani che presentano simili caratteristiche temporali. Si impiegano priors di tipo spike e slab, Markov Random Field e processi di Dirichlet per identificare l'attività cerebrale, modellare la complessa correlazione spaziale del cervello e clusterizzare serie temporali correlate. Il comportamento del modello proposto viene valutato su simulazioni di dati fMRI.

Key words: Bayesian Nonparametrics, Bayesian Variable Selection, Markov Random field prior, Dirichlet Process Prior, fMRI data.

Michele Guindani
UT MD Anderson Cancer Center e-mail: mguindani@mdanderson.org

Linlin Zhang
Rice University, e-mail: lz17@rice.edu

Francesco Versace
UT MD Anderson Cancer Center e-mail: fversace@mdanderson.org

Marina Vannucci
Rice University e-mail: marina@rice.edu

1 Introduction

The complex temporal and spatial dynamics captured by fMRI experiments demand the use of advanced statistical methods. We provide a joint analytical framework that allows us to detect regions of the brain exhibiting neuronal activity in response to a stimulus and, simultaneously, infer the association, or clustering, of spatially remote voxels that exhibit fMRI time series with similar characteristics.

In fMRI experiments, the blood oxygenation level dependent (BOLD) signal contrast is used to measure neuronal activation in response to an input stimulus. In the formulation we adopt, the stimulus pattern is convolved with a Poisson hemodynamic response function (HRF), with a voxel-dependent shape parameter, that characterizes the delay between the onset of the stimulus and the arrival of blood to the activated brain regions. In addition, we assume a long memory process to model the correlated errors and take into account the time course structure of voxel responses, and then exploit the whitening properties of discrete wavelet transforms.

Our model allows us to cluster the time course responses of distant brain regions via a Dirichlet process (DP) prior. The induced spatio-temporal clustering can be viewed as an aspect of “functional” connectivity, as it naturally captures statistical dependencies among remote neurophysiological events (Friston, 2011). In addition, we detect activation in response to a stimulus by using mixture priors with a spike at zero on the coefficients of the regression model characterizing the association between response and stimulus. The selection of activated voxels takes into account the complex spatial correlation structure of the brain through a Markov Random Field (MRF) prior. For inference, we use Markov Chain Monte Carlo (MCMC) sampling techniques that combine Metropolis-Hastings schemes employed in Bayesian variable selection with sampling algorithms for nonparametric DP models.

Our inferential strategy results in the detection of activated regions of the brain and clustering of similar time courses. We illustrate the performances of our model on a simulated study, designed on the basis of a real fMRI experiment.

The rest of the paper is organized as follows. Our Hierarchical Bayes statistical framework is presented in Section 2. In section 3 we present the simulation study and discuss the results.

2 A Wavelet-based Bayesian Hierarchical Model

In an fMRI experiment, a subject’s brain is scanned at multiple time points while the subject performs a set of tasks. Thus, a time series of BOLD response is acquired for each voxel of the brain. Let $Y_v = (Y_{v1}, \dots, Y_{vT})^T$ be the $T \times 1$ vector of the response data, with Y_{vi} the BOLD image intensity at time $i = 1, \dots, T$, for voxel v , with $v = 1, \dots, V$. The BOLD response for a single subject is often modeled with a linear regression model of the type

$$Y_v = X_v \beta_v + \varepsilon_v, \quad \varepsilon_v \sim N_T(0, \Sigma_v), \quad (1)$$

where X_v is a known $T \times p$ covariate matrix and $\beta_v = (\beta_{v1}, \dots, \beta_{vp})^T$ a $p \times 1$ vector of regression coefficients. Without loss of generality, we can assume that the data are centered and no intercept is needed in the model. Furthermore, we assume $p = 1$.

The covariate X_v is modeled as the convolution of the outside stimulus with a function $h(\cdot)$, that represents the delay of the BOLD response with respect to the stimulus onset, as

$$X_v = \int_0^t x(s) h_{\lambda_v}(t-s) ds, \quad (2)$$

where $x(s)$ represents the time dependent stimulus and $h_{\lambda_v}(t) = \exp(-\lambda_v) \lambda_v^t / t!$ (Friston et al., 1994). We assume λ_v to be voxel-dependent.

In order to take into consideration the temporal dependence of the patterns observed across the fMRI scans, we assume Σ_v is a matrix with elements $\Sigma_v(i, j) = [\gamma(|i-j|)]$ for some auto-covariance function $\gamma(h)$. More specifically, we assume a long-memory process, i.e.

$$\gamma(h) \sim Ch^{-\alpha}, \quad (3)$$

for some $C > 0$, $0 < \alpha < 1$ and h large.

We employ a discrete wavelet transform (DWT) to decorrelate the data and thus simplify the cross-covariance structure induced by the long memory process. Thus, we can rewrite (1) in the wavelet domain as

$$Y_v^* = X_v^* \beta_v + \varepsilon_v^*, \quad \varepsilon_v^* \sim N_T(0, \Sigma_v^*), \quad (4)$$

where $Y_v^* = WY_v$, $X_v^* = WX_v$, and $\varepsilon_v^* = W\varepsilon_v$. Because of the wavelet transformation, we can assume that the $(T \times T)$ covariance matrix Σ_v^* is diagonal and write its diagonal elements as $\psi_v \sigma_{mm}^2$, indicating the variance of the n th wavelet coefficient at the m th scale. More specifically, we assume the variance progression formula for the covariance structure of the wavelet coefficients as

$$\psi_v \sigma_{mm}^2 = \psi_v (2^{\alpha_v})^{-m}, \quad (5)$$

with ψ_v the innovation variance and $\alpha_v \in (0, 1)$ the long memory parameter (Jeong et al., 2013).

The two parameters (ψ_v, α_v) characterize the behavior of the fMRI time series at each voxel. In particular, for any set of voxels, similar values indicate close neural activity dynamics. Hence, it is reasonable to investigate the temporal association of the brain voxels by clustering the voxel-dependent vectors of long memory parameters. We assume a Dirichlet Process (DP) prior on the pairs (ψ_v, α_v) . The DP is commonly used in many applications of Bayesian nonparametric methods to provide an unsupervised method of clustering. Let G indicate a realization of the DP prior. The DP can be characterized by specifying two parameters, the base measure G_0 and the total mass parameter η . Typically, G_0 is a parametric model assumed as reference, since $E(G) = G_0$. The total mass parameter characterizes prior uncertainty. Any realization G of the DP defines a discrete probability distribution almost surely. Thus, any sample from G has a positive probability of ties, inducing clustering among the sampled values. In our model, the prior specification for (ψ_v, α_v)

can be summarized as

$$\begin{aligned} (\psi_v, \alpha_v) | G &\sim G \\ G | \eta, G_0 &\sim DP(\eta, G_0) \\ G_0 &= \text{IG}(a_0, b_0) \times \text{Beta}(a_1, b_1) \end{aligned} \quad (6)$$

where the base measure G_0 is specified as the product of inverse gamma (IG) and beta distributions. Since V is large, we fix η so to favor a parsimonious representation of the data, without substantially impacting the inference conclusions. See also Zhang et al. (2014).

In addition to clustering voxels characterized by similar temporal dynamics, we are interested in detecting regions of the brain that show activation in response to a given stimulus. We cast such problem as a problem of variable selection, to identify the nonzero β_v 's in model (4). We use mixture priors with a spike at zero on the regression coefficients (George and McCulloch, 1993; Sha et al., 2004). Let γ_v be a binary random variable such that $\gamma_v = 0$ if $\beta_v = 0$ and $\gamma_v = 1$ otherwise, then we assume

$$\beta_v \sim \gamma_v N(0, \tau) + (1 - \gamma_v) \delta_0, \quad v = 1, \dots, V, \quad (7)$$

with δ_0 a point mass at zero and τ a relatively large variance term for the non-null component. In order to account for spatial correlation in the selection procedure, we further place a Markov Random Field (MRF) prior on the selection parameter γ_v . We follow Li and Zhang (2010) and Stingo et al. (2011) and parameterize the conditional probability of γ_v as

$$P(\gamma_v | d, e, \gamma_k, k \in N_v) \propto \exp(\gamma_v (d + e \sum_{k \in N_v} \gamma_k)) \quad (8)$$

with N_v the set of neighboring voxels of voxel v . According to the prior (8), a voxel has a greater probability of being activated if more of the neighboring voxels are activated. The parameter $d \in (-\infty, \infty)$ represents the expected prior number of activated voxels and controls the sparsity of the model, whereas $e > 0$ affects the probability of identifying a voxel as active according to the activation of its neighbors and therefore acts as a smoothing parameter.

Finally, we complete our prior model by considering a uniform distribution on the interval (u_1, u_2) as a prior distribution for the delay parameter λ_v in the hemodynamic function, as suggested by Quirós et al. (2010),

$$\lambda_v \sim \text{U}(u_1, u_2), \quad v = 1, \dots, V. \quad (9)$$

Posterior inference is obtained by resorting to Markov Chain Monte Carlo (MCMC) sampling algorithms.

3 Illustrative Data Analysis

In this section we simulate a synthetic fMRI data set based on a real experiment, which was originally described by Büchel and Friston (1997). The experiment was performed on a single subject, under 4 different tasks (“Fixation”, “Attention”, “No Attention”, and “Stationary”). Here, we focused on a slice captured under the “no attention” condition and then generated synthetic data as $Y_{syn} = Y + w$, where Y denotes the component simulated from our model as described below and w is the selected slice from the real fMRI study (slice 27). In order to simulate Y , we first obtained the design matrix X for a block design. Then, we convolved X with a Poisson hemodynamic function and applied the wavelet transform. The true activation map defines three rectangular active regions (i.e., $\gamma_v = 1$ in each region), as shown in Figure 1(a). The true values of the β ’s in these regions were sampled from a $\text{Uniform}(-80, 80)$. We set the values of (ψ_v, α_v) at $(10, 0.8)$, $(5, 0.5)$, $(1, 0.2)$, leading to three different clusters of voxels. Finally, we obtained Y as $Y = W^T Y^*$. Our final synthetic dataset comprised time series data for 256 scans of 64×64 voxels, mirroring a real block-design fMRI study. Figure 1(b) shows the first scan.

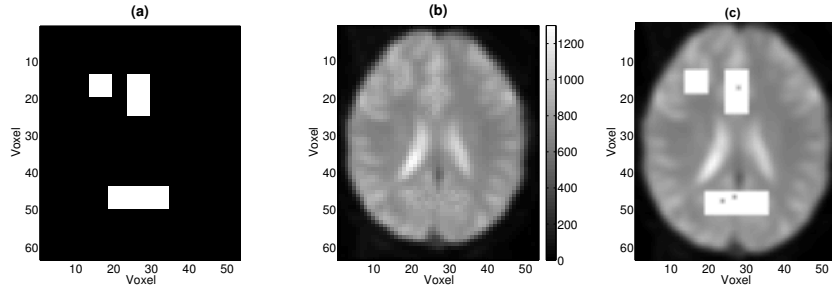


Fig. 1 Synthetic data: (a) True map of γ ; (b) First scan of our synthetic data. (c) Overlay on slice 27 of the posterior activation map obtained by assigning value 1 to those voxels with $p(\gamma_v = 1|y) > 0.8$ and value 0 otherwise

For model fitting, we ran a MCMC algorithm, with 20,000 iterations, discarding the first 10,000 iterations as burn-in. We assessed convergence by the Raftery-Lewis diagnostic test implemented in the R package “coda”. We assigned the non-informative priors to all the parameters of the model and applied, at each iteration, discrete wavelet transforms using Daubechies minimum phase wavelets with 4 vanishing moments. Figure 1(c) shows the posterior activation map, obtained by assigning value 1 to those voxels with $p(\gamma_v = 1|y) > 0.8$ and value 0 otherwise. The accuracy of our detection is 99.86%. Figure 2 shows the posterior clustering map of the voxels and the posterior mean maps of the estimates of the parameters ψ and α . We are able to recover the true clustering of the voxels, with all maps showing three large clusters and a few misclassified voxels, and to well estimate the true values of the parameters.

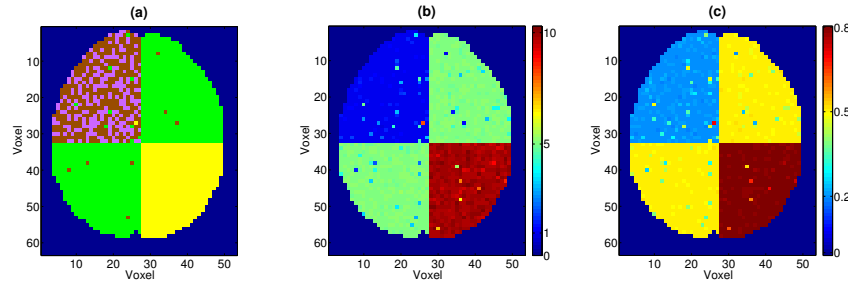


Fig. 2 Synthetic data: (a) Posterior clustering map - different colors correspond to different clusters; (b) posterior mean map of ψ ; (c) posterior mean map of α .

References

- Büchel, C., Friston, K., 1997. Modulation of connectivity in visual pathways by attention: Cortical interactions evaluated with structural equation modelling and fMRI. *Cerebral Cortex* 7 (8), 768–778.
- Friston, K., 2011. Functional and effective connectivity: a review. *Brain Connectivity* 1 (1), 13–36.
- Friston, K., Jezzard, P., Turner, R., 1994. Analysis of functional MRI time-series. *Human Brain Mapping* 1 (2), 153–171.
- George, E., McCulloch, R., 1993. Variable selection via Gibbs sampling. *Journal of the American Statistical Association* 88 (423), 881–889.
- Jeong, J., Vannucci, M., Ko, K., 2013. A wavelet-based Bayesian approach to regression models with long memory errors and its application to fMRI data. *Biometrics* 69, 184–196.
- Li, F., Zhang, N., 2010. Bayesian variable selection in structured high-dimensional covariate spaces with applications in genomics. *Journal of the American Statistical Association* 105 (491), 1202–1214.
- Quirós, A., Diez, R., Gamerman, D., 2010. Bayesian spatiotemporal model of fMRI data. *NeuroImage* 49 (1), 442–456.
- Sha, N., Vannucci, M., Tadesse, M. G., Brown, P., Dragoni, I., Davies, N., Roberts, T., Contestabile, A., Salmon, M., Buckley, C., Falciani, F., 2004. Bayesian variable selection in multinomial probit models to identify molecular signatures of disease stage. *Biometrics* 60 (3), 812–819.
- Stingo, F., Chen, Y., Tadesse, M., Vannucci, M., 2011. Incorporating biological information into linear models: A Bayesian approach to the selection of pathways and genes. *The Annals of Applied Statistics* 5 (3), 1978–2002.
- Zhang L., Guindani M., Versace, F. and Vannucci M., 2014. A Spatio-Temporal Nonparametric Bayesian Variable Selection Model of fMRI Data for Clustering Correlated Time Courses. *Submitted*.