Image segmentation using Dirichlet process mixture model

Anay Pattanaik, Anupreet Porwal, Saurabh Kataria
Indian Institute of Technology Kanpur
{anay,anupreet,saurabhk}@iitk.ac.in

Final Report
Supervisor: **Prof. Piyush Rai**
Department of Computer Science and Engineering
Indian Institute of Technology Kanpur

Abstract

Non-parametric Bayesian (NPB) methods allows us to devise conceptually simpler models that can adapt to the size of training data without committing to fixed number of parameters. Dirichlet process (DP) is one of the NBP techniques which is used primarily for clustering. DP is well suited for image segmentation purposes where it is difficult to fix the number of clusters/ segments a priori. It can also be coupled with Markov random field (MRF) to incorporate spatial coherence in the image. In this project, we wish to study and implement MDP-MRF model for image segmentation, and compare it with other popular segmentation techniques. We also evaluate the performance of algorithm on a subset of IBSR dataset using an error metric defined later in the report. This will be followed by a discussion about distance dependent Chinese restaurant process (CRP) as an alternative to MRF constrained MDP model.

1 Literature Survey

1.1 Introduction

Any core image segmentation approach focus on two core issues: the information captured by a single segment and the choice of number of segments. Dirichlet process mixture models(MDP) deals the second problem i.e. the choice of number of segments. MDP provides a Bayesian framework for clustering problems with unknown number of clusters. It differ from the conventional parametric methods due to the fact it treats number of cluster as a random variable with a control parameter. A Dirichlet process(DP) is a distribution over probability distributions. Formally, a DP is parametrized by a concentration parameter α and base distribution H and a distribution G is distributed over Θ according to the DP(α,H) iff

$$(G(T_1), G(T_2), \ldots, G(T_K)) \sim Dirichlet(\alpha H(T_1), \alpha H(T_2), \ldots, \alpha H(T_K)) \tag{1}$$

where $(T_1, T_2, \ldots, T_K)$ is any finite partition of Θ. In MDP models, DP distributed discrete random measure is used as a prior over the components of a mixture models.

1.2 Unconstrained MDP model

In image segmentation the goal is to cluster local histograms that are extracted as features from the image(process of feature extraction is explained briefly in the next section). Each histogram is described by a vector $\mathbf{h_i} = (h_{i1}, \ldots, h_{iN_{bins}})$ of non negative integers. DP can be mathematically formulated as:

$$\begin{aligned} G &\sim DP(\alpha, H) \\ \theta_i | G &\sim G \\ \mathbf{h_i} | \theta_\mathbf{i} &\sim F(\theta_i) \end{aligned} \tag{2}$$

where $\mathbf{h_i}$ are the local histograms that we want to cluster, θ_i are the probability of the histogram bins. An element of vector θ_i,θ_{ij} denotes the probability for a value to occur in j bin for histogram at site i. Given θ_i the histograms are considered to be multinomially distributed. The prior distribution over the parameter vectors is assumed to be a Dirichlet distribution of dimension N_{bins}. However, MDP models neglect the spatial smoothness of the natural images which may lead to incoherent segments in noisy images like that of MRI. This problem can be addressed by constrained model using Markov random field (MRF) described in next sub-section.

1.3 MDP-MRF model

The notion of spatial dependence between neighborhood image features forms the basis of MRF. A joint distribution Π on the parameters $\theta_1, \ldots, \theta_n$ is called a MRF w.r.t. a neighborhood graph (For image segmentation, vertices representing a location in an image and an edge indicating locations are adjacent) if

$$\Pi(\theta_i|\theta_{-i}) = \Pi(\theta_i|\theta_{\delta(i)}) \tag{3}$$

where θ_{-i} denotes the parameter set with θ_i removed and $\theta_{\delta(i)}$ the index set of neighbours of a particular site of an image. Thus, intuitively speaking, the Markov property tells us that the sites are dependent but the dependence is local, restricted to the neighborhood of the site. Due to such dependency among location sites of the image captured by MRF, the MDP/MRF model now forms a new cluster, say K+1, with probability proportional to α and remains in one of the old cluster $\{$1,2,...,K$\}$ with probability proportional to n_i, where i=1,..., K , i.e. number of points already in that cluster and number of sites in its neighborhood having the same cluster assignment. Thus, the generative model for a MRF constrained MDP can be written as follows

Generative Story of the model

$$\begin{aligned} G &\sim DP(\alpha, H) \\ \theta_i | G &\sim M(\theta_i|\theta_{-i})G(\theta_i) \\ \mathbf{h_i} &\sim Mult(h_i|\theta_i) \end{aligned} \tag{4}$$

is the MRF contribution to the prior and is of form $M(\theta_i|\theta_{-i}) \propto exp(-H(\theta_i|\theta_{-i}))$ where $H(\theta_i|\theta_{-i})$ is the cost function defined on the neighborhood graph defined as follows

$$H(\theta_i|\theta_{-i}) = \Sigma_{l \in \delta(i)} \delta_{\theta_i, \theta_l}$$

This ensures our motive of spatial smoothness while image segmentation and fulfils the characteristics of MDP models by learning the number of clusters depending on the data provided.

1.4 Distance dependent Chinese restaurant process

An alternative to MRF is to take distance into account explicitly. This is achieved through distance dependent Chinese restaurant process (ddCRP). Thus, ddCRP can be used for spatial setting. As opposed to CRP wherein each customer is assigned to a table/cluster, in ddCRP each customer is assigned to another customer. The distribution of customer assignment can be given by.

$$p(c_i = j|D, f, \alpha) \propto \begin{cases} f(d_{ij}) & j \neq i \\ \alpha & j = i \end{cases} \tag{5}$$

Here d_{ij} represents the distance between the two customers and $f(d_{ij})$ is the decay function. This function determines how the distance between two customers affects the probability of one being connected to another and being assigned the same table/cluster. In the context of image segmentation, the pixels/windows can be thought of as customers. The distance between windows/pixels is the number of jumps required to reach one window/pixel from another window/pixel. The generative model can be described as.

$$\begin{aligned} G &\sim DP(\alpha, H) \\ \theta_i | G &\sim G \\ c_i &\sim ddCRP(\alpha, f, d) \\ \mathbf{h_i} | c_i &\sim F(\theta_{c_i}) \end{aligned} \tag{6}$$

A distribution over parameters is generated by DP followed by sampling all the parameters for each cluster from this distribution. Then the current window/pixel assignment is done. The histogram of a data point is generated corresponding to the cluster of the assigned window/pixel. Thus, the window/pixel assignment has been done taking into account the spatial distance.

2 Implementation details

Experiments on image segmentation of a MR image is done using the code available in [7]. Brief description of code is as follows. First step is feature extraction. Code takes as input the grayscale image $(n \times m)$, size of window (r) and meshsize (m_0). Edgemap of the image can also be given as an optional input. Using sliding window method, subregions of the input image are considered and their histograms of gray values are stored. Centre of such windows are called sites. For processing pixels on boundaries, original image is padded with neighbouring pixels accordingly. A list of neighbours of each site is also stored. The final feature vector of each site consists of its histogram of gray values, neighbourhood list, $\frac{n}{m_0}$, and $\frac{m}{m_0}$.

Second step is to use the extracted features and concentration parameter α of DP to do the clustering. Inference of the parameters can be done easily using Gibbs sampling. Total number of iterations are fixed beforehand. π is chosen to be uniformly distributed. Initially, all the sites are assigned the same cluster i.e. cluster $\#1$. Number of data points in each cluster is also maintained. Probabilities of histogram bins i.e. θ given β, π, and feature matrix, are drawn from Dirichlet distribution. For the current iteration, all data points are considered one at a time. The data point under consideration is taken out from the cluster it was previously assigned. $\log$ of Markov random field cost(MRFC) is calculated which contains the lambda times the frequency of its actual neighbours assigned to each cluster. Then, the probabilities of that histogram going to existing clusters as well as probability of histogram going to a new cluster is calculated. Through a simple sampling routine, we determine where the data point will go given the probabilities we just computed. After we have updated the cluster assignment of each data point, we delete the empty clusters, update the θ parameters, and re-compute cross entropies.

After completion of all iterations, final cluster assignment is done by looking at the mode of the assignment in last 50 iterations. Below we present the segmentation results using MDP model. We have set number of bins, window radius, and mesh size equal to 8, 2, and 2 respectively.

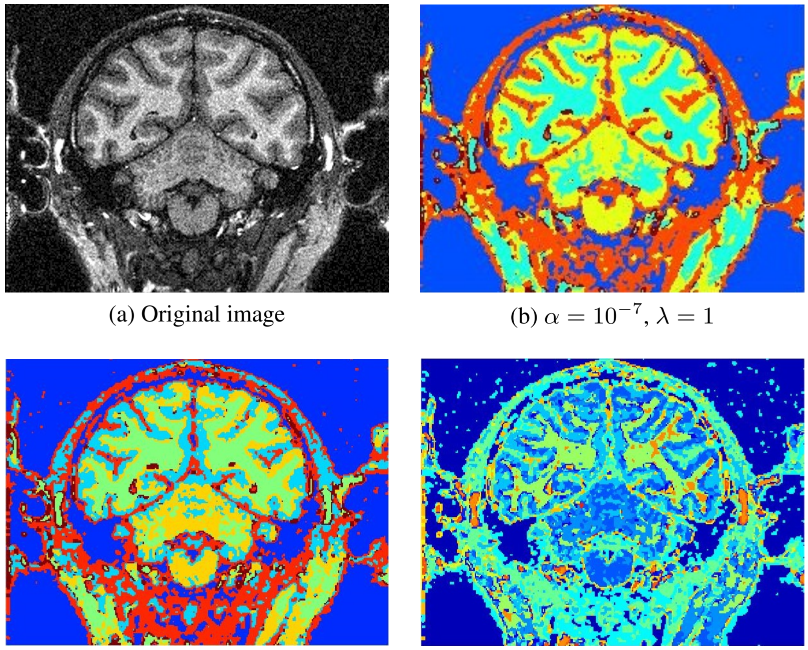
In Figure 1, the effect of α parameter of Dirichlet process is illustrated. As α increases, the number of clusters also increase leading to over-segmented and granular type image. MR images are difficult to segment because the noisy and spatially incoherent structure triggers the number of segments to increase even further. Therefore, the choice of α is crucial.

To ensure spatial compatibility and deal with the noise in MR images, the MRF smoothing parameter λ must be tuned appropriately. This parameter plays its role in the MRF cost function as follows.

$$H(\theta_i|\theta_{-i}) = -\lambda \sum_{l \in \delta(i)} \delta_{\theta_i, \theta_l} \tag{7}$$

In Figure 2, we illustrate the effect of λ keeping $\alpha = 10^{-7}$ fixed. As can be observed, decreasing λ results in poorer spatial coherency in resultant image.

In Figure 3, comparison of three popular image segmentation techniques are presented namely, MRF constrained MDP, mean-shift, and k-means (left to right). Visually, the MDP/MRF model gives best result.



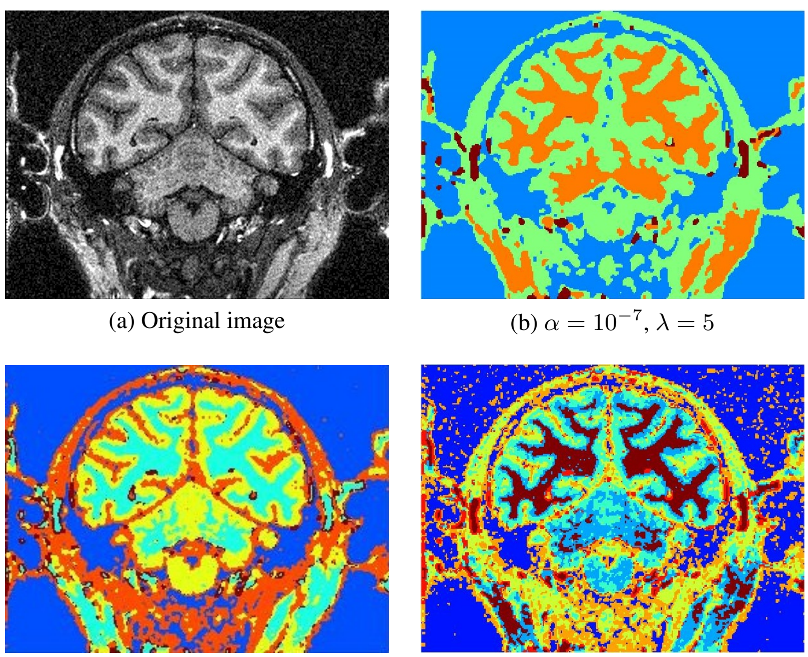
(a) Original image

(b) $\alpha = 10^{-7}$, $\lambda = 1$

(c) $\alpha = 10^{-4}$, $\lambda = 1$

(d) $\alpha = 1$, $\lambda = 1$

Figure 1: The following images shows the original image (top-left) and segmented images with α and λ mentioned below the images. Increasing the value of α leads to more clusters and therefore over-segmentation.



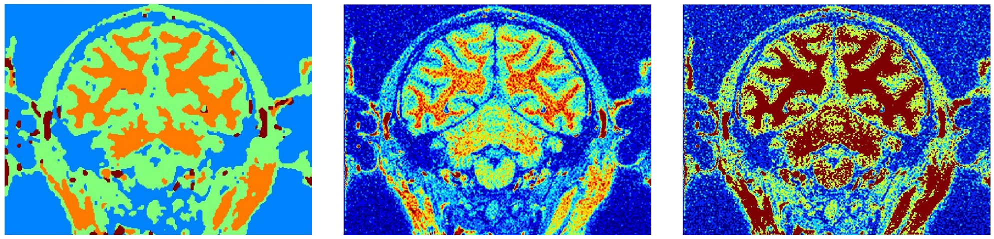
(a) Original image

(b) $\alpha = 10^{-7}$, $\lambda = 5$

(c) $\alpha = 10^{-7}$, $\lambda = 1$

(d) $\alpha = 10^{-7}$, $\lambda = 0$

Figure 2: The following images shows the original image (topleft) and segmented images with α and λ mentioned below the images. Increasing the value of λ leads to strong smoothing and ensure local smoothness while doing image segmentation.



(a) MDP/MRF; $\alpha = 10^{-7}$, $\lambda = 5$

(b) Mean shift; Kernel bandwidth $bw = 0.2$

(c) K-means; Number of clusters $K = 5$

Figure 3

3 Performance evaluation and results

Here we evaluate the performance of MDP-MRF model on image segmentation task. Performance of the algorithm was evaluated with real MR images consisting of 15 different individual slices of a single patient's data provided by the Internet Brain Segmentation Repository (IBSR) and was tested with its ground truth classification performed by trained investigators using a semi- automated intensity contour mapping algorithm. The 20 normal MR brain data sets and their manual segmentations were provided by the Center for Morphometric Analysis at Massachusetts General Hospital and are available at http://www.cma.mgh.harvard.edu/ibsr/. To compare the ground truth classification image with the algorithm generated segmented image, we use the following technique. Let us say that the two cluster configuration of n points is denoted by C and C'. Let n_{01} denotes the number of pair of points that are clustered together in C but not in C', and n_{10} denotes the number of pair of points that are clustered together in C' but not in C. Then error is evaluated as

$$e = \frac{(n_{01} + n_{10})}{\binom{n}{2}} \tag{8}$$

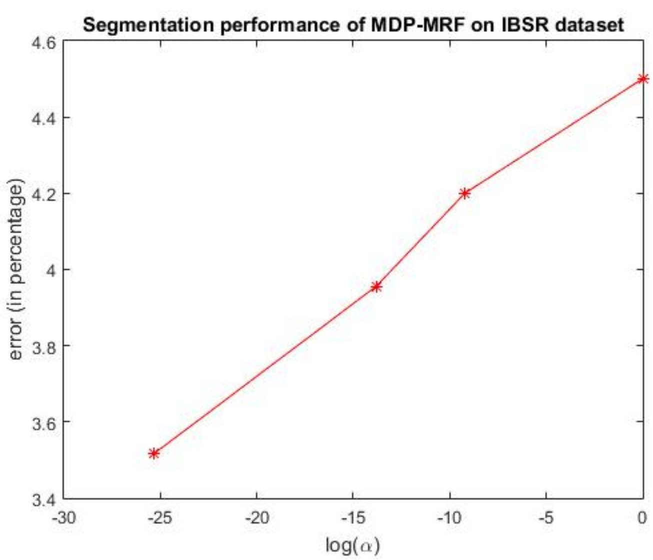
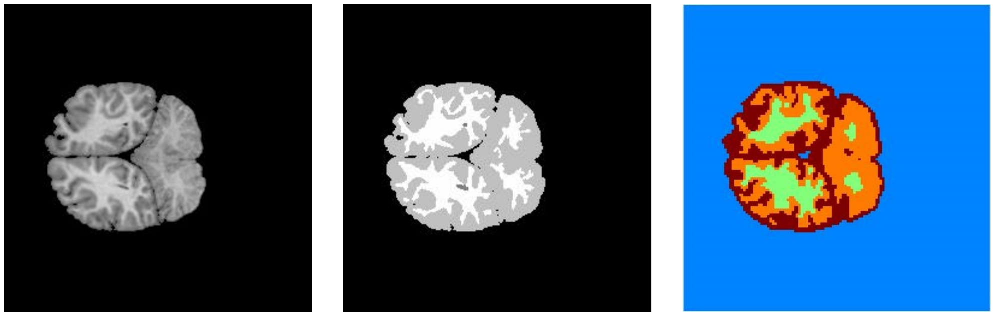


Figure 3 shows the variation of error with increasing α. Since we do not expect the number of clusters in a brain MR image to be too large, the unnecessary clusters created results in larger error. For instance, $\alpha = 10^{-11}$ gives an error of about 3.5%.

Figure 4 illustrates the kind of result obtained from the algorithm run on IBSR dataset.



(a) Sample original MR image from IBSR dataset

(b) Ground truth of segmentation

(c) Segmentaton result of algorithm

Figure 4

4 Conclusion

In this project, we have looked at the unconstrained MDP model, MRF constrained MDP model and how different hyperparameters of these models affect their performance in the image segmentation task. We have experimentally verified the fact that with increasing the value of concentration parameter α, there is a increase in number of clusters formed, and increasing the value of MRF cost parameter λ increases the spatial dependence of a site on its neighbors and thus gives more spatially coherent segmentation. We also understand that choice of these parameters becomes crucial and can be learnt through data using a Bayesian prior on the same. We also looked at distance dependent CRP which models the customer assignment using a decay function of distance between the nodes and thus works well even when the data is not exchangeable which is the case in image segmentation tasks.

A substantial amount of progress is being done in the area of non- parametric Bayesian models. This project helped us to gain sufficient exposure to non parametric Bayesian techniques and leaves us with many interesting directions to explore in future like regional ddCRP which takes color and texture properties of image into account while segmenting the images or DP means which talks about asymptotic relation between K mean and MDP models.

5 Acknowledgment

We would like to express our gratitude to Prof. Piyush Rai for his mentorship, guidance, and encouragement throughout the course and project. We truly appreciate his humility and constant support that enabled us to successfully achieve the proposed objectives.

References

[1] Gershman, Samuel J., and David M. Blei. ”A tutorial on Bayesian nonparametric models.” Journal of Mathematical Psychology 56.1 (2012): 1-12.

[2] da Silva, Adelino R. Ferreira. ”A Dirichlet process mixture model for brain MRI tissue classification.” Medical Image Analysis 11.2 (2007): 169-182.

[3] Orbanz, Peter, and Joachim M. Buhmann. ”Nonparametric Bayesian image segmentation.” International Journal of Computer Vision 77.1-3 (2008): 25-45.

[4] Ghosh, Soumya, et al. ”Spatial distance dependent Chinese restaurant processes for image segmentation.” Advances in Neural Information Processing Systems. 2011.

[5] Blei, David M., and Peter I. Frazier. ”Distance dependent Chinese restaurant processes.” The Journal of Machine Learning Research 12 (2011): 2461-2488.

[6] Kulis, Brian, and Michael I. Jordan. ”Revisiting k-means: New algorithms via Bayesian nonparametrics.” arXiv preprint arXiv:1111.0352 (2011).

[7] http://stat.columbia.edu/ porbanz/IJCV.html