

Tractography Segmentation Using a Hierarchical Dirichlet Processes Mixture Model

Xiaogang Wang¹, W. Eric L. Grimson¹, and Carl-Fredrik Westin²

¹Computer Science and Artificial Intelligence Lab,
Massachusetts Institute of Technology, Cambridge, MA, 02139, USA

²Computational Research Laboratory, Childrens Hospital,
Harvard Medical School, Boston, MA 02115, USA

{xgwang,welg}@csail.mit.edu, westin@bwh.harvard.edu

Abstract. In this paper, we propose a new nonparametric Bayesian framework to cluster white matter fiber tracts into bundles using a hierarchical Dirichlet processes mixture (HDPM) model. The number of clusters is automatically learnt from data with a Dirichlet process (DP) prior instead of being manually specified. After the models of bundles have been learnt from training data without supervision, they can be used as priors to cluster/classify fibers of new subjects. When clustering fibers of new subjects, new clusters can be created for structures not observed in the training data. Our approach does not require computing pairwise distances between fibers and can cluster a huge set of fibers across multiple subjects without subsampling. We present results on multiple data sets, the largest of which has more than 120,000 fibers.

1 Introduction

Diffusion Magnetic Resonance Imaging (dMRI) is an MRI modality that has gained tremendous popularity over the past five years and is one of the first methods that made it possible to visualize and quantify the organization of white matter in the human brain in vivo. Extracting connectivity information from dMRI, termed “tractography”, is an especially active area of research, as it promises to model the pathways of white matter tracts in the brain, by connecting local diffusion measurements into global trace-lines. In neurological studies of white matter using tractography it is often important to identify anatomically meaningful fiber bundles. Similar fibers form clusters of points, where each cluster is identified as a “fiber bundle”.

In this paper, we propose a nonparametric Bayesian framework to cluster fibers into bundles. The 3D space of the brain is quantized into voxels. A bundle is modeled as a multinomial distribution over voxels and orientations. This probabilistically models the spatial variation of the pathways of fibers. The models of bundles are learnt from how voxels are connected by fibers instead of comparing distances between fibers. If two voxels are connected by many fibers, both of the voxels have large weights in the model of the same bundle, which means that they are on the same pathway of white matter tracts. Many existing approaches

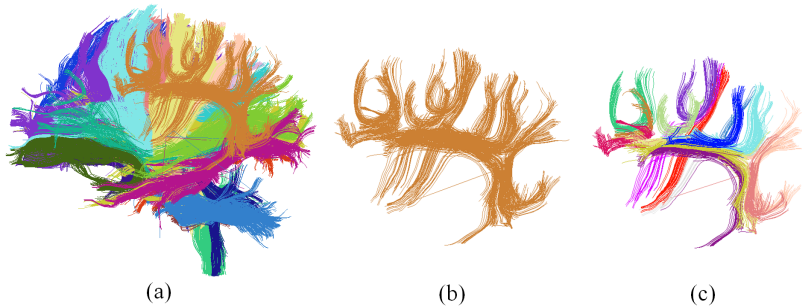


Fig. 1. An example of multiscale clustering. The spatial range of the whole brain is $200 \times 200 \times 200$. (a): The clustering result when the space is quantized into voxels of size $11 \times 11 \times 11$. The bundles correspond to structures at a large scale. (b): One bundle from (a). (c): The space is quantized into voxels of size $3 \times 3 \times 3$ and the bundle in (b) is further clustered into smaller bundles corresponding to structures at a finer scale.

have difficulty in determining the number of clusters and in clustering a very large set of fibers. Our approach automatically learns the number of clusters from data with a Dirichlet process (DP) prior [1]. While the space and time complexities of existing distance-based fiber clustering approaches are at least $O(M^2)$, where M is the number of fibers, the space complexity of our approach is $O(M)$ since it does not compute and store pairwise distances between fibers.

After the models of bundles have been learnt from training data without supervision, they are used as priors to cluster/classify new fibers. When clustering fibers of new subjects, our approach adapts the models of bundles to new data and creates new clusters for structures which are not observed in training data, instead of fixing the number of clusters as current methods do. Our framework can be extended to multiscale clustering. First cluster fibers using a large size of voxels and bundles correspond to structures at a large scale. Then each bundle can be further clustered using a smaller size of voxels, leading to structures at a finer scale. An example is shown in Figure 1. Multiscale clustering makes it easier for experts to identify white matter structures across different scales.

1.1 Related Work

Automatically clustering fibers has drawn a lot of attention in recent years. A typical framework is to first define a pairwise similarity/distance between fibers and to input the similarity matrix to standard clustering algorithms. Brun et al. [2] computed the Euclidean distances between 9-D fiber shape descriptors. Jonasson et al. [3] measured the similarity between two fibers by counting the number of points sharing the same voxel. Gerig et al. [4] proposed three measures related to Hausdorff distance: closest point distance, mean of closest distances and Hausdorff distance. Various clustering algorithms, such as hierarchical clustering (single-link and complete-link) [4, 5], fuzzy c-means [6], k-nearest neighbors [7], normalized cuts [2] and spectral clustering [3, 8] were used. Mean of closest distances and spectral clustering were popular among possible choices [8, 9].

These clustering algorithms required manually specifying the number of clusters or a threshold for deciding when to stop merging/splitting clusters, both of which are difficult to know especially when the data sets are complicated and noisy. Moberts et al. [9] showed that the performance of clustering varied dramatically when different numbers of clusters were chosen. To avoid this difficulty, O'Donnell and Westin [8] first chose a large cluster number for spectral clustering and then manually merged clusters to obtain models for white matter structures.

Another drawback of this framework is the high space and time complexities of computing pairwise distances between fibers when the data set is large. Whole brain tractography produces between 10,000 and 100,000 fibers per subject. It is difficult to compute a $100,000 \times 100,000$ similarity matrix or even to store it in memory. Some clustering algorithms, such as spectral clustering, need to compute the eigenvectors of this huge similarity matrix. This problem becomes more serious when clustering fibers of multiple subjects. The current solutions are to cluster only a small portion of the whole data set after subsampling or to do some numerical approximation based on the sampled subset [8]. However, important information from the full data set may be lost after subsampling.

Maddah et al. [10] proposed a probabilistic approach to cluster fibers without computing pairwise distances. They used a Dirichlet distribution¹ as a prior to incorporate anatomical information. This approach is different from ours. It used a parametric model, assuming that the number of clusters is known and required manual initialization of cluster centers. [10] required establishing point correspondence which was difficult, while our approach does not.

Dirichlet process mixtures (DPM) models were applied to medical image analysis in recent years because of their capability to learn the number of clusters and their flexibility to adapt to a wide variety of data. Adelino [11] used a DPM model for brain MRI tissue classification. In [12, 13] DPM models were used to model spatial brain activation patterns in functional magnetic resonance imaging. In [14], Jbabdi et al. modeled the connectivity profiles of a brain region as an infinite mixture of multivariate Gaussian distributions with a DP prior. To the best of our knowledge, our work is the first to use HDPM for tractography segmentation to automatically learn the number of clusters from data. Our approach is related to the work [15] where HDPM models were used for word-document analysis. HDPM was also used for trajectory analysis in visual surveillance [16].

2 Method

We begin by introducing DP in Section 2.1. In Section 2.2 and 2.3 we propose our HDPM model for clustering fibers and use Gibbs sampling for inference. In Section 2.3, we explain how to use the learned models of bundles as a prior to cluster new data.

¹ Dirichlet distribution is used as a prior of finite mixture models. These models can only well adapt to data from particular distributions. Dirichlet process in our approach is used as a prior in infinite mixture models. These models can well adapt to a wide variety of data.

2.1 Dirichlet Process

DP [1] is used as a prior to sample probability measures. It is defined by a concentration parameter α , which is a positive scalar, and a base probability measure H . A probability measure G randomly drawn from $DP(\alpha, H)$ is always a discrete distribution,

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k}, \quad (1)$$

which can be obtained from a stick-breaking construction [17]. In Eq (1), ϕ_k is a parameter vector sampled from H , δ_{ϕ_k} is a Dirac delta function centered at ϕ_k , and π_k ($\sum_{k=1}^{\infty} \pi_k = 1$) is a non-negative scalar constructed by $\pi_k = \pi'_k \prod_{l=1}^{k-1} (1 - \pi'_l)$, $\pi'_k \sim \text{Beta}(1, \alpha)$.

G can be used as a prior for infinite mixture models. Let $\{w_i\}$ be a set of observed data points. w_i is sampled from a density function $p(\cdot|\theta_i)$ parameterized by θ_i , and θ_i (which is one of the ϕ_k s in Eq (1)) is sampled from G . Data points sharing the same parameter vector ϕ_k are clustered together under this mixture model. Given parameter vectors $\theta_1, \dots, \theta_N$ of N data points, the parameter vector θ_{N+1} of data point w_{N+1} can be sampled from a prior by integrating out G ,

$$\theta_{N+1}|\theta_1, \dots, \theta_N, \alpha, H \sim \sum_{k=1}^K \frac{n_k}{N + \alpha} \delta_{\theta_k^*} + \frac{\alpha}{N + \alpha} H. \quad (2)$$

There are K distinct parameter vectors $\{\theta_k^*\}_{k=1}^K$ (identifying K components) among $\theta_1, \dots, \theta_N$. n_k is the number of points with parameter vector θ_k^* . θ_{N+1} can be assigned as one of the existing components (w_{N+1} is assigned to one of the existing clusters) or can sample a new component from H (a new cluster is created for w_{N+1}). The posterior of θ_{N+1} is

$$p(\theta_{N+1}|w_{N+1}, \theta_1, \dots, \theta_N, \alpha, H) \propto p(w_{N+1}|\theta_{N+1})p(\theta_{N+1}|\theta_1, \dots, \theta_N, \alpha, H). \quad (3)$$

It is likely for the Dirichlet process mixture (DPM) model to create a new component if existing components cannot well explain the data. There is no limit to the number of components. These properties make DP ideal for modeling data clustering problems when the number of clusters is not well-defined in advance.

2.2 Hierarchical Dirichlet Process Mixture Model

In probability theory, statistics, and machine learning, a graphical model is a graph that represents independences among random variables. The graphical model of our HDPM model is shown in Figure 2. There are M fibers and each fiber j has N_j points which are ordered sequentially. $o_{ji} = (\mathbf{u}_{ji}, \Delta\mathbf{u}_{ji})$ is the observed 3D coordinate $\mathbf{u}_{ji} = (x_{ji}, y_{ji}, z_{ji})$ and shift $\Delta\mathbf{u}_{ji} = \mathbf{u}_{ji+1} - \mathbf{u}_{ji}$ of point i on fiber j . The 3D space of the brain is uniformly quantized into voxels and shifts are quantized into three orientations $\Delta\mathbf{u}_1 = (1, 0, 0)^T$, $\Delta\mathbf{u}_2 = (0, 1, 0)^T$ and $\Delta\mathbf{u}_3 = (0, 0, 1)^T$. A codebook is built, in which codes (entries of the codebook)

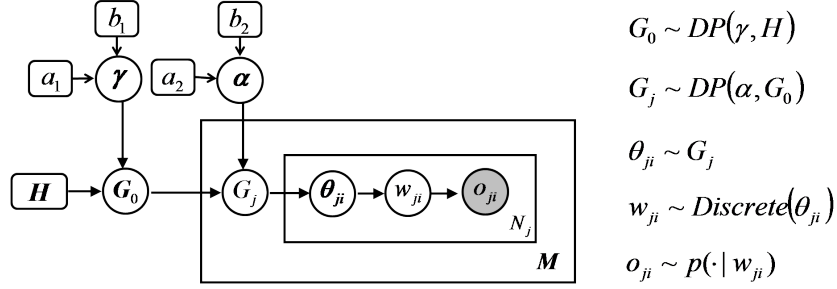


Fig. 2. The graphical model of our HDPM model. The right side list the distributions where the random variables are sampled from. G_0 is a prior on the whole data set. G_j is a prior on fiber j . Both G_0 and G_j are sampled from DP. θ_{ji} is the model of a bundle sampled for a point. w_{ji} and o_{ji} are the code and observation of a point.

are indices of voxels and orientations. Let \mathbf{u}_w be the centroid of the voxel and d_w be the index of the orientation vector corresponding to code w . Quantization is done in a probabilistic way,

$$p(o_{ji}|w) = p(\mathbf{u}_{ji}|\mathbf{u}_w)p(\Delta\mathbf{u}_{ji}|d_w), \quad (4)$$

$$p(\mathbf{u}_{ji}|\mathbf{u}_w) \propto \begin{cases} \cos^2\left(\frac{\|\mathbf{u}_{ji}-\mathbf{u}_w\|^2}{2R^2}\pi\right), & \|\mathbf{u}_{ji}-\mathbf{u}_w\| \leq R \\ 0, & \|\mathbf{u}_{ji}-\mathbf{u}_w\| > R \end{cases}, \quad (5)$$

$$p(\Delta\mathbf{u}_{ji}|d_w) \propto \begin{cases} 1, & d_w = \arg \max_d \frac{|\Delta\mathbf{u}_{ji} \cdot \Delta\mathbf{u}_d|}{\|\Delta\mathbf{u}_{ji}\| \cdot \|\Delta\mathbf{u}_d\|} \\ 0, & \text{otherwise} \end{cases}. \quad (6)$$

Since we do not distinguish the starting and ending points of a fiber, the sign of the correlation between $\Delta\mathbf{u}_{ji}$ and $\Delta\mathbf{u}_d$ is ignored in Eq (6). The statistical model ϕ_k of a bundle is a multinomial distribution over voxels and orientations. Optionally, if the symmetry across hemispheres is considered, we can do bilateral clustering as in [8]. Assuming that the brain is aligned and $x = 0$ is the midsagittal plane, we modify observed 3D coordinates as $\mathbf{u}_{ji} = (|x_{ji}|, y_{ji}, z_{ji})$ ignoring the signs of x coordinates. Thus, learnt models of bundles are symmetric to the midsagittal reflection.

A prior G_0 on the whole data set is sampled from a DP, $G_0 \sim DP(\gamma, H)$, where the base measure H is a Dirichlet distribution. $G_0 = \sum_{k=1}^{\infty} \pi_{0k} \delta_{\phi_k}$ is a infinite mixture in which components $\{\phi_k\}$ are models of bundles. For a fiber j , a prior G_j is sampled from a DP, $G_j = DP(\alpha, G_0)$. It was shown that in HDPM all the G_j share the same set of components $\{\phi_k\}$ as G_0 , however, they have different weights π_j over $\{\phi_k\}$, i.e. $G_j = \sum_{k=1}^{\infty} \pi_{jk} \delta_{\phi_k}$ [15]. Thus the models of bundles are learnt from all the fibers, and fibers have different distributions over bundles. For a point i on fiber j , the model θ_{ji} ($\theta_{ji} \in \{\phi_k\}$) of a bundle is sampled from G_j , $\theta_{ji} \sim G_j$. Its index of voxel and orientation w_{ji} is sampled from the model of a bundle, $w_{ji} \sim \text{Discrete}(\theta_{ji})$. Observation o_{ji} is sampled from $p(o_{ji}|w_{ji})$. Concentration parameters γ and α are sampled from gamma priors,

$\gamma \sim \text{Gamma}(a_1, b_1)$, $\alpha \sim \text{Gamma}(a_2, b_2)$. In Figure 2, H , a_1 , a_2 , b_1 and b_2 are hyperparameters. The clustering performance is quite robust to the choice of their values in a large range. $\{o_{ji}\}$ are observations. The remaining are hidden variables to be inferred. A fiber j is assigned to a bundle k with maximum π_{jk} .

The data likelihood is higher when the distribution of a fiber concentrates on fewer bundles instead of being uniform. So if two voxels are connected by many fibers, both of them have large weights in the model of the same bundle.

The size of voxels determines the scale of the structures to be learnt. Our framework can be extended to multiscale clustering. First cluster fibers using a large size of voxels and bundles correspond to structures at a large scale. Then each bundle can be further clustered using a smaller size of voxels, showing structures at a finer scale. Multiscale clustering makes it easier for experts to identify white mater structures across different scales.

2.3 Inference

We use the Gibbs sampling inference proposed in [15], which is based on Chinese restaurant franchise. First we introduce some notations. c_{ji} is the index of the bundle assigned to point i on fiber j . n_{jk} is number of points assigned to bundle k on fiber j . n_j is the number of point on fiber j . m_{kw} is the number of points with code w and being assigned to bundle k . m_k is the total number of points assigned to bundle k . n_{jk}^{-ji} , m_{kw}^{-ji} and m_k^{-ji} mean that they are statistics without counting c_{ji} . $H = \text{Dir}(h, \dots, h)$ is a flat Dirichlet prior. L is the size of the codebook.

During the sampling procedure, suppose that K models of bundles (clusters) have been created and assigned to data. Then,

$$G_0 = \sum_{k=1}^K \pi_{0k} \delta_{\phi_k} + \pi_{0u} G_u, \quad G_u \sim DP(\gamma, H) \quad (7)$$

The Gibbs sampling scheme proposed in [15] integrated out $\{\pi_{jk}\}$ and $\{\phi_k\}$ without sampling them. The posterior of c_{ji} is given by

$$p(c_{ji} | \{c_{j'i'}\}_{j'i' \neq ji}, \{w_{ji}\}, \{\pi_{0k}\}, \alpha) \propto \begin{cases} (n_{jk}^{-ji} + \alpha\pi_{0k}) \cdot \frac{n_{kw_{ji}}^{-ji} + h}{n_k^{-ji} + Lh}, & k \in \{1, \dots, K\} \\ \alpha\pi_{0u} \cdot \frac{1}{L}, & k \text{ is new} \end{cases} \quad (8)$$

This posterior is the product of two terms which explain how many points on fiber j are assigned to bundle k ($n_{jk}^{-ji} + \alpha\pi_{0k}$), and how well the code w_{ji} fits the model of an existing bundle $((n_{kw_{ji}}^{-ji} + h)/(n_k^{-ji} + Lh))$ or a flat distribution $(1/L)$ to create a new bundle. This shows that points on the same fiber tend to choose the same bundle. The posteriors of $\{\pi_{0k}\}$, γ and α involve more details of the Chinese restaurant franchise. They can be found in [15]. We need one more step to sample w_{ji} which is not observed in our model but observed in [15],

$$p(w_{ji} | o_{ji}, \{w_{j'i'}\}_{j'i' \neq ji}, \{c_{j'i'}\}) \quad (9)$$

$$\propto p(o_{ji} | w_{ji}) p(w_{ji} | \{w_{j'i'}\}_{j'i' \neq ji}, \{c_{j'i'}\}) \propto p(o_{ji} | w_{ji}) \frac{n_{c_{ji}w_{ji}}^{-ji} + h}{n_{c_{ji}}^{-ji} + Lh}. \quad (10)$$

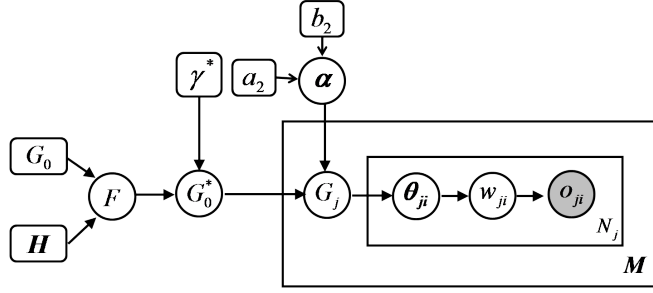


Fig. 3. The graphical model of our HDPM model for clustering testing data. F is used as the base measure to sample the prior G_0^* on the new data set. F includes the models of bundles in G_0 learnt from training data.

where $p(o_{ji}|w_{ji})$ is given by Eq (4).

Although $\{\phi_k\}$ and $\{\pi_{jk}\}$ are not explicitly sampled during the Gibbs sampling procedure, they can be estimated from any single sample,

$$\hat{\phi}_{kw} = \frac{m_{kw} + h}{m_k + Lh}, \quad \hat{\pi}_{jk} = \frac{n_{jk} + \alpha\pi_{0k}}{n_j + \alpha}.$$

The space complexity of our approach is $O(M)$. The time complexity of each Gibbs sampling iteration is $O(M)$. It is difficult to provide theoretical analysis on the convergence of Gibbs sampling. In practice, we stop burn-in when the data likelihood converges. From our empirical observation, the time complexity of our approach is much lower than $O(M^2)$. Recently some more efficient inference approaches, such as variational inference [18], and parallel sampling [19], have been proposed and applied to DPM and HDPM models. In future work, we will study how to improve the inference of our model using these schemes.

2.4 Clustering New Data

After the models of bundles have been learnt from training data without supervision, we can fix G_0 and the number of clusters to classify new fibers. Thus our model is converted to a parametric model.

Optionally, we can use the models learnt from training data as priors to cluster instead of classifying new data. The pre-learnt models can adapt to new data, and new clusters can be created for structures not observed in the training data. Our HDPM model for clustering testing data is shown in Figure 3. Suppose that G_0 represented in Eq (7) has been learnt from training data and K clusters are created. A prior G_0^* on testing data is to be learnt. Different from the model shown in Figure 2, where G_0 is generated from a DP with a flat base measure H , G_0^* is generated from $DP(\gamma^*, F)$, where the based measure F is constructed from G_0 and includes models learnt from training data.

$$F = \omega^* \sum_{k=1}^K \hat{\pi}_{0k} \delta_{\phi_k^*} + (1 - \omega^*)H \quad (11)$$

F is composed of two parts: the models learnt from training data and a flat prior. ω^* is a scalar between 0 and 1. $\{\hat{\pi}_{0k}\}$ are normalized weights in G_0 ,

$$\hat{\pi}_{0k} = \frac{\pi_{0k}}{\sum_{k'=1}^K \pi_{0k'}}.$$

This assumes that before observing any testing data, there already exist K models of bundles $\{\phi_k^*\}_{k=1}^K$. However, instead of letting ϕ_k^* be equal to ϕ_k in Eq (7), we sample ϕ_k^* from a Dirichlet distribution choosing ϕ_k as prior,

$$\phi_k^* \sim \text{Dir}(\xi_k^* \cdot \phi_k + H),$$

where ξ_k^* is a positive scalar. Thus the models of bundles can adapt to testing data instead of being fixed.

The choice of γ^* , ω^* and ξ_k^* controls how much the models learnt from the training data affect the clustering of testing data. The two extreme cases are that the pre-learnt models have no effect on clustering new data ($\omega^* = 0$, $\xi_k^* = 0$) and that the models learnt from new data are exactly the same as those learnt from training data ($\gamma^* = \infty$, $\omega = 1$, $\xi_k^* = \infty$).

Suppose there are K^* models of bundles assigned to testing data. Then an explicit construction of G_0^* is given by

$$G_0^* = \sum_{k=1}^K \pi_{0k}^* \delta_{\phi_k^*} + \sum_{k=K+1}^{K^*} \pi_{0k}^* \delta_{\phi_k^*} + \pi_{0u}^* G_{0u}^*. \quad (12)$$

Models $\{\phi_k^*\}_{k=1}^K$ have been seen in training data. They are sampled from priors $\text{Dir}(\xi_k^* \cdot \phi_k + H)$ and are updated using testing data. $\{\phi_k^*\}_{k=K+1}^{K^*}$ are new models not found in training data. They are sampled from a flat prior $\text{Dir}(H)$. The remaining parts are the same as described in Section 2.2.

3 Results

We evaluate our approach on multiple data sets. The spatial range of the whole brain is roughly $200 \times 200 \times 200$. The size of voxels is $11 \times 11 \times 11$. We choose the hyperparameters in Figure 2 as $a_1 = a_2 = b_1 = b_2 = 1$, $h = 0.3$. We do bilateral clustering. Running on a computer with 3GHz CPU, it takes around one minute to cluster 1,000 fibers and around four hours to cluster 60,000 fibers.

The first data set has 3,152 fibers with ground truth. They are manually labeled to six anatomical structures. Figure 4 (a)-(d) plots the clustering results of our approach and a spectral clustering approach, compared with the ground truth. Colors are used to distinguish clusters. Since clusters may be permuted in different results, the meaning of colors is not consistent across different results. The spectral clustering approach uses the mean of closest distances as the distance measure, which was found the most effective in previous studies [9, 8]. The clustering result of our approach is close to the ground truth. Although the correct number of clusters has been set, two anatomical structures are merged in

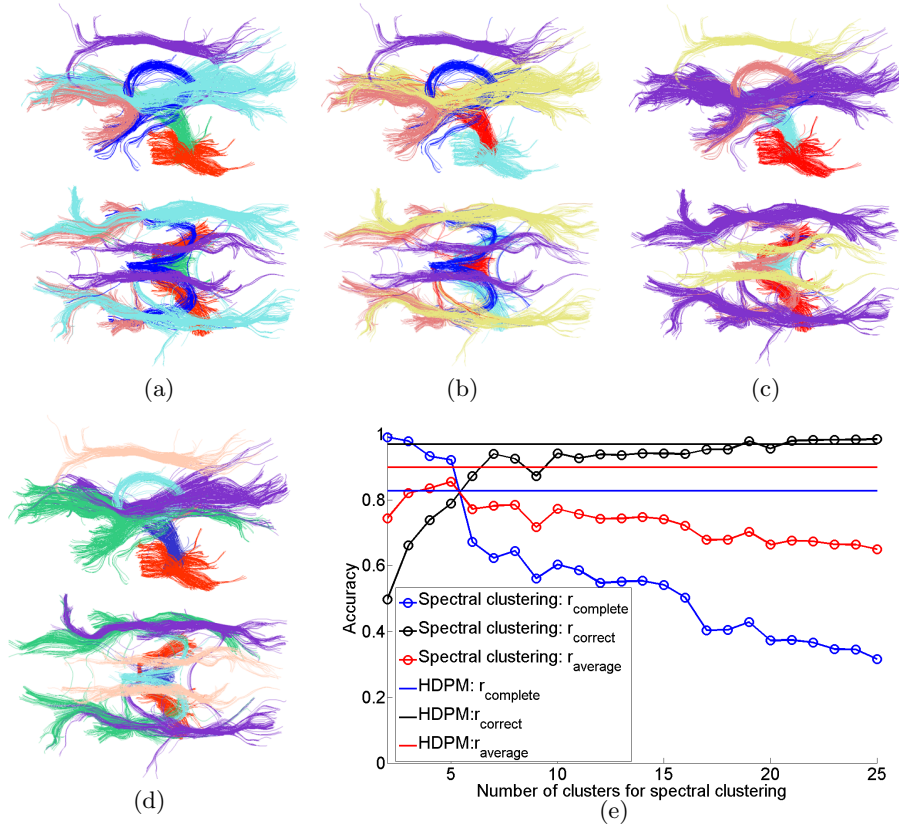


Fig. 4. Compare the results of two clustering approaches with the ground truth on a data set with 3,152 fibers. Two views are plotted for each result. (a) Ground truth. (b) Our approach. (c) Spectral clustering when the number of clusters is 6. (d) Spectral clustering when number of clusters is 7. (e) The accuracies of completeness and correctness of spectral clustering and our approach (HDPM).

the result of the spectral clustering approach. A few outlier fibers form a small cluster. As the number of clusters increases to 7, the two anatomical structures still cannot be separated, instead, another structure splits into two clusters.

There are two important aspects, called *correctness* and *completeness*, to be considered when comparing a clustering result with the ground truth [9]. Correctness implies that fibers of different anatomical structures are not clustered together. Completeness means that fibers of the same anatomical structures are clustered together. Putting all the fibers into the same cluster results in 100% completeness and 0% correctness. Putting every fiber into a singleton cluster results in 100% correctness and 0% completeness. To measure correctness, we randomly sample 5,000 pairs of fibers which are in different anatomical structures according to the ground truth and calculate the accuracy ($r_{correct}$) of they are also in different clusters according to the clustering result. To measure

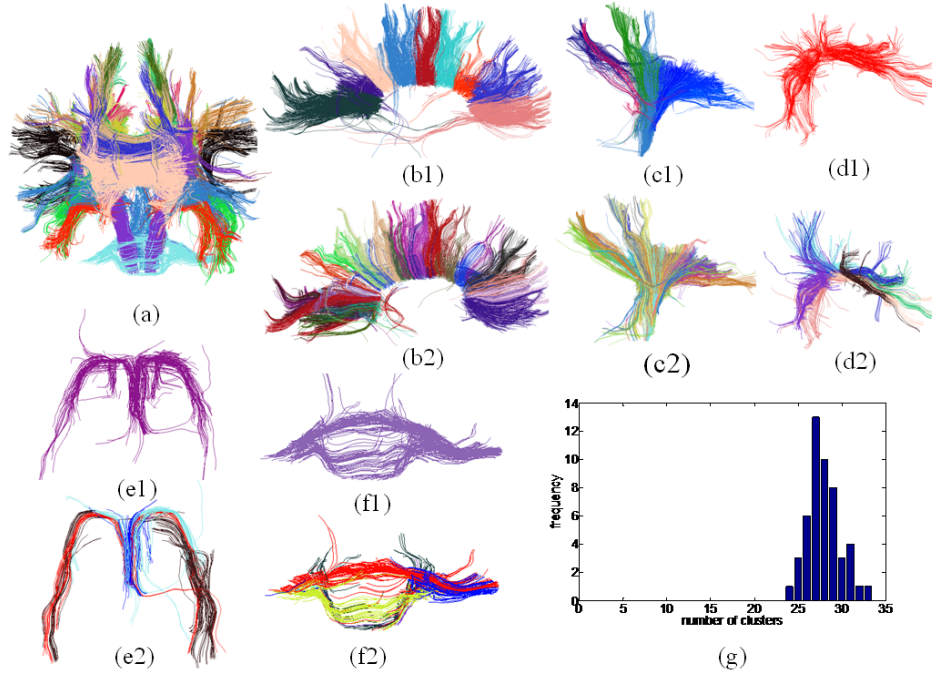


Fig. 5. Compare results of our approach and the approach proposed in [8], in which experts manually merged the clusters from spectral clustering to obtain anatomical structures. (a) Clustering all the fibers using our approach. (b1)-(f1) show the obtained anatomical structures by merging clusters from our approach (totally 27 clusters). (b2)-(f2) show the obtained anatomical structures by merging clusters from spectral clustering (totally 200 clusters). Colors are used to distinguish clusters. (g) plots the frequency of the numbers of clusters learnt by our approach when running 50 trials of Gibbs sampling with random initializations.

completeness, we randomly sample 5,000 pairs of fibers which are in the same anatomical structures and calculate the accuracy ($r_{correct}$) of they are also in the same clusters. $r_{average} = (r_{correct} + r_{complete})/2$ is also computed. The accuracies of our approach and spectral clustering are plotted in Figure 4 (e). As we increase the number of clusters from 2 to 25, the correctness of spectral clustering increases and its completeness decreases. Its best $r_{average}$ is found when the number of clusters is five, which is close to the ground truth, and it is lower than $r_{average}$ of our approach. The correctness of our approach is almost consistently better than spectral clustering until spectral clustering chooses more than 20 clusters. The completeness of our approach is significantly better than spectral clustering when the number of clusters of spectral clustering is larger than 5.

We compare our approach with the approach proposed in [8] on a larger data set with 12,420 fibers. In [8], fibers were first grouped into a large number of clusters (200) and then experts merged these clusters to obtain anatomical structures. In this data set there are 10 anatomical structures. Our approach clusters these fibers to 27 clusters. We also manually merge them to these 10 anatomical

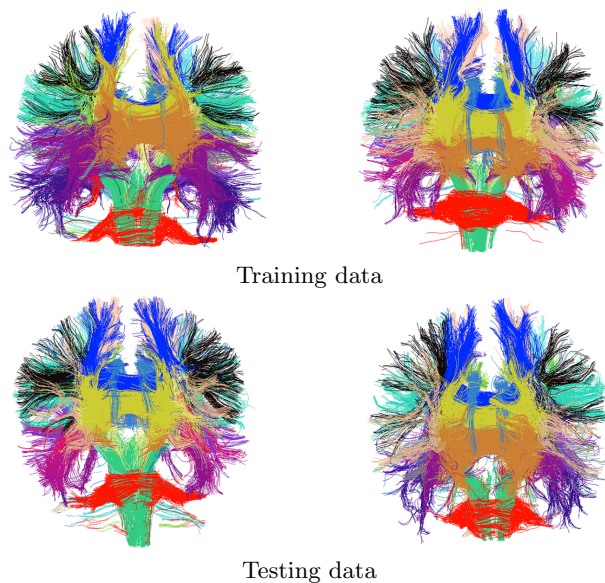


Fig. 6. Cluster fibers across multiple subjects.

structures, however it takes much less effort than [8] since the number of clusters is smaller. Figure 5 shows some of the anatomical structures obtained by the two approaches. 83.2% fibers have consistent anatomical labels according to the two results. To evaluate how our approach is sensitive to initialization, we run 50 trials of Gibbs sampling with random initializations. Figure 5 (g) plots the frequency of the numbers of clusters learnt from data.

Figure 6 shows the results of clustering fibers across multiple subjects. The training data has 63,751 fibers of two subjects. The models of bundles are learnt from all these fibers. The testing data has 61,572 fibers of two subjects.

4 Conclusion

We propose a nonparametric Bayesian framework for tractography segmentation. The number of clusters is automatically learnt from data through DP. This method has much lower space complexity than distance-based clustering methods can cluster a very large set of fibers. Our Bayesian model is very flexible to include knowledge from experts as priors. In future work, we plan to incorporate anatomical information in the model to guide tractography segmentation.

Acknowledgments

This work was supported by NIH grants NIH R01 MH074794, NIH P41 RR13218, NIH U54 EB005149, and NIH U41 RR019703. The authors also want to acknowledge valuable discussions with Lauren O’Donnell regarding techniques for clustering fiber tracts.

References

1. Ferguson, T.S.: A bayesian analysis of some nonparametric problems. *The Annals of Statistics* **1** (1973) 209–230
2. Brun, A., Knutsson, H., Park, H.J., Shenton, M.E., Westin, C.-F.: Clustering fiber traces using normalized cuts. In: *Proc. of MICCAI*. (2004) 368-375
3. Jonasson, L., Hagmann, P., Thiran, J.P., Wedeen, V.J.: Fiber tracts of high angular resolution diffusion mri are easily segmented with spectral clustering. In: *International Society for Magnetic Resonance in Medicine*. (2005) 1310
4. Gerig, G., Gouttard, S., Corouge, S.: Analysis of brain white matter via fiber tract modeling. In: *Proc. of IEEE Engineering in Medicine and Biology*. (2004) 4421-4424
5. Xia, Y., Turken, A.U., Whitfield-Gabrieli, S.L., Gabrieli, J.D.: Knowledge-based classification of neuronal fibers in entire brain. In: *Proc. of MICCAI*. (2005) 205-212
6. Maddah, M., Grimson, W.E.L., Warfield, S.K., Wells, W.M.: A unified framework for clustering and quantitative analysis of white matter fiber tracts. *Medical Image Analysis* **12** (2008) 191–202
7. Ding, Z., Gore, J.C., Anderson, A.W.: Classification and quantification of neuronal fiber pathways using diffusion tensor mri. *Magnetic Resonance in Medicine* **49** (2003) 716–721
8. O’Donnell, L.J., Westin, C.F.: Automatic tractography segmentation using a high-dimensional white matter atlas. *IEEE Trans. on Medical Imaging* **26** (2007) 1562–1575
9. Moberts, B., Vilanova, A., Jake, J.W.: Evaluation of fiber clustering methods for diffusion tensor imaging. In: *Proc. of IEEE Visualization*. (2005) 65-72
10. Maddah, M., Zollei, L., Grimson, W.E.L., Wells III, W.M.: Modeling of anatomical information in clustering of white matter fiber trajectories using dirichlet distribution. In: *MMBIA*. (2008) 1-7
11. Adelino, R., Ferreira, S.: A dirichlet process mixture model for brain mri tissue classification. *Medical Image Analysis* **11** (2006) 169–182
12. Kim, S., Smyth, P.: Hierarchical dirichlet processes with random effects. In: *Proc. of NIPS*. (2006) 697-704
13. Thirion, B., Tucholka, A., Keller, M., Pinel, P.: High level group analysis of fmri data based on dirichlet process mixture models. In: *Proc. of IPMI*. (2007) 482-494
14. Jbabdi, S., Woolrich, M.W., Behrens, T.: Multiple-subjects connectivity-based parcellation using hierarchical dirichlet process mixture models. *NeuroImage* **44** (2009) 373–384
15. Teh, Y.W., Jordan, M.I., Beal, M.J., Blei, D.M.: Hierarchical dirichlet processes. *Journal of the American Statistical Association* **101** (2006) 1566-1581
16. Wang, X., Ma, K.T., Ng, G.W., Grimson, W.E.L.: Trajectory analysis and semantic region modeling using a nonparametric Bayesian model. In: *Proc. of CVPR*. (2008) 1-8
17. Sethuraman, J.: A constructive definition of dirichlet priors. *Statistica Sinica* **4** (1994) 639–650
18. Blei, D.M., Jordan, M.I.: Variational inference for dirichlet process mixtures. *Bayesian Analysis* **1** (2006) 121–144
19. Asuncion, A., Smyth, P., Welling, M.: Asynchronous distributed learning of topic models. In: *Proc. of NIPS*. (2008)