Bayesian Nonparametric Priors for Hidden Markov Random Fields: Application to Image Segmentation

Hongliang Lu, Julyan Arbel, Florence Forbes

Bayesian Nonparametric Priors for Hidden Markov Random Fields: Application to Image Segmentation

Hongliang Lü, Julyan Arbel, Florence Forbes
Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France

I. Background and motivation

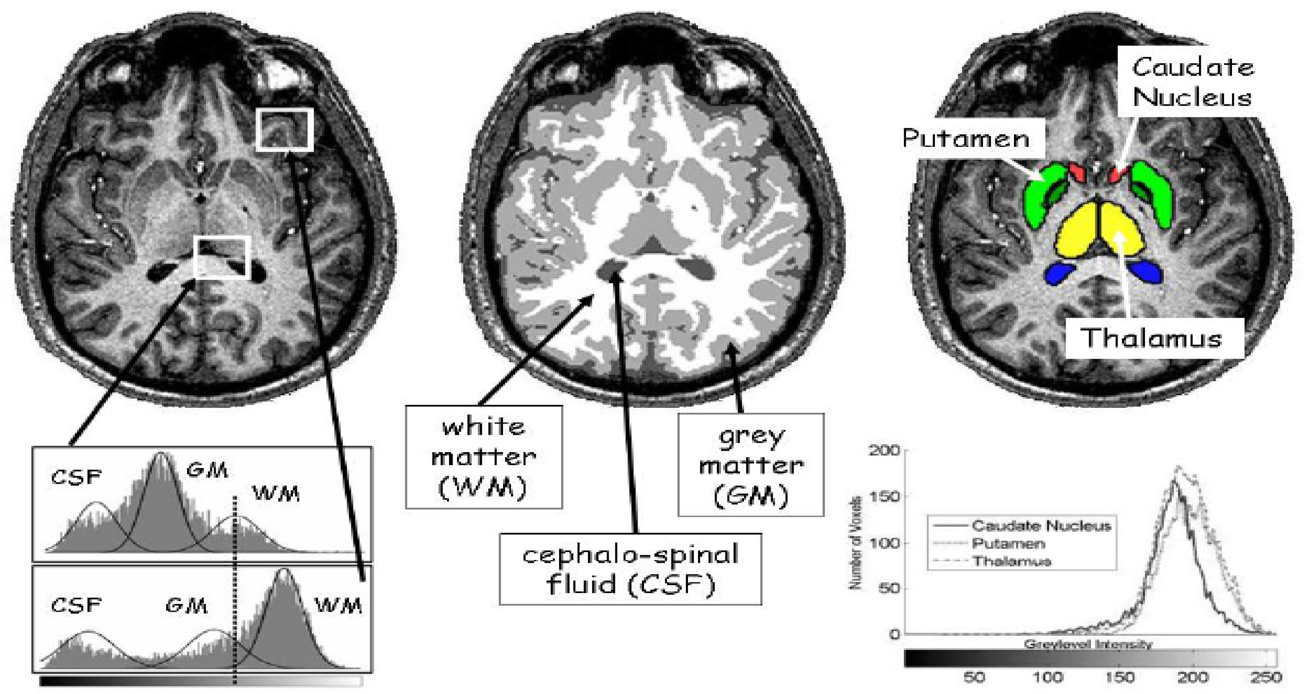


Figure 1: Illustration of challenges for unsupervised image segmentation: blur, noise, color/contrast imperfection, partial volume effect (large slice thickness), anatomic variability and complexity, number of segments...

Image segmentation in real-world applications is typically performed on noisy images. To achieve better segmentation performance, several extensions of the usual **Bayesian nonparametric (BNP) mixture model** with spatial regularization are therefore necessary.

II. BNP Priors

The **Dirichlet process (DP)** is one of the most commonly used BNP priors. It is a **random process** G defined over a probability space $\mathcal{Y}$ and characterized by a **concentration parameter** α and a **base distribution** G_0, such that for any finite partition $\{A_1, \dots, A_p\}$ of $\mathcal{Y}$, the random vector $(P(A_1), \dots, P(A_p))$ is **Dirichlet distributed**:

$$(P(A_1), \dots, P(A_p)) \sim \mathrm{Dir}(\alpha G_0(A_1), \dots, \alpha G_0(A_p))$$

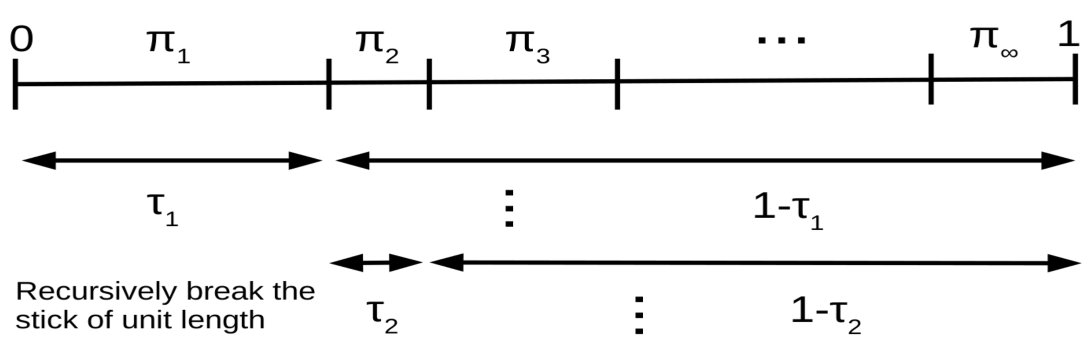
which is often denoted by $G \sim \mathrm{DP}(\alpha, G_0)$.

III. Stick-breaking construction

The **DP** has almost surely **discrete** realizations. It can be built by the **stick-breaking construction**:

$$G = \sum_{k=1}^{\infty} \pi_k(\boldsymbol{\tau})\, \delta_{\theta_k^*} = \sum_{k=1}^{\infty} \left[\tau_k \prod_{l<k} (1-\tau_l) \right] \delta_{\theta_k^*}$$

where $\theta_k^* \overset{\text{iid}}{\sim} G_0$ and $\tau_k \overset{\text{iid}}{\sim} \mathcal{B}(1, \alpha)$.



References

[1] P. Orbanz and J. M. Buhmann, IJCV, vol. 77, no 1-3, p. 25-45, 2008.

[2] M. Albughdadi *et al.*, Signal Processing, 135:132-146, 2017.

[3] F. Forbes and N. Peyrard, IEEE TPAMI, 25(9):1089-1101, 2003.

[4] P. Arbelaez *et al.*, IEEE TPAMI, Vol. 33, No. 5, pp. 898-916, May 2011.

[5] S. P. Chatzis, Pattern Recognition, 46(6): 1595-1603, 2013.

IV. DP-Potts mixture model

The usual **DP mixture model** assumes that a set of data points $\mathbf{y} = \{y_1, \dots, y_N\}$ with $y_i \in \mathbb{R}^D$ (*e.g.*, pixels) can be generated through the following hierarchical representation:

- $G \sim \mathrm{DP}(\alpha, G_0)$
- $\theta_i | G \sim G,\ i = 1, \dots, N$
- $y_i | \theta_i \sim F(y_i|\theta_i),\ i = 1, \dots, N$

where $\boldsymbol{\theta} = \{\theta_1, ..., \theta_N\}$ denotes a set of model parameters. To take into account spatial constraints, we introduce a **Potts model** component using a set of assignment variables $\mathbf{z} = \{z_1, \dots, z_N\}$ with $z_i = z(\theta_i)$ so as to favor spatial agregation [1]:

$$M(\boldsymbol{\theta}) \propto \exp\left(\beta \sum_{i \sim j} \delta_{z(\theta_i) = z(\theta_j)} \right)$$

with β being the **regularization parameter**. The **DP mixture model** is thus extended to become the **DP-Potts mixture model**:

- $G \sim \mathrm{DP}(\alpha, G_0)$
- $\boldsymbol{\theta} | M, G \sim M(\boldsymbol{\theta}) \times \prod_i G(\theta_i)$
- $y_i | \theta_i \sim F(y_i|\theta_i),\ i = 1, \dots, N$

Accordingly, the **stick-breaking construction** of the **DP-Potts mixture model** can be summarized as follows:

- $\theta_k^* | G_0 \sim G_0$ and $\tau_k | \alpha \sim \mathcal{B}(1, \alpha),\ k = 1, 2, \dots$
- $\pi_k(\boldsymbol{\tau}) = \tau_k \prod_{l=1}^{k-1} (1 - \tau_l),\ k = 1, 2, \dots$
- $p(\mathbf{z}|\boldsymbol{\tau}; \beta) \propto \prod_i \pi_{z_i}(\boldsymbol{\tau}) \exp(\beta \sum_{i \sim j} \delta_{z_i = z_j}),\ z_i = 1, 2, \dots$
- $y_i | z_i, \boldsymbol{\theta}^* \sim F(y_i | \theta^*_{z_i}),\ i = 1, \dots, N$

V. Variational Bayes

In a Bayesian setting, we need to evaluate the intractable posterior distribution $p(\mathbf{z}, \boldsymbol{\tau}, \alpha, \boldsymbol{\theta}^* | \mathbf{y}; \phi)$ (ϕ denotes a set of hyperparameters) which can be estimated by means of the **mean-field approximation**:

$$q(\mathbf{z}, \boldsymbol{\tau}, \alpha, \boldsymbol{\theta}^*) \simeq q_{\mathbf{z}}(\mathbf{z}) q_{\boldsymbol{\tau}}(\boldsymbol{\tau}) q_\alpha(\alpha) q_{\boldsymbol{\theta}^*}(\boldsymbol{\theta}^*)$$

Variational Bayes (VB) consists of alternating maximization of **free energy**

$$\mathcal{F}(q_{\mathbf{z}}, q_{\boldsymbol{\tau}}, q_\alpha, q_{\boldsymbol{\theta}^*}; \phi) = \mathbb{E}_{q_{\mathbf{z}} q_{\boldsymbol{\tau}} q_\alpha q_{\boldsymbol{\theta}^*}} \left[\log \frac{p(\mathbf{z}, \boldsymbol{\tau}, \alpha, \boldsymbol{\theta}^*, \mathbf{y}; \phi)}{q_{\mathbf{z}} q_{\boldsymbol{\tau}} q_\alpha q_{\boldsymbol{\theta}^*}} \right]$$

which implies [2]

- **E-steps**: VE-$\mathbf{z}$, VE-α, VE-$\boldsymbol{\tau}$ and VE-$\boldsymbol{\theta}^*$.
- **M-steps**: ϕ updating is straightforward except for β.

Here, **M-β step** leads to the estimation of β:

$$\hat{\beta} = \arg\max_{\beta} \mathbb{E}_{q_{\mathbf{z}} q_{\boldsymbol{\tau}}} [\log p(\mathbf{z}|\boldsymbol{\tau}; \beta)]$$

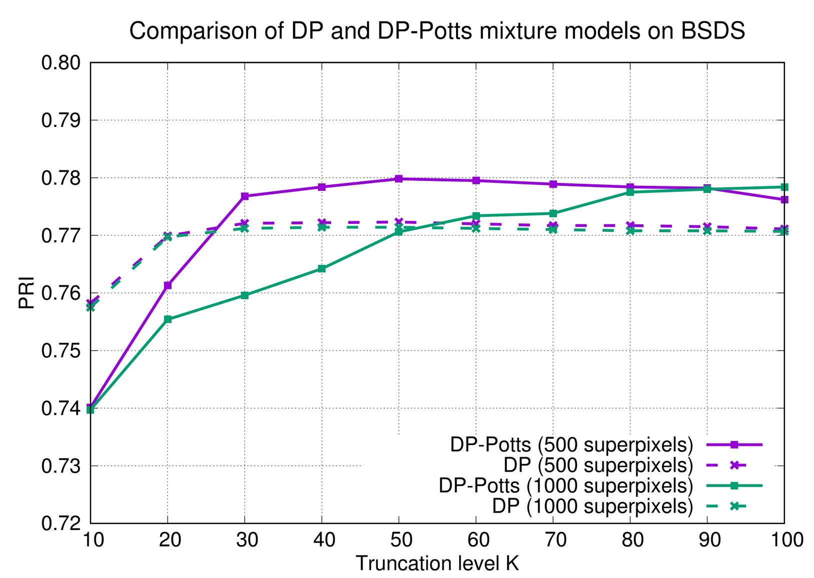
which involves $p(\mathbf{z}|\boldsymbol{\tau}; \beta) = \mathcal{K}(\beta, \boldsymbol{\tau})^{-1} \exp\left(V(\mathbf{z}; \boldsymbol{\tau}, \beta)\right)$ with the **normalization constant** $\mathcal{K}(\beta, \boldsymbol{\tau})$ and the **potential function**

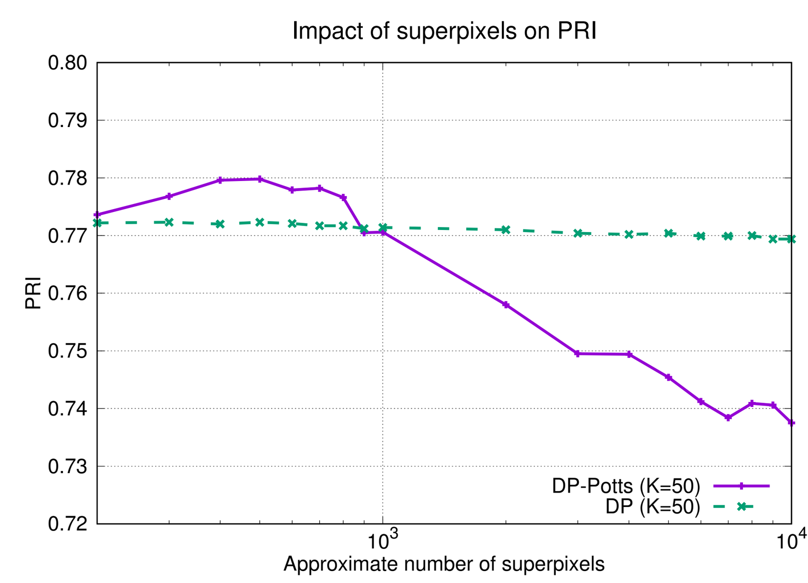
$$V(\mathbf{z}; \boldsymbol{\tau}, \beta) = \sum_i \log \pi_{z_i}(\boldsymbol{\tau}) + \beta \sum_{i \sim j} \delta_{z_i = z_j}$$

To find the optimal value of β, further approximations, such as the mean-field-like approximation [3] of $q_{\mathbf{z}}$ and replacing $\boldsymbol{\tau}$ with a fixed $\tilde{\boldsymbol{\tau}} = \mathbb{E}_{q_{\boldsymbol{\tau}}}[\boldsymbol{\tau}]$, are required.

VI Some experiments and results

Experiments were performed using superpixels on a subset (154 images) of the Berkeley segmentation data set (BSDS) [4]. Regarding the performance evaluation, the probabilistic rand index (PRI) was computed under different conditions:





We also compared our best results ($K = 50$ and about 500 superpixels) with those given in the literature:

	Proposed models		Results given in [5]		
PRI	**DP**	**DP-Potts**	iHMRF	MRF-PYP	Graph Cuts
Mean (%)	**77.23**	**77.98**	75.50	76.49	76.10
Median (%)	**79.05**	**79.56**	76.89	78.08	77.59

VII Conclusions and future work

A general **DP-Potts mixture model** and the associated **VB algorithm** were proposed. The model was successfully applied to image segmentation on different types of datasets. We also investigated the impact of β on the segmentation results and presented an **estimation procedure for β**.

In the sequel, we plan to survey **how β affects the inferred number of components**. Other types of priors (**Pitman-Yor process**, **Gibbs-type priors**, etc.) and other variational approximations (truncation-free) will also be considered. On the other hand, it is crucial to study theoretical properties of BNP priors under **structural constraints (temporal or spatial)**. Other applications may also be possible, such as **discovery probability** and **community detection in graphs**.