Resting State fMRI Analysis using a Spatial Regression Mixture Model

Vangelis P. Oikonomou, Konstantinos Blekas and Loukas Astrakas‡

Abstract

Functional MRI (fMRI) is one of the most important techniques to study the human brain. A relatively new problem to the analysis of fMRI data is the identification of brain networks when the brain is at rest i.e. no external stimulus is applied to the subject. In this work a method to find the Resting State Networks (RSNs), using fMRI time series, is proposed. To achieve that our method uses the Regression Mixtures Models (RMMs). RMMs are mixture models specifically design to cluster time series. Furthermore, our method takes into account the spatial correlations of fMRI data by using a new functional for the responsibilities of the mixture. Experimental results have showed the usefullness of the proposed approach compared to other methods of the field such as the k - means algorithm.

1. Introduction

fMRI is a technique that help us to understand the function of the brain by using task - based or stimulus driven paradigms. This is achieved by measuring the relative change of the BOLD signal from the baseline during the performance of a task or in response to a stimulus. In recent years there is an increase interest in the application of the method when the subject is at rest. This approach is called resting state fMRI (RS fMRI) or functional connectivity of MR imaging (fc fMRI). RS fMRI investigates syncronous activations between distinct spatially regions in order to identify RSNs. This technique focuses to the analysis of spontaneous low frequency fluctuations of the BOLD signal. The significance of these fluctuations was first presented by [1].

---
*V.P. Oikonomou is with the Laboratory of Knowledge and Intelligent Computing, Department of Informatics and Telecommunication Technology, TEI of Epirus, 47100, Arta, Greece, viknmu@gmail.com

†K. Blekas is with the Faculty of Computer Science and Engineering, University of Ioannina, Greece, kblekas@cs.uoi.gr

‡L. Astrakas is with the Faculty of Medical School, University of Ioannina, Greece, astrakas@uoi.gr

From a physiological viewpoint, in the literature have been reported many RSNs. The most notable of them is the Default Mode Network (DMN). Studies have showed that the functions of brain areas within the DMN are crusial to human mental activity. More specifically, it has been observed the connections between brain areas belonging in the same network are disrupted in pathological case such as Altzheimer disease [2]. Also, several other RSNs have been identified. These networks include the somatosensory network, the visual network and the auditory network among others [3]. These networks have demonstrated high concistency and reproducibility across subjects.

Although the study of resting state fMRI time series is a relatively new field, enough works have been published in last years concerns the analysis of time series from a methological point of view. Methods based on correlation analysis, such as the seed - based analysis [1], are proposed. Also, multivariate approaches such as ICA are used to analyze RS fMRI time series. More specifically, the ICA approach is used to identify spatially independent resting networks [4]. Also, Graph based methods are used to study the time series. These approaches represent the brain as a collection of nodes connected with edges [5]. Finally, clustering methods, such as the k-means, are used to divide the time series into distinct groups where the members inside a group have large similarity while each group compare to others present great differences [6].

In this work we proposed a new mixture model to find the RSNs of the brain. More specifically, a regression mixture model is utilized. Regression mixture models are designed to dealt with various aspects of time series analysis such as variable size of time series. The spatial characteristics of fMRI data are introduce into our model through the mixing probabilities of the mixture. In addition a new prior for the mixing probabilities is proposed. This prior acts as a mean filter on the neighborhood of a time series. The remainder of this paper is organized as follows: In section 2 we present the regression mixture model while in Section 3 the regression mixture model with spatial properties is

described. In section 4 results using simulated and real data are presented. Furthermore, a comparison of the proposed approach with the k-means algorithm is provided. Finally, in Section 5 we give some concluding remarks and future directions.

2. Mixture Modeling of Time Series

Let $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_N\}$ be a set of N fMRI time series of length T, i.e. each element $\mathbf{y}_n$ is a sequence of values measured at T successive time instances x_l, i.e. $\mathbf{y}_n = \{y_{nl}\}_{l=1,\cdots,T}$. It must be noted that the data analysis in our study has been made on a single slice. We assume that the time series can be described using a linear regression model. Hence each time series can be formulated as:

$$\mathbf{y}_n = \mathbf{X}\mathbf{w}_n + \mathbf{e}_n ,\qquad (1)$$

where $\mathbf{w}_n$ is the vector of M (unknown) linear regression coefficients, while $\mathbf{e}_n$ corresponds to the stochastic noise term that is assumed to be zero mean Gaussian with variance σ_n^2, i.e. $\mathbf{e}_n \sim \mathcal{N}(0, \sigma_n^2\mathbf{I})$. Finally, $\mathbf{X}$ is the design matrix where its selection plays an important role for the data analysis. A typical design matrix scheme is the Vandermonde or B-splines matrix dealing with polynomial or splines models, respectively. Also, predetermined dictionaries such as the Fourier Transform, the Discrete Cosine Transfofm and the Wavelet Transform can be used to construct the design matrix. Finally, another approach is to assume a kernel design matrix using an appropriate kernel basis function over time instances $\{x_l\}_{l=1}^{T}$. A common choice is to use the Gaussian kernel

$$[X]_{lk} = K(x_l, x_k; \lambda) = \exp\left(-\frac{(x_l - x_k)^2}{2\lambda}\right) ,$$

where λ is a scalar parameter. In addition, the design matrix may contain information about the experimental paradigm of fMRI experiment [7].

According to this model, the conditional probability density of the observative sequence $\mathbf{y}_n$ is also Gaussian

$$p(\mathbf{y}_n|\theta_n) = \mathcal{N}(\mathbf{X}\mathbf{w}_n, \sigma_n^2\mathbf{I}) .$$

where θ_n is the set of model parameters, i.e. $\theta_n = \{\mathbf{w}_n, \sigma_n^2\}$.

In this study we consider the task of fMRI time-series clustering, i.e. the division of the input set of time series Y into K clusters, in such a way that each cluster contains similar in nature elements. This is equivalent of assuming that each cluster has its own regression generative mechanism, as given by a conditional density with parameters $\theta_j = \{\mathbf{w}_j, \sigma_j^2\}$. Mixture modeling provides a powerful platform of establishing the clustering

procedure. It is described with the following probability density:

$$f(\mathbf{y}_n|\Theta) = \sum_{j=1}^{K} \pi_j p(\mathbf{y}_n|\theta_j) ,\qquad (2)$$

where π_j are the weights (prior probabilities) of every cluster that satisfy the constraints: $\pi_j \geq 0$ and $\sum_{j=1}^{K} \pi_j = 1$, while Θ is the set of all mixture model parameters, i.e $\Theta = \{\pi_j, \theta_j\}_{j=1}^{K}$. Assignment of the data to the K groups is then achieved according to the maximun of the posterior probabilities of component membership:

$$z_{jn} = \frac{\pi_j p(\mathbf{y}_n|\theta_j)}{f(\mathbf{y}_n|\Theta)} .\qquad (3)$$

Based on the above formulation, the task of clustering can be converted into a parameter estimation problem where we need to maximize the log-likelihood function:

$$l(\Theta) = \log p(Y|\Theta) = \sum_{n=1}^{N} \log\left\{\sum_{j=1}^{K} \pi_j p(\mathbf{y}_n|\theta_j)\right\}.\qquad (4)$$

3. Spatially Variant Mixture Modeling of fMRI time-series

The fMRI time series are structures that involve spatial properties, where adjacent voxels tend to have similar activity behavior. This effect is introduce into our model by treating carefully the mixing probabilities. More specifically, we assume that any sequence $\mathbf{y}_n$ has a set of labels π_{nj} that provides with the degree of belonginess to each cluster j. We treat these parameters as random variables that have also the contraints $\pi_{nj} \geq 0$ and $\sum_{j=1}^{K} \pi_{nj} = 1$.

Following the theory of Markov Random Fields (MRF), we consider a Gibbs distribution over these mixing labels parameters. This establishes spatial dependencies and offers smoother solutions. More specifically, we assume a Gibbs potential function of the following formulation:

$$\vartheta_{jn} = \sum_{m \in N_n} z_{jn} z_{jm}\qquad (5)$$

This function describes the influence of the neighborhood to the decision process. More specifically, it acts as a smooth filter to the posteriors and it works like a voting system i.e. a particular time series will be grouped into the cluster where the most time series of the neighborhood belong. Finally, the following prior over mixing probabilities is given by:

$$\pi_{nj} = \frac{\vartheta_{jn}}{\sum_{k=1}^{K} \vartheta_{kn}}\qquad (6)$$

An important role in using a regression model is how to estimate its order M, i.e. the size of linear regression coefficients $\mathbf{w}_j$. Estimating the proper value of M depends on the shape of data to be fitted, where models of small order may lead to underfitting, while large values of M may become responsible for data overfitting. This may deteriorate significantly the clustering performance. Bayesian regularization framework provides an elegant solution to this problem [8, 9]. It initially assumes a large value of order M. Then, a heavy tailed prior distribution $p(\mathbf{w}_j)$ is imposed upon the regression coefficients that will enforce most of the coefficients to be zero out after training.

The sparsity can be achieved in an hierarchical way [8] by considering first a zero-mean Gaussian distribution over the regression coefficients:

$$ p(\mathbf{w}_j|\alpha_j) = \mathcal{N}(\mathbf{w}_j|\mathbf{0}, A_j^{-1}) = \prod_{l=1}^{M} \mathcal{N}(w_{jl}|0, \alpha_{jl}^{-1}) , \quad (7) $$

where A_j is a diagonal matrix containing the M components of the precision (inverse variance) vector $\alpha_j = (a_{j1}, \ldots, a_{jM})$. At a second level, precision can be seen as hyperparameters that follow a Gamma prior distribution:

$$ p(\alpha_j) = \prod_{l=1}^{M} \Gamma(\alpha_{jl}|b, c) \propto \prod_{l=1}^{M} \alpha_{jl}^{b-1} \exp^{-c\alpha_{jl}} . \quad (8) $$

Note that both Gamma parameters b and c are a priori set to zero so as to achieve uninformative priors. The above two-stage hierarchical sparse prior is actually the Student's-t distribution enforcing most of the values α_{jl} to be large and thus eliminating the effect of the corresponding coefficients w_{jl} by setting to zero. In such way the regression model order for every cluster is automatically selected and overfitting is avoided.

The clustering procedure becomes now a Maximum-A-Posteriori (MAP) estimation problem, where the MAP log-likelihood function is given by

$$ l_{MAP}(\Theta) = \sum_{n=1}^{N} \log\{\sum_{j=1}^{K} \pi_{nj} p(\mathbf{y}_n|\theta_j)\} + \log p(\Pi) $$
$$ + \sum_{j=1}^{K} \{\log p(\mathbf{w}_j|\alpha_j) + \log p(\alpha_j)\} . \quad (9) $$

where $\Theta = \{\{\pi_{nj}\}_{n=1}^{N}, \theta_j = (\mathbf{w}_j, \alpha_j, \sigma_j^2)\}_{j=1}^{K}$ and $p(\Pi) = \prod_{n=1}^{N} \prod_{j=1}^{K} \pi_{nj}$. Employing the EM algorithm to MAP estimation requires at each iteration the conditional expectation values z_{nj} of the hidden variables to be computed first (E-step):

$$ z_{jn} = \frac{\pi_{nj} p(\mathbf{y}_n|\theta_j)}{\sum_{k=1}^{K} \pi_{nk} p(\mathbf{y}_n|\theta_k)} . \quad (10) $$

Also, during the M- step of the algorithm we obtained:

$$ \mathbf{w}_j = \left[\left(\sum_{n=1}^{N} z_{nj}\right)\frac{1}{\sigma_j^2}\mathbf{X}^T\mathbf{X} + \mathbf{A}_j\right]^{-1} $$
$$ \cdot \frac{1}{\sigma_j^2}\mathbf{X}^T\left(\sum_{n=1}^{N} z_{nj}\mathbf{y}_n\right), \quad (11) $$

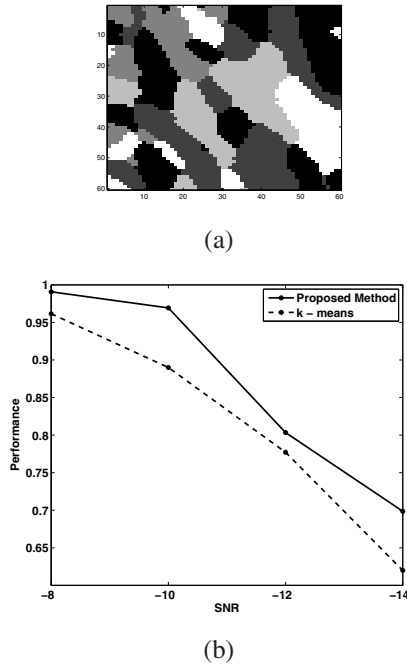$$ \alpha_{jl} = \frac{1+2c}{w_{jl}^2 + 2b}, \quad (12) $$

$$ \sigma_j^2 = \frac{\sum_{n=1}^{N} z_{nj}\|\mathbf{y}_n - \mathbf{X}\mathbf{w}_j\|^2}{T\sum_{n=1}^{N} z_{nj}} . \quad (13) $$

To conclude the proposed algorithm consists from Eqs. (10) - (13). These equations are applied iteratively until converged. Furthermore, π_{nj} are calculated by using Eq. (6). Finally, to assign labels in each time series the posterior probabilities are used (see Eq. 10).

4. Experiments

We have conducted experiments using simulated fMRI time series. The proposed algorithm has been compared with the k-means algorithm. During the experiments with simulated fMRI data, we have created 3-D datasets of time series using linear regression models with known design matrix and regression coefficients. In these time series, we have added white Gaussian noise of various SNR levels. The spatial correlation between the time series is achieved through the regression coefficients. The spatial patterns are drawn in Fig. 1(a). Finally, for each SNR level we studied the performance of the comparative methods by executing 50 Monte Carlo simulations. To measure the quality of each clustering approach, we have used the performance, calculated as the percentage of correctly classified time series. As we can observe the proposed method provides us with better results compared to k-means algorithm (see Fig. 1(b)). More specifically, the performance of the proposed approach for -8 dB is 0.9908 while for the k - means is 0.9616. As the SNR is becoming smaller the performance of both approaches is detoriated. However, the detoriation rate of the proposed approach is smaller than that of k - means.

We have test the proposed approach in real data and we have compared with the k-means algorithm. These data have been downloaded from NITRC (http://www.nitrc.org), under the project name: NYU CSC TestRetest. A detailed description of the dataset can be found in [10]. From this dataset a subject was chosen to study the performance of our method. We have applied the proposed method and the k - means to the raw time - series without any preprocessing steps. This will showed how robust are the two approaches in
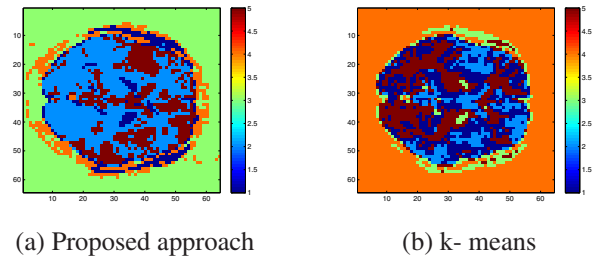
(a)

(b)

Figure 1. (a) Spatial patterns with 5 clusters employed in our experiments with simulated time series and (b) Performance of the two methods for various SNR levels

difficult and noisy situations. Both methods are applied by define the number of clusters to 5. The obtained results are shown in Fig. (2). As we can see both approach find the DMN (cluster 5 for the proposed approach and cluster 2 for the k-means), however the proposed approach provides us with a cluster more compact than the k-means. Also, this feature can be observed to others provided clusters.

5. Conclusions

In this work a new regression mixture model was proposed. More specifically, a new spatial prior over the mixture propabilities of the mixture model was defined. Furthermore, to estimate the model parameters the EM framework was adopted. The proposed method has been tested in real resting state fMRI time series and simulated cases and it has been compared with k - means, while the provided results have shown its superiority. In the future we intend to study more carefully the concept of model order (i.e. number of clusters) and the behavior of the method in larger resting state fMRI datasets.

(a) Proposed approach   (b) k- means

Figure 2. Obtained clusters in the cases of real data. (In colorbars the cluster indexes)

References

[1] B. Biswal, F. Z. Yetkin, V. M. Haughton, and J. S. Hyde, "Functional connectivity in the motor cortex of resting human brain using echo-planar mri," *Magn. Reson. Med.*, vol. 34, pp. 537–541, 1995.

[2] Y. Liu, K. Wang, C. Yu, Y. He, Y. Zhou, M. Liang, L. Wang, and T. Jiang, "Regional homogeneity, functional connectivity and imaging markers of Alzheimer's disease: a review of resting-state fMRI studies.," *Neuropsychologia*, vol. 46, no. 6, pp. 1648–1656, 2008.

[3] J. Damoiseaux, S. A. R. B. Rombouts, F. Barkhof, P. Scheltens, C. J. Stam, S. M. Smith, and C. F. Beckmann, "Consistent resting-state networks across healthy subjects.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 37, pp. 13848–13853, 2006.

[4] C. F. Beckmann, M. DeLuca, J. T. Devlin, and S. M. Smith, "Investigations into resting-state connectivity using independent component analysis," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 360, no. 1457, pp. 1001–1013, 2005.

[5] J. Wang, X. Zuo, and Y. He, "Graph-based network analysis of resting-state functional MRI.," *Frontiers in systems neuroscience*, vol. 4, 2010.

[6] Y. Golland, P. Golland, S. Bentin, and R. Malach, "Data-driven clustering reveals a fundamental subdivision of the human cortex into two global systems," *Neuropsychologia*, vol. 46, pp. 540–553, 2008.

[7] V. P. Oikonomou and K. Blekas, "An adaptive regression mixture model for fmri cluster analysis," *IEEE Trans. Med. Imaging*, vol. 32, pp. 649–659.

[8] M. E. Tipping, "Sparse Bayesian Learning and the Relevance Vector Machine," *Journal of Mach. Learn. Research*, vol. 1, pp. 211–244, 2001.

[9] C. M. Bishop, *Pattern Recognition and Machine Learning*. 2006.

[10] Z. Shehzad, A. M. C. Kelly, P. T. Reiss, D. G. Gee, K. Gotimer, L. Q. Uddin, S. H. Lee, D. S. Margulies, A. K. Roy, B. B. Biswal, E. Petkova, F. X. Castellanos, and M. P. Milham, "The resting brain: Unconstrained yet reliable," *Cerebral Cortex*, vol. 19, no. 10, pp. 2209–2229, 2009.