31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

On Clustering fMRI Using Potts and Mixture Regression Models

Jing Xia, Feng Liang and Yongmei Michelle Wang

Abstract— **In this paper, we propose a model based clustering method for functional magnetic resonance imaging (fMRI) data to detect the functional connectivity network. The Potts model, which represents spatial interactions of neighboring voxels, is introduced to integrate the temporal mixture regression modeling into one single unified model. The estimation of the parameters is achieved through a restoration maximization (RM) algorithm for computation efficiency and accuracy. Additional features of our method include: the optimal number of clusters can be automatically determined; global trends and informative paradigms of the data are extracted by a dimension reduction algorithm based on principal component analysis (PCA) and a statistical significance test. Experimental results demonstrate that our approach can lead to robust and sensitive detection of functional networks.**

I. INTRODUCTION

Analysis of functional interrelations among brain regions from functional magnetic resonance imaging (fMRI) is of increasing interest in neuroscience research. A variety of model-dependent, such as the general linear model (GLM) [1], and model-independent approaches, such as principal component analysis (PCA) [2], independent component analysis (ICA) [3] and partial least square (PLS) [4], have been proposed to determine task-specific activations and to identify functionally connected regions (i.e., functional connectivity) of the brain. However, GLM is based on univariate analysis and does not take the spatial correlation into account. PCA, ICA and PLS are based on multivariate analysis, but it is sometimes difficult to interpret so many eigenimages and one single feature of interest may split into multiple components.

In this paper, we propose a stochastic clustering method using a finite mixture regression model and a Potts model to identify functional clusters or networks, based on the seed voxels/regions. Comparing with the aforementioned methods, our method has several distinctive advantages. First, in the traditional seed-voxel approach, the time series of seed voxels/regions serve as reference functions to identify the connectivity map. It is impractical to use time series of all voxels from the seed region since the collinearity among the large number of time series will lead to an unstable model. Therefore, extracting the significant signals from the seed region is crucial for learning the functional connectivity network. We propose to use key principal components (PCs) from seed regions, with the number of PCs selected by statistical significance test. Compared with the mean time series, the PCs form a parsimonious, but yet informative summary of the signals from the seed region, which makes our model more robust and less sensitive to noise. Second, we model the functional dependence on the seed regions via a mixture regression model, where different mixture components represent different functional clusters of activation and connectivity. Our mixture model is more flexible than a single parametric model used in most model based methods and can provide a satisfactory result when local variations are present in the data [5]. Third, we model the neighborhood dependence through a Potts model, a generalization of the Ising model [6], which can utilize shared information across neighboring voxels to fit the temporal regression model. For a finite mixture model, it is standard to use the expectation maximization (EM) algorithm to estimate parameters. However, the incorporation of the dependence among neighboring voxels is computationally intractable because of the incalculable normalizing constant in the Potts model. Instead, we adapt a restoration maximization (RM) algorithm [7] to estimate the parameters in our mixture regression model with spatial structure constraints, for efficiency and accuracy.

The remainder of the paper is organized as follows. Section 2 formulates the functional connectivity analysis as a mixture regression model. Section 3 introduces the RM algorithm that incorporates spatial dependence among neighboring voxels via the Potts model. Section 4 presents the experimental results, and Section 5 closes this paper with conclusions.

II. A MIXTURE MODEL FOR FMRI REGRESSION

Let $Y_i = [y_{i1}, y_{i2}, \cdots, y_{iT}]'$ denote the intensity of the fMRI time series at voxel $i = 1, \cdots N$, where N is the total number of voxels and T is the number of time points. We model the observed time series for voxel i by a linear model as

$$Y_i = \mathbf{X}\boldsymbol{b}_i + \boldsymbol{\varepsilon_i}, \tag{1}$$

J. Xia is with the Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA `jingxia2@illinois.edu`

F. Liang is with the Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA, `liangf@illinois.edu`

Y. M. Wang is with the Departments of Statistics, Psychology and Bioengineering, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA, `ymw@illinois.edu`

978-1-4244-3296-7/09/$25.00 ©2009 IEEE

where $\mathbf{X}$ is a design matrix consisting of significant paradigms from different stimuli, seed regions or the whole image, $\boldsymbol{b}_i$ is an unknown linear regression coefficient, and $\boldsymbol{\varepsilon_i}$ is some noise process usually modeled by a AR(1) or ARMA(1,1) model. For example, in the standard GLM approach for connectivity analysis, the design matrix includes the intercept, the cosine bases which represent global trends with low frequencies, and the mean time series of the regions of interest (ROIs) or seed regions [1].

In our approach, instead of using parametric cosine bases to represent global trends and mean time series to represent seed regions, we form our regressors via a nonparametric data-dependent approach based on principal component analysis (PCA). Since the majority of voxels are not involved in activation or connectivity, we use the top PCs from the whole fMRI data to represent the global trends. For each of the ROIs or seed regions, we extract the seed paradigms using PCA after taking away global trends. In the standard PCA, the number of PCs is often chosen to keep a certain fraction, say 95%, of the total variation in the data, so the reduced data still contain about 95% of the original information. However, our goal is different here: the selected PCs should represent the signals that are significantly different from noise. Therefore, we adapt a χ^2 test from [8] for the selection instead of using a fixed variance percentage cut-off.

To model the functional dependence of voxels on the seed regions, we introduce a latent variable $Z_i \in \{1, 2, \cdots, K\}$ to denote the functional cluster label for voxel i . Voxels in the same functional cluster should depend on signals from various sources (background and seed regions) in a similar way, which is reflected on the similarity of their regression coefficients. To identify functionally dependent voxels, we model all the regression coefficients $\boldsymbol{b}_i$'s by a mixture model: $\boldsymbol{b}_i = \boldsymbol{\beta}_k, Z_i = k$ with probability w_k. Consequently, the data Y_i follows a finite mixture regression model

$$Y_i|\boldsymbol{\Theta} \sim \sum_{k=1}^{K} w_k \text{MVN}(\mathbf{X}\boldsymbol{\beta}_k, \sigma^2\Sigma)$$

where $\boldsymbol{\Theta} = \{\omega_1, \cdots, \omega_K, \boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_K, \sigma^2\}$ denotes the collection of all unknown parameters. To simplify the computation, we assume a common temporal correlation matrix Σ for all voxels, which is known or can be estimated by plug-in.

III. PARAMETER ESTIMATION VIA RM

A. Potts Model

A special feature of fMRI data is the similarity between neighboring voxels. To incorporate this spatial information into the inference on functional clusters, we introduce a Potts model as a prior distribution on the cluster labels $\boldsymbol{Z} = [Z_1, Z_2, \cdots, Z_N]^{'}$. The Potts model is a model of interacting spins on a lattice system and is a generalization of the well-known Ising model. The joint density function is given by $p(\boldsymbol{Z}) \propto \exp(\alpha H(\boldsymbol{Z}))$, where $H(\boldsymbol{Z})$ is the energy function representing the interaction among voxels (more similar the neighboring voxels, the higher the energy), and α is a tuning parameter, also known as the inverse temperature. When the temperature is low (i.e., α is large), all labels tend to be identical, which corresponds to a single cluster. As the temperature increases, the single cluster will split into multiple small ones. Therefore, different choices of the tuning parameter α reflect our prior belief on the size of the clusters. Later we will introduce a data-dependent tuning approach for α.

In the context of fMRI functional connectivity network, the energy function based on functional distance is more meaningful since it allows disjoint brain regions to link together [6]. Therefore, we modify the energy function used in our work as $H(\boldsymbol{Z}) = \sum_{i\sim j} \kappa_{ij}(1 - \delta_{Z_i Z_j})$, where $i \sim j$ indicates voxel i and j are neighbors (we use the 2D/3D Euclidean spatial neighborhood system throughout), $\delta_{Z_i Z_j}$ is the delta function that equals to 1 if $Z_i = Z_j$ and 0 otherwise, $\kappa_{ij} = \frac{1}{n}\exp\left(-f_{ij}^2/2a^2\right)$ with n being the average number of neighbors per voxel, a being the average functional distance of the nearest neighbor and $f_{ij} = 1 - corr(Y_i, Y_j)$ denoting the functional distance between two voxels i and j. Since the correlation coefficient $corr(Y_i, Y_j)$ measures the functional similarity of the acquired fMRI time series, the new energy function scales the interaction between voxels by their similarity presented in the data: the more similar the data, the higher the interaction.

B. Restoration-Maximization (RM) Algorithm

The EM algorithm is a popular maximum likelihood estimation method for mixture models, however, it faces two obstacles when applied to our model. In the E-step we have to evaluate

$$E\left[\log(f(\boldsymbol{Z}, \mathbf{Y}|\boldsymbol{\Theta}, \alpha)|\mathbf{Y}, \hat{\boldsymbol{\Theta}}^{(m)}, \hat{\alpha}^{(m)}\right]$$

where the expectation is taking over the latent variable $\boldsymbol{Z}$ conditioning on the data $\mathbf{Y}$ and the value of the parameters $\boldsymbol{\Theta}$ and α from the m^{th} iteration. With a large number of voxels, the joint distribution of all Z_i's is difficult to work with. Further, the probability function of the Potts model is defined up to a normalizing constant, which is needed in the EM algorithm, but practically impossible to be calculated when the number of voxels is large.

Instead of averaging out the unobserved latent variable $\boldsymbol{Z}$ as that in the E-step, an alternative approach is to replace each Z_i by its most likely value, i.e., the probability mode. Finding the mode of the joint distribution of $\boldsymbol{Z}$ is still challenging, so we do it sequentially for each i conditioning on other Z_j's, which leads to the following Restoration

Maximization (RM) algorithm.

R-step: for each i, set $Z_i^{(m+1)}$ equal to the mode of $p\left(Z_i|\mathbf{Y}, \hat{\alpha}^{(m)}, \boldsymbol{Z}_{[-i]}\right)$, where $\boldsymbol{Z}_{[-i]}$ denotes all cluster labels except Z_i. Due to the special form of Potts models, the conditional density function can be further simplified to a form which only depends on the neighboring voxels.

M-step: choose $\boldsymbol{\Theta} = \hat{\boldsymbol{\Theta}}^{(m+1)}$ to maximize $f(\mathbf{Y}|\boldsymbol{Z}^{(m+1)}, \boldsymbol{\Theta})$ and choose $\alpha = \hat{\alpha}^{(m+1)}$ to maximize the pseudo-likelihood $\prod_{i=1}^{N} f\left(Z_i^{(m+1)}|\boldsymbol{Z}_{[-i]}^{(m+1)}, \alpha\right)$, where the pseudo-likelihood is used to avoid the complicated normalizing constant as suggested in [7].

C. Details on the updating step

At the $(m+1)^{\text{th}}$ iteration, after updating $\boldsymbol{Z}^{(m+1)}$ sequentially, update ω_k, $\boldsymbol{\beta}_k$ and σ^2 as following: for $k = 1, \ldots, K$,

$$\begin{aligned}
\hat{\omega}_k^{(m+1)} &= \frac{1}{N}\sum_{i=1}^{N} I\left(Z_i^{(m+1)} = k\right), \\
\hat{\boldsymbol{\beta}}_k^{(m+1)} &= \frac{\sum_{i:Z_i^{(m+1)}=k}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}Y_i}{N\hat{\omega}_k^{(m+1)}}, \\
(\hat{\sigma}^2)^{(m+1)} &= \frac{1}{NT}\sum_{k=1}^{K}\sum_{i:Z_i^{(m+1)}=k}\left(Y_i - \mathbf{X}\hat{\boldsymbol{\beta}}_k^{(m+1)}\right)' \\
&\quad \Sigma^{-1}\left(Y_i - \mathbf{X}\hat{\boldsymbol{\beta}}_k^{(m+1)}\right).
\end{aligned}$$

Numerical optimization methods are used to estimate α and the updating step for α is as following:

$$\alpha^{(m+1)} = \alpha^{(m)} - \left(\frac{\partial^2 l\left(\alpha^{(m)}\right)}{\partial^2\alpha^{(m)}}\right)^{-1}\frac{\partial l\left(\alpha^{(m)}\right)}{\partial\alpha^{(m)}},$$

where the log pseudo-likelihood function is given by

$$\begin{aligned}
l(\alpha) &= \sum_{i=1}^{N}\log p(Z_i = k|\boldsymbol{Z}_{[-i]}, \alpha) \\
&= \sum_{i=1}^{N}\sum_{k=1}^{K} I(Z_i = k)\log(p(Z_i = k|\boldsymbol{Z}_{[-i]}, \alpha))
\end{aligned}$$

IV. RESULTS

A. Simulated Results

In the simulated experiments, we compare the results with and without spatial Potts priors at different noise levels. When no Potts model is embedded, the standard EM algorithm and the GLM method are used for comparison. The synthetic data size is 40×40 and the number of time points is 120 with a block design. There are two global trends in the image, and three seed time series are generated by the SPM package for three functional clusters or networks (Figure 1). The functional cluster sizes for the three networks are 7.86%, 7.11% and 7.74%, respectively. We performed the experiments at three different signal-to-noise ratio (SNR) levels ($\text{SNR} = 0.1$, 0.2 and 0.3).

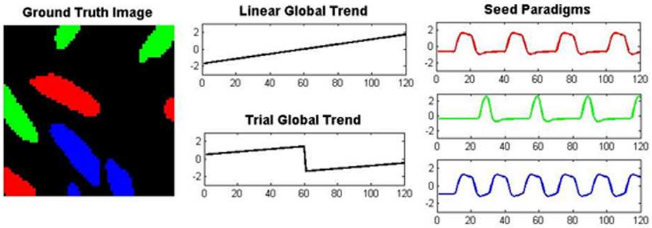

Fig. 1. Ground truth clusters of the simulated image (left); two global trends (central); and three seed paradigms (right).

The first step in our method is to use PCA to extract the global trends and the seed paradigm for each seed region. From a χ^2 test on the PCA eigenvalues, we choose one PC as the global trend and one PC in each of the seed regions as the paradigm.

Figure 2 plots the clustering results from the three methods at three different SNR levels. Table 1 is a list of the corresponding pairwise misclassification rate, which is defined as $1 - \frac{1}{N}\sum_{i=1}^{N} I(\hat{Z}_i = Z_i)$, where $\hat{Z}_i$ is the estimated value and Z_i is the true value. As SNR decreases, the EM algorithm fails to detect correct functional clusters, since global trends dominate the entire activity and no spatial model is available for additional information. However, the RM algorithm works well by modeling the neighborhood dependence through a Potts model. The GLM is not sensitive to global trends but the performance is not as good as the RM. Note that here we assume the true number of clusters is known, i.e., 4.

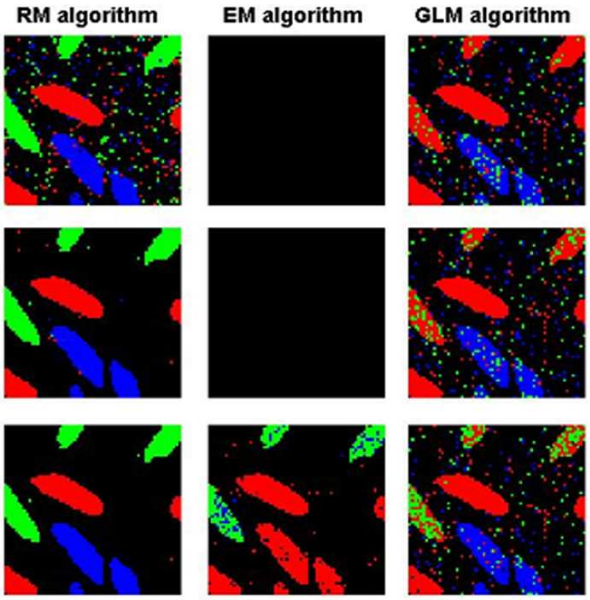

Fig. 2. Comparison Results of RM algorithm, EM algorithm and GLM at SNR=0.1 (top), 0.2 (middle), 0.3 (bottom).

Next, without the assumption of the known and true number of clusters, K, we run our RM algorithm for a range of K values from 2 to 7. Then we use AIC/BIC criteria to determine K. The optimal number of clusters turns out to be

TABLE I

COMPARISON OF PAIRWISE MISCLASSIFICATION RATES

SNR	RM algorithm	EM algorithm	GLM algorithm
0.1	0.0867	0.2319	0.1482
0.2	0.0040	0.2319	0.1262
0.3	0.0012	0.1023	0.0950

4 for both AIC and BIC, which is consistent with the ground truth. The clustering results for different K are displayed in Figure 3, and it indicates that $K = 4$ produces the best fitting.

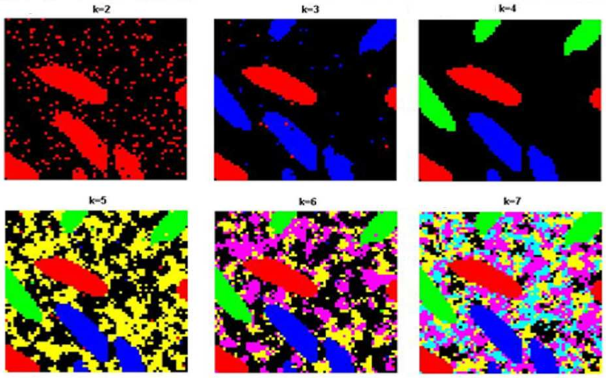

Fig. 3. Comparison of RM algorithm at different number of networks when SNR=0.2.

B. Empirical Results

The real fMRI data set ($53 \times 63 \times 46 \times 360$) is obtained from the SPM data site (*http://www.fil.ion.ucl.ac.uk/spm/data/attention.html*) with a visual motion task. The subject was scanned during four runs, with 90 image volumes in each run. Four conditions - “fixation”, “attention”, “no attention” and “stationary” - were used and there were 10 multi-slice volumes per condition. The SPM package is used for the standard preprocessing.

We use PCA to extract the global trends of the data. Based on the χ^2 test, three principal components are selected as the regressors. The first component corresponds to a linear trend, the second one corresponds to a low frequency trend with some task paradigms, and the third one represents a jump trend resulting from different trials. From AIC/BIC, the optimal cluster or network number is chosen to be $K = 4$. Figure 4 is a demonstration of the posterior cingulate cortex (PCC) functional network. In the figure, we can find several regions of interest, such as left inferior parietal cortex (left IPC: region “A”), right inferior parietal cortex (right IPC: region “B”), ventral anterior cingulate cortex (vACC: region “C”), medial prefrontal cortex (MPFC: region “D”) and left dorsolateral prefrontal cortex (left DLPFC: region “E”). This demonstrates that such regions are functionally connected with the PCC even in visual tasks [9].

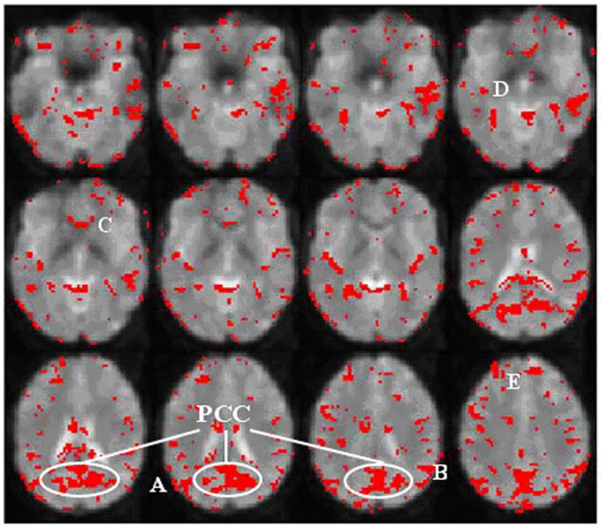

Fig. 4. Functional connectivity network of PCC regions in real fMRI data.

V. CONCLUSION

This paper presents a novel fMRI functional network detection method based on a finite mixture regression model, incorporating the spatial interactions of neighboring voxels via a Potts model. The estimation of the parameters is achieved through the RM algorithm for computation efficiency and accuracy. In addition, the global trends and the informative paradigms of the data are extracted from PCA and statistical χ^2 test. Experimental results on simulated and real fMRI data show that the proposed approach can lead to a robust and sensitive detection of functional clusters and networks.

REFERENCES

[1] K. Friston, P.Jezzard, and R. Turner, “Analysis of functional MRI time-series,” *Human Brain Mapping*, pp. 153–171, 1994.

[2] A.H. Andersen, D.M. Gash, and M.J. Avison, “Principal component analysis of the dynamic response measured by fMRI: A generalized linear systems framework,” *Magnetic Resonance Imaging*, pp. 795–815, 1999.

[3] M.J. McKeown and T.J. Seijnowski, “Independent component analysis of fMRI data: Examining the assumptions,” *Human Brain Mapping*, pp. 368–372, 1998.

[4] A.R. MaIntosh, W.K. Chau, and A.B. Protzner, “Spatiotemporal analysis of event-related fmri data using partial least squares,” *NeuroImage*, pp. 764–775, 2004.

[5] G. Mclachlan and D. Peel, *Finite Mixture Models*, John Wiley and Sons Inc, Canada, first edition, 2000.

[6] Larissa Stanberry, Alejandro Murua, and Dietmar Cordes, “Functional connectivity mapping using the ferromagnetic potts spin model,” *Human Brain Mapping*, pp. 422–440, 2008.

[7] W. Qian and D.M. Titterington, “Estimation of parameters in hidden markov models,” *Philosophical Transactions of the Royal Society of London*, pp. 407–428, 1991.

[8] D. F. Morrison, *Multivariate Statistical Methods*, Mc Graw-Hill, New York, first edition, 1990.

[9] M.D. Greicius, B. Krasnow, A.L. Reiss, and V. Menon, “Functional connectivity in the resting brain: A network analysis of the default mode hypothesis,” *Proceedings of the National Academy of Sciences*, pp. 253–258, 2003.