Multi-Level Functional Connectivity by Markov Random Field

Wei Liu[a], Suyash P. Awate[a], P. Thomas Fletcher[a]

[a] *Scientific Computing and Imaging Institute, University of Utah, USA*

Abstract

Identifying functional networks from resting-state functional MRI is a challenging task, especially for multiple subjects. Most current studies estimate the networks in a sequential approach, i.e., they identify each individual subject's network independently to other subjects, and then estimate the group network from the subjects networks. This one-way flow of information prevents one subject's network estimation benefiting from other subjects. We propose a hierarchical Markov Random Field model, which takes into account both the within-subject spatial coherence and between-subject consistency of the network label map. Both population and subject network maps are estimated simultaneously using a Gibbs sampling approach in a Monte Carlo Expectation Maximization framework. We compare our approach to two alternative groupwise fMRI clustering methods, based on K-means and Normalized Cuts, using both synthetic and real fMRI data. We show that our method is able to estimate more consistent subject label maps, as well as a stable group label map.

Keywords: resting-state functional MRI, segmentation, functional connectivity

1. Introduction

Resting-state functional MRI (rs-fMRI) is widely used for detecting the intrinsic functional networks of the human brain. The analysis of rs-fMRI data is a challenging task, due to the scanner and physiological noise, head motion, and subject's random thoughts during data acquisition. The availability of large rs-fMRI databases open the door for systematic group studies of functional connectivity. Despite that the inherent noise in rs-fMRI poses difficulty of functional network estimation at the individual level, combining many subjects' data together and jointly estimating the common functional networks is more robust.

For identifying functional networks by using rs-fMRI of a single subject, One of the most common approaches is independent component analysis (ICA) and its variants Calhoun et al. (2001b). ICA recover the statistically independent functional networks without *a priori* knowledge of the regions of interest. The functional networks can also be identified by solving an image segmentation problem. By using clustering algorithms (also know as parcellations), the brain voxels are partitioned into disjoint regions that represent the functional networks.

As for study of multiple subjects, Group ICA Calhoun et al. (2001b) is a generalization of single-subject ICA. in group ICA, all subjects are assumed to share a common spatial component map but have distinct time courses. The time courses from all subjects are concatenated temporally, followed by a single subject ICA. The subject component maps are obtained by a back-reconstruction procedure. Alternatively, to generalize single-subject clustering methods to group study, the time courses of all subjects can be concatenated just like group ICA, resulting in a single group segmentation. Or, segmentations are estimated from each single subject, and then are averaged to obtain a group affinity matrix, based on which a second level segmentation is followed (Bellec et al., 2010; Van Den Heuvel et al., 2008).

However, many of the currently used methods have one or more of the drawbacks listed below:

- Due to the anatomical difference among subjects, the existing registration routines can not achieve perfect alignment. Hence even after project each subjects fMRI images into a common space, the time courses from different subjects may not map to same anatomical regions even they share same spatial coordinates. Together with the individual random thoughts during scan, these difference results in the inter-subject variability of the functional networks that must be modeled explicitly. Some methods (Yeo et al., 2011) do not produce estimates of individual functional connectivity map. Such individual estimates are an important step in understanding functional networks not just on average, but also how these networks vary across individuals. Other methods such as the group ICA (Calhoun et al., 2001a,b), does not have an explicit statistical model of the variability between the group and subject component maps. (((we should add here that: spatial smoothing is often used for better matching between the subject's functional data)))
- In the group analysis, subjects may exhibit similar but not exactly same spontaneous BOLD fluctuation. Current group studies typically first identify each subject's

Email addresses: weiilbu@sci.utah.edu (Wei Liu), suyash@sci.utah.edu (Suyash P. Awate), fletcher@sci.utah.edu (P. Thomas Fletcher)

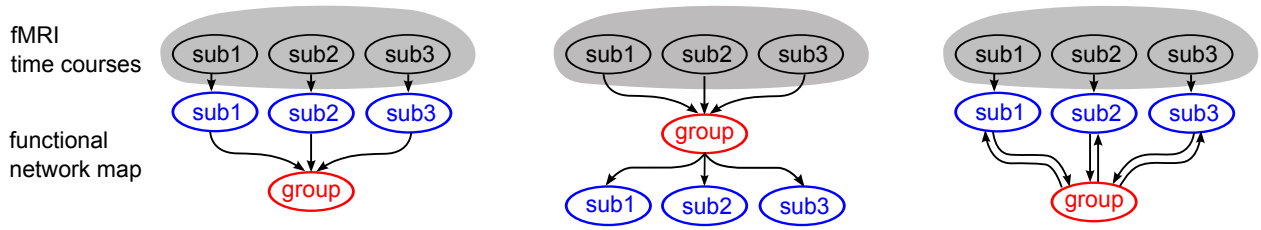

Figure 1: Comparison of segmentation methods for group study of fc-fMRI. Most methods are one-way approach, either in a bottom-up order in the left diagram, or the top-down order in the middle diagram. In left diagram, each subject's functional network map are estimated independent of other subjects, and a summary of the group's map is estimated from subject's map. In the middle diagram, a group map is obtained first, and it is used as a guide or a prior information for subject maps' estimation. Our methods aim at a joint estimation of both level of network maps, where group and subjects map help each other in a bidirectional flow, as show in the right diagram.

connectivity separately regardless of other subjects, and then estimate a *pooled* summary of the group connectivity map (Van Den Heuvel et al., 2008; Craddock et al., 2011). Or, a group map is estimated first, then is back-reconstructed to obtained the subject network map (Calhoun et al., 2001a). See figure 1 for an illustration. Such approaches are sub-optimal, since estimation of one subject's connectivity does not benefit from other subjects. From a Bayesian point of view, once the population distribution is known, it can help each subject's estimation as a prior. Subjects network estimates also gives feedback on group estimation.

We need a data-driven, unified probabilistic framework and put the connectivity variables of both group and subjects jointly into this model. Inference can be made from the posterior distribution of the variables in both (subject and group) levels given the observed time series data.

In this paper we propose a Bayesian hierarchical model including both subject and group levels in order to identify the functional networks from rs-fMRI data. We assume a group network label map that acts as a prior to the label maps for all subjects in the population. This Bayesian perspective provides a natural regularization of the estimation problem of a single subject using information from the entire population. The variability between the subjects and group are taken into account through the conditional distributions between group and subjects. The within-subject spatial coherence is modeled by a Markov Random Field (MRF). [[talk about the disadvantage of spatial smoothing, and our methods mitigate that.]]

[[Give a few examples of applications of MRF in neuroimaging, or give a one or two sentences of introduction for MRF.]] MRF is a statistical tool for modeling the dependency of multiple hidden variables on a graph. Here we extend the conventional concept of the spatial regularization with MRF by redefining the graph including the network variables in the group and subject levels.

Furthermore, the full Bayesian model provides us an opportunity to study the variability of the functional network by inference of the posterior. Besides the traditional variance and confidence interval analysis, the mode of spatial variability can be inferred from multivariate analysis. For the first time, to our best knowledge, it would be possible to visualize how functional homogeneous regions change along the principal direction of their posterior variability, and compare these change across subjects.

[[Optimization of MRF. Past method, our method, exact optimization intractable, approximation]] Both the group clustering and subject clusterings are estimated simultaneously with a Monte Carlo Expectation Maximization (MCEM) algorithm. The model is data-driven in that all parameters, regularized by two given hyper-parameters, are estimated from the data, and the only parameter that must be specified is the number of networks.

[[Talk about how the experiments is designed, both the synthetic data and real data. Test-retest data.]]

[[talk about this is the extension of the preliminary results reported in miccai 2012. The difference: alternative formulation of the graphical model??? parameter estimation?]]

[[Our model is essentially a hidden Markov model with a carefully defined graph that represent the hierarchical structure of the group fMRI data.]]

[[Need a bullet list of the contribution of the paper]]

[[what does each section talk about in detail]]

2. Hierarchical Model for Functional Networks

We define each subject's network label map as a Markov Random Field (MRF) with the neighborhood structure given by a regular lattice. The statistical dependency between adjacent voxels acts as a prior model favoring spatial coherence of functional regions. An additional group label map is defined on top of all subject label maps. The group label map has the same Markov structure as the individuals, again to encourage spatial coherence of the functional regions in the group level. In addition, each voxel in the group map is connected to the corresponding voxel in each subject. These connections model the relationship between the group and the individuals. Hence, all voxels of subjects and group label map are jointly connected into a single MRF. See figure 2 for an illustration.

More specifically, define a undirected graph $\mathbf{G} = (\mathbf{V}, \mathbf{E})$. The set of vertices $\mathbf{V} = (\mathbf{V}_G, \mathbf{V}_1, \cdots, \mathbf{V}_J)$ includes all the voxels in the gray matter volumes $\mathbf{V}_j$ of all J subjects as well as those in the group volume $\mathbf{V}_G$. An edge $(s,t) \in \mathbf{E}$ is defined if 1) $s \in \mathbf{V}_G, t \in \mathbf{V}_j$ and s, t have the same physical coordinates,

or 2) $s,t \in \mathbf{V}_G$, and s,t are spatial neighbors, or 3) $s,t \in \mathbf{V}_j$, and s,t are spatial neighbors. In our model we use 6-neighbor system in a 3D volume image. On each node $s \in \mathbf{V}$, a discrete random variable $y_s \in \mathbf{L} = \{1, \cdots, L\}$ is defined to represent the functional network labels.

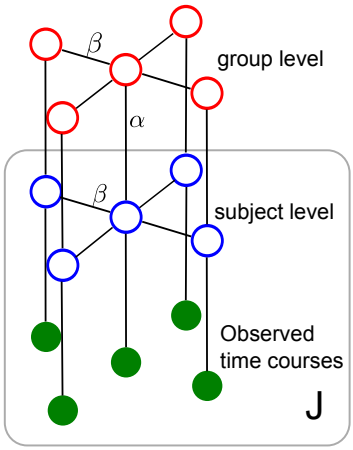

Figure 2: Markov random filed is define on a graph that includes the voxels of all subjects as well as the group map. An edge is added when two voxels are spatial neighbors within one subject, or is added between two voxels if they are at different levels (that is, one in group map, the other in subject map), and share same physical coordinates. The square on subject level and time courses repeats the nodes in the square J times, representing all the subjects nodes. For the cross-level connections, only the center voxels connection are shown, though the connections exist on all other voxels.

2.1. MRF Prior

MRF is a principal method for modeling the spatial context information. Here we use it as a prior distribution of network label set $Y = \{y_s \in \mathbf{L} | s \in \mathbf{V}\}$. Our MRF prior on the hierarchical model is essentially a Potts model (Potts, 1952) with different weights on the within-subject connections and the cross-level connections. Because of the equivalence of MRFs and Gibbs fields (Besag, 1974), we define the prior as $p(Y) = \frac{1}{Z}\exp\{-U(Y)\}$, where the energy function $U(Y)$ is given by the sum of potential functions defined on the cliques:

$$U(Y) = \sum_{s,r \in \mathbf{V}_G} \beta\psi(y_s, y_r) + \sum_{j=1}^{J} \left(\sum_{s \in \mathbf{V}_G, \tilde{s} \in \mathbf{V}_j} \alpha\psi(y_s, y_{\tilde{s}}) + \sum_{s,r \in \mathbf{V}_j} \beta\psi(y_s, y_r) \right). \quad (1)$$

Here ψ is a binary function taking zero when the two inputs are equal and one otherwise, and α and β are parameters determining the strength of the connections. (s,r) is a pair of neighboring voxels at the same level of the hierarchy (see figure 2), and $(s,\tilde{s})$ is a pair of neighboring voxels at different level in the hierarchy, but share same physical coordinates. However, according to our definition of the graph including two levels, both (s,r) and $(s,\tilde{s})$ are neighbors, or, cliques in the MRF, and can be treated equally in the following inference procedure.

This regularization encodes two physiologically meaningful *a priori* assumptions on the functional networks under investigation: 1) The networks are spatially coherent within single subject map and within group map. This is modeled by the β potential term . 2) The networks are similar between subjects, and therefore between the group and subjects. This is modeled by the α potential term. The proposed energy function represents both priors without introducing blurring artifacts.

[[need to add more advantage of the graph and the model]]

2.2. Likelihood Model

In the generative model, for any individual subject, the observed time course at each voxel is assumed to be generated from a distribution conditioned on the network label at that voxel. In fMRI analysis the time series at each voxel is usually normalized to be zero mean and unit norm, so the analysis is robust to shifts or scalings of the data. This results in the data being projected onto a high-dimensional unit sphere. After normalization, the sample correlation between two time series is equal to their inner product. Therefore the problem of clustering voxels with high correlations can be reformulated to finding clusters with small within-cluster distance on the sphere, where the distance is defined by the inner product of two vectors.

We use $X = \{(x_1, \ldots, x_N) |\ x_s \in S^{p-1}\}$ to denote the set of *normalized* time series on p-sphere. Given Y, the random vectors x_s are conditionally independent, hence $\log p(X|Y) = \sum_{s \in \mathbf{V}_j} \log p(x_s|y_s)$. The likelihood function $p(x_s|y_s)$ is naturally modeled by a von Mises-Fisher (vMF) distribution

$$f(x_s | y_s = l; \mu_l, \kappa_l) = C_p(\kappa_l) \exp\left(\kappa_l \mu_l^\intercal x_s\right), \quad x_s \in S^{p-1}, \quad l \in \mathbf{L}, \quad (2)$$

where for the network cluster label l, μ_l is the mean time course, $\kappa_l \geq 0$ is the *concentration parameter*, and C_p is the normalization constant. The larger the κ_l, the greater the density concentrated around the mean. Since (2) depends on x only through $\mu^T x$, the vMF distribution is unimodal and rotationally symmetric around μ.

3. Bayesian Inference

We solve the inference problem in a *maximum a posteriori* (MAP) framework. That is, given the observed time course data X, we estimate the posterior mode of $p(Y|X)$. This consists of the following components.

3.1. Parameter Estimation

In this data-driven model, we propose to estimate the parameters $\theta = \{\alpha, \beta, \kappa, \mu\}$ from the data using an Expectation Maximization (EM) algorithm. However, the high-dimensionality and dependency between spatially adjacent voxels in MRF make it infeasible to obtain a closed form

solution of the expectation of $\log p(X,Y)$ with respect to $p(Y|X)$. Here we propose to approximate the expectation using Monte Carlo EM (MCEM), in which a set of samples, $(Y^1,\cdots,Y^M)$, generated from density $p(Y|X)$ is used to approximate the expected value by the empirical average $\frac{1}{M}\sum_{m=1}^{M}\log p(X,Y^m)$.

3.2. Gibbs Sampling

[[talk about conditional independency in order to introduce the Gibbs sampling]]

Gibbs sampling converts a multivariate sampling problem into a consecutive univariate sampling, hence is well adapted to draw the Monte Carlo samples from $p(Y|X)$. In our hierarchical structure, the sampling procedure is also done in a multi-level fashion. At the image level, a sample of the group label map, Y_G^m, is drawn given the previous subject label map, Y_j^{m-1}. Next, a sample for each subject map, Y_j^m, is generated given the current group label map, Y_G^m. At the voxel level, we draw samples of the label y_s given the rest of nodes fixed, and update $y_s, \forall s \in V$. The conditional probability used to generate samples at the group and subject voxels are given as

$$p(y_s|y_{-s},X) \propto \frac{1}{Z_s}\exp\{-U(y_s|y_{-s})\}\cdot p(x_s|y_s) \tag{3}$$
$$= \frac{1}{Z_s}\exp\{-U_p(y_s|y_{N(s)},x_s)\}$$
$$U_p = \alpha\sum_{j=1}^{J}\psi(y_s,y_{\tilde{s}}^j)+\beta\sum_{r\in N(s)}\psi(y_s,y_r), \forall s\in\mathbf{V}_G, \tag{4}$$
$$U_p = \alpha\psi(y_s,y_{\tilde{s}})+\beta\sum_{r\in N(s)}\psi(y_s,y_r) - \kappa_l\mu_l^{\mathsf{T}}x_s - \log C_p,\ \forall s\in\mathbf{V}_j, \tag{5}$$

where $-s$ is the set of all nodes excluding s, Z_s the normalization constant, U_p is the posterior energy, and $N(s)$ is the set of neighbors of s. $y_{\tilde{s}}^j$ in (4) is the network label of subject j's voxel with the same physical coordinates with s, and $y_{\tilde{s}}$ in (5) is the label of group's voxel with the same physical coordinates with s. Because of the dependency on previous samples, the sequence of samples will be a Markov Chain, hence our method falls into Markov Chian Monte Carlo (MCMC) sampling. After a sufficient burn-in period, a series of samples $Y^m, m = 1\cdots M$ is saved for approximating the expectation $\mathbb{E}[\log p(X,Y)]$.

3.3. Pseudo likelihood

To evaluate $\log p(X,Y^m;\theta) = \log p(Y^m;\theta) + \log p(X|Y^m;\theta)$ as a function of θ, we face the difficulty of evaluating the partition function Z in $p(Y^m)$. In practice the Gibbs field is approximated by pseudo-likelihood (Besag, 1974), which is defined as the product of the conditional distribution $p(y_s|y_{-s}), \forall s\in\mathbf{V}$. Therefore the energy function can be written as

$$\begin{aligned}U(Y) &\approx \sum_{s\in\mathbf{V}}U(y_s|y_{-s})\\ &= \sum_{s\in\mathbf{V}_G}\left(\alpha\sum_{j=1}^{J}\psi(y_s,y_{\tilde{s}}^j)+\beta\sum_{r\in N(s)}\psi(y_s,y_r)\right)\\ &+\sum_{j=1}^{J}\sum_{s\in\mathbf{V}_j}\left(\alpha\psi(y_s,y_{\tilde{s}})+\beta\sum_{r\in N(s)}\psi(y_s,y_r)\right),\end{aligned}$$

where $y_{\tilde{s}}^j$ and $y_{\tilde{s}}$ has the same definition with (4) and (5).

3.4. Hierarchical MRF algorithm using MCEM

With all the preparation above, parameter estimation can be done by maximizing $\frac{1}{M}\sum_{m=1}^{M}\log p(X,Y^m)$. The α and β in the MRF prior can be optimized by maximizing $\frac{1}{M}\sum_{m=1}^{M}\log p(Y^m)$ with a Newton-Raphson method. We assume a Gaussian prior distribution on α with hyperparameters μ_α and σ_α, which is given manually and does not have significant impact on the model. In order for MCMC sampling to converge quickly to the posterior, we need a reasonably good initial network label map. Here the K-means clustering on a concatenated group dataset is used for the initial maps of both the group and subjects. After the EM parameter estimation iterations are done, an Iterated Conditional Modes (ICM) on the current sample map gives the final label maps. Putting this all together, the groupmrf algorithm to estimate the group and individual label maps is given in Algorithm 1.

Algorithm 1: Monte Carlo EM for group MRF

Data: Normalized fMRI, initial group label map
Result: Group label map Y_G, subjects label map $Y_j, j\in\{1,\cdots,J\}$, parameters $\{\alpha,\beta,\mu,\sigma\}$
while $\mathbb{E}[\log p(X,Y)]$ *not converge* **do**
  repeat
    foreach $s\in\mathbf{V}_G$ **do** Draw consecutive samples of y_s from $p(y_s|y_{-s},X;\theta)$ using (4) ;
    foreach $j = 1\ldots J$ **do**
      foreach $s\in\mathbf{V}_j$ **do** Draw consecutive samples of y_s from $p(y_s|y_{-s},X;\theta)$ using (5) ;
    Save sample Y^m after B burn-ins;
  until $B+M$ *times*;
  foreach $l = 1\cdots L$ **do**
    Estimate $\{\mu_l,\kappa_l\}$ by maximizing $\frac{1}{M}\sum_{m=1}^{M}\log p(X|Y^m)$;
  Estimate $\{\alpha,\beta\}$ by maximizing $\frac{1}{M}\sum_{m=1}^{M}\log p(Y^m)$;
Run ICM on current samples to estimate final label maps.

4. Experiments

[[This section should cover synthetic data, both 7 and 17 networks of real data?, bootstrap, posterior map. Do we

want to split experiments and results in two sections? For now let me split them into two sections.]]

4.1. Synthetic Data

Since our groupmrf is an extension of existing clustering methods, we compare our method with two other clustering method: K-Means and normalized-cuts (N-Cuts), as well as two degenerated version of groupmrf algorithm: the groupmrf1 and groupmrf2.

The K-Means algorithm, as a simple and fast clustering method, was applied on paradigm fMRI study in Baumgartner et al. (1998), and later used by Bellec et al. (2010) for bootstrap analysis of a group of rs-fMRI. In our experiment, the distance metric of *K-Means* is defined as $1 - x_s^\intercal x_r$. we apply K-Means on each subject fMRI time courses 20 times, and choose one segmentation map with the minimal ratio of inter-cluster sum-of-distance and intra-cluster sum-of-distance. For group study, we obtain a *group* dataset by concatenating all subjects time courses. K-Means is applied 20 times on this group dataset again to obtain a group network label map. the initial cluster centers for both subject and group K-Means are chosen randomly but maximizing the between-center distance (Arthur and Vassilvitskii, 2007).

Normalized cuts treats the fMRI image segmentation as a graph partitioning problem, and propose a global criterion Shi and Malik (2000). It is used by Van Den Heuvel et al. (2008); Craddock et al. (2011) for group rs-fMRI study. following Van Den Heuvel et al. (2008), we also apply N-Cuts in two stages. First N-Cuts is run on each subject's affinity matrix, as computed as the pairwise correlation between time courses. A second N-Cuts is applied on a group affinity matrix, which is computed by summing up all of the subjects' affinity matrix derived from their segmentation map. We use the Ncutclustering_9 toolbox Shi and Malik (2000), a newer version of the one used in Van Den Heuvel et al. (2008).

groupmrf1 is different with groupmrf in that it does not have the within-subject and within-group links (i.e. the β term in (1)). groupmrf2 is different with groupmrf in that it does not have the between-level links (the α terms in (1)). Hence groupmrf2 amounts to define a MRF on each single subject and estimate each subject's network label map indepedent of other subjects. Such strategy is equivalent to the hidden Markov model we proposed in Liu et al. (2011). Both groupmrf1 and groupmrf2, as degenerated case of groupmrf, serve to check if a simplified model be able to achieve same or better performance compared to the proposal model.

It is infeasible for our method, a clustering-based method to compare a decomposition-based method, such as group ICA, since the former is more of a disjoint partition and the later has overlapping membership assigned to each voxel. However, a posterior map from our method will be shown for a qualitative comparison with ICA component map.

We simulate synthetic rs-fMRI data in two steps: generating functional network pattern maps, and generating time courses. Since we assume the network map a MRF, we generate the group network map by random sampling a Potts model with $\beta = 2.0$ and 500 scans. Given the group map, a subject map is generated with $\alpha = 0.5$ and $\beta = 2.0$ based on the definition of distribution in equation 1. The subject map generation steps are repeated 25 times to obtain 25 subjects in total. To simulate the time courses given the network label map, we first generate mean time courses $\mu_l, l = \{1, \ldots, L\}$ from a first-order auto-regressive process $x_t = \varphi x_{t-1} + \varepsilon$ with $\varphi = 0.8$ and noise variance $\sigma_\varepsilon = 0.1$. The sample correlation between the mean time series is in the range of $(-0.15, 0.3)$. Then, independent Gaussian white noise of $\sigma^2 = 50$ are added on each cluster's mean time course.

5. Results

5.1. Synthetic Data

The simulated true networks patterns map and the estimated maps from various methods are shown in figure 3. These maps show that methods such as K-Means, N-Cuts and groupmrf1 are less sensitive to noise as these methods apply on data with spatial smoothing, which however also results in the losing of some fine scale patterns. groupmrf2's single-level MRF helps to recover some fine-scale patterns, but one subject's estimation could not benefit from other subjects due to the lack of group level. Our proposed groupmrf is able to recover most details of the networks. As a quantitative comparison, the Rand index between the true group label map and estimated map from K-Means, N-Cuts, groupmrf1 and groupmrf are 92.9%, 91.8%, 94.2% and 95.6% respectively. The average Rand indices of all subjects label maps for these methods are given in figure 4. It is noted due to the multi-level regularization, groupmrf1's Rand index has less variance and thus more consistent compared to K-Means and N-Cuts, though the average is not significantly improved. groupmrf2 improves the overall segmentation accuracy, but has larger variance across subjects. This is because the information of one subject's good estimates does not propagate to other subjects, due to the lack of hierarchy in the model. And at last our groupmrf with a full-fledged hierarchical MRF model, has highest average accuracy, as well as least variance across subjects.

5.2. Real Data

Appendix A. Plan for Experiments

Zuo et al. (2010) is the group that first use New York test-retest dataset. They use intraclass correlation (ICC) to compute the reliability at each voxel. Their goal is to compare the ICC for different independent components (ICs). Our goals are to show the hierarchical MRFs can get more consistent maps compared to other methods. There are a few questions we need to address:

1) What methods that the groupmrf can compare with?

2) What measurement to use? There are again a few options:

If we use ICC, we can compute the ICCs for each voxel, by using all three sessions of all subjects. ICC is usually used

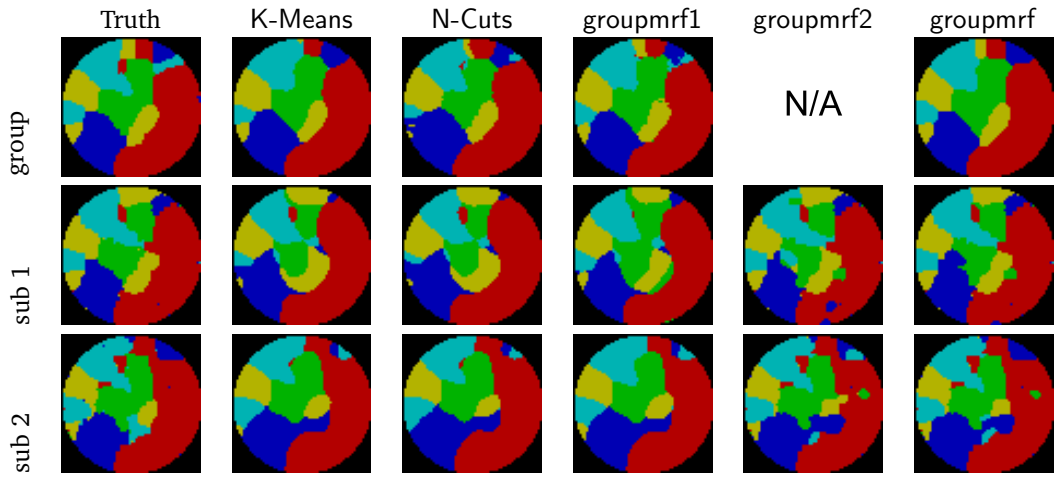

Figure 3: Compare the group and subject network label maps estimated from the proposal groupmrf and other methods. Only two from all 25 subjects map are shown as a illustration. The simulated true maps, with regions of different sizes, and the average 90% Rand index between group and subjects, well represent the functional network patterns in real data.

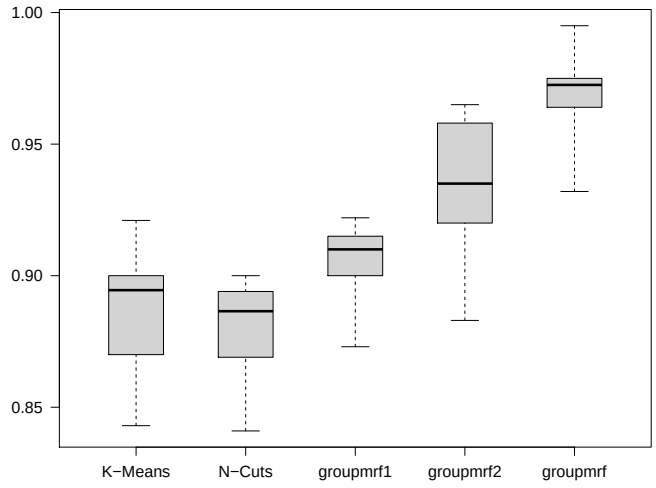

Figure 4: A box-Whisker plot of the Rand index value of all subjects label map for various methods. The bottom and top of the boxes are the 25th and 75th percentile, and the whiskers extends to the whole range of the data.

on a continuous random variable, so we can use the postior probability that we estimate from MC samples as the r.v in ICC (I know it is not the best definition). To be specific, following Zuo et al. (2010), we define y_{ij} as the jth measurement made on the ith subject. Here $j = 1, 2, 3$. We have the following decomposition

$$y_{ij} = \mu + p_i + t_j + \epsilon_{ij}$$

It is possible that we use the indicator variable of each MC sample as the measurement, but that will add another level of variance and make things more complicated. So we just use the posterior probability as a single summary of all MC samples for each subject in each session.

For groupmrf, we will get a ICC map for each component. We will compute the same map for other clustering method we want to compare with. Now we need to compute a single number, i.e. a summary statistics from the ICC map, for the purpose of comparison.

Instead, we can use *Rand index* (RI), *normalized Rand index* (NRI) and *normalized mutual information* (MNI) as the measurements for comparing two clusterings. Rand index is applied on discrete segmentations of a dataset or image. NRI is a normalize version of RI such that it has zero expectation. All 3 measures give single number for the similarity of two clusterings. We can do the following steps

- We can show groupmrf has more consistent segmentations between subjects, but within session. We compute the pairwise similarity for all subject in one session, and average the RI values to get a mean RI. For other clustering methods, also compute the mean RI. Comparing these mean RI will show groupmrf's advantage. The procedure can repeat for all 3 sessions.

  This comparison will show the groupmrf's advantage for sure.

- We can show groupmrf gives more consistent segmentation across sessions. For each subject, we compute the similarity for each pair of 3 sessions, and get $\binom{3}{2} = 3$ similarity numbers. Then I can average the numbers for all subjects and get a single similarity value. We do the same thing for other clustering methods.

  This comparison will probably show groupmrf's advantage.

- We can also compare the group map estimated from groupmrf and other methods. For our method, we have 3 group maps, and we compute the pairwise similarity for each pair of them. If other methods also has 3 group maps, we can compute the similarity also.

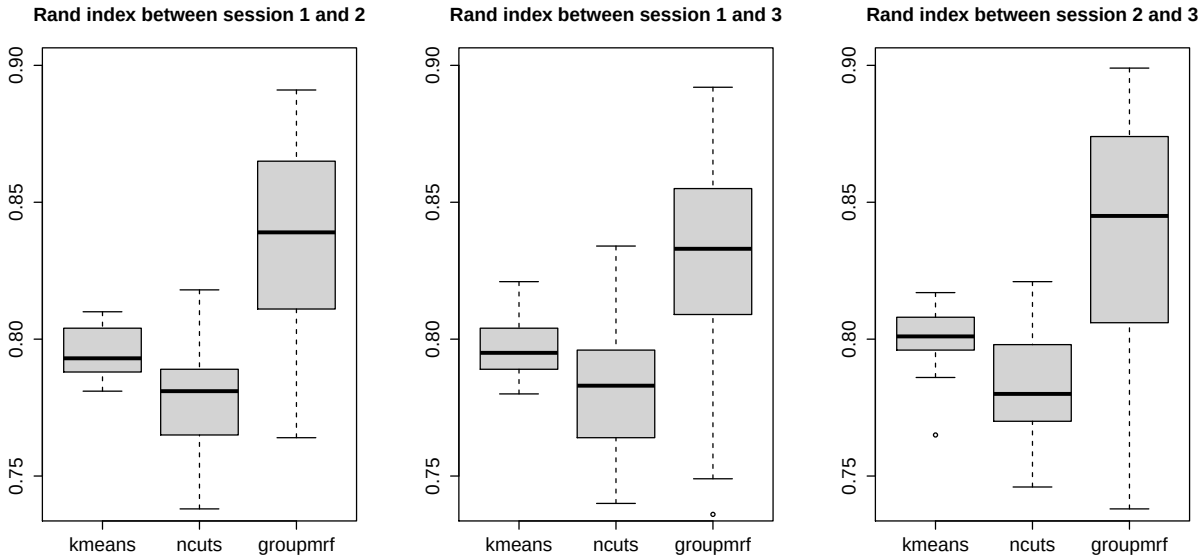

Figure 5:

Our method probably will not stand out in this experiment.

References

Arthur, D., Vassilvitskii, S., 2007. k-means++: The advantages of careful seeding, in: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, Society for Industrial and Applied Mathematics. pp. 1027–1035.

Baumgartner, R., Windischberger, C., Moser, E., 1998. Quantification in functional magnetic resonance imaging: fuzzy clustering vs. correlation analysis. Magnetic Resonance Imaging 16, 115–125.

Bellec, P., Rosa-Neto, P., Lyttelton, O., Benali, H., Evans, A., 2010. Multi-level bootstrap analysis of stable clusters in resting-state fMRI. Neuroimage 51, 1126–1139.

Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society. Series B (Methodological) 36, 192âĂŞ–236.

Calhoun, V., Adali, T., Pearlson, G., Pekar, J., 2001a. A method for making group inferences from functional mri data using independent component analysis. Human brain mapping 14, 140–151.

Calhoun, V., Adali, T., Pearlson, G., Pekar, J., 2001b. Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. HBM 13, 43–53.

Craddock, R., James, G., Holtzheimer III, P., Hu, X., Mayberg, H., 2011. A whole brain fMRI atlas generated via spatially constrained spectral clustering. Human brain mapping .

Liu, W., Awate, S., Anderson, J., Yurgelun-Todd, D., Fletcher, P., 2011. Monte Carlo Expectation Maximization with Hidden Markov Models to Detect Functional Networks in Resting-State fMRI. Machine Learning in Medical Imaging , 59–66.

Potts, R., 1952. Some generalized order-disorder transformations, in: Proceedings of the Cambridge Philosophical Society, Cambridge Univ Press. pp. 106–109.

Shi, J., Malik, J., 2000. Normalized cuts and image segmentation. IEEE PAMI 22, 888–905.

Van Den Heuvel, M., Mandl, R., Pol, H., 2008. Normalized cut group clustering of resting-state FMRI data. PLoS One 3, e2001.

Yeo, B., Krienen, F., Sepulcre, J., Sabuncu, M., Lashkari, D., Hollinshead, M., Roffman, J., Smoller, J., Zöllei, L., Polimeni, J., et al., 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. Journal of Neurophysiology 106, 1125–1165.

Zuo, X., Kelly, C., Adelstein, J., Klein, D., Castellanos, F., Milham, M., 2010. Reliable intrinsic connectivity networks: test-retest evaluation using ica and dual regression approach. Neuroimage 49, 2163.