Bayesian Nonparametric Priors for Hidden Markov Random Fields: Application to Image Segmentation
Hongliang Lu, Julyan Arbel, Florence Forbes

To cite this version:
I. BACKGROUND AND MOTIVATION

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 \( \alpha \) and a base distribution \( G_0 \), such that for any finite partition \( \{A_1, \ldots, A_J\} \) of \( \mathcal{Y} \), the random vector \((P(A_1), \ldots, P(A_J))\) is Dirichlet distributed:

\[
P(A_1), \ldots, P(A_J) \sim \text{Dir}(\alpha G_0(A_1), \ldots, \alpha G_0(A_J)),
\]

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

III. STICK-BREAKING CONSTRUCTION

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

\[
G = \sum_{k=1}^{\infty} \tau_k \delta_{G_0} = \sum_{k=1}^{\infty} \tau_k \prod_{i=0}^{k-1} (1-\tau_i) \delta_{G_0},
\]

where \( \tau_k \sim \text{Beta}(1, \alpha) \) and \( \delta_{G_0} \) is the Dirac measure at \( G_0 \).

IV. DP-POTTS MIXTURE MODEL

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

\[
\begin{align*}
G & \sim \text{DP}(\alpha, G_0) \\
\theta_1, \theta_N & \sim G, \text{ i.i.d.} \\
y_{ij} & \sim F(\theta_{y_{ij}}), \text{ i.i.d.}
\end{align*}
\]

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

\[
M(\theta) \propto \exp \left( \beta \sum_{i=\sim} \delta_{z_i(z_j)} \right)
\]

with \( \beta \) being the regularization parameter. The DP mixture model is thus extended to become the DP-Potts mixture model:

\[
\begin{align*}
G & \sim \text{DP}(\alpha, G_0) \\
\theta & \sim M(\theta) \otimes \prod G(\theta_i) \\
y_{ij} & \sim F(\theta_{y_{ij}}), \text{ i.i.d.}
\end{align*}
\]

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

\[
\begin{align*}
\theta_k & \sim G_0 \quad k = 1, 2, \ldots \quad \text{IID} \\
s_k(\tau) \sim \text{Beta}(1, \tau) \\
p(\tau, \beta) \propto \prod_{i=1}^{K} \delta_{z_i(z_j)} \\
y_{ij} & \sim F(\theta_{y_{ij}}), \text{ i.i.d.}
\end{align*}
\]

V. VARIATIONAL BAYES

In a Bayesian setting, we need to evaluate the intractable posterior distribution \( p(\alpha, \beta, \theta^* | y, \phi) \) which can be estimated by means of the mean-field approximation:

\[
q(\alpha, \beta, \theta^*) \approx q(\alpha)(\theta^* | \phi) \propto \frac{1}{Z} \exp \left( \sum_{k=1}^{K} \log p(\tau_0, \beta, \theta^*) \right)
\]

Variational Bayes (VB) consists of alternately maximizing the free energy

\[
F(\alpha, \beta, \theta^*) = \mathbb{E}_{q(\alpha, \beta, \theta^*)} \left[ \log \frac{p(\alpha, \beta, \theta^*, y | \phi)}{q(\alpha, \beta, \theta^*)} \right]
\]

which implies [2]

\[
\begin{align*}
\text{E-steps: } & \quad \text{VE}-\alpha, \text{ VE}-\phi, \text{ VE}-\tau \quad \text{and VE}-\theta^* \\
\text{M-steps: } & \quad \phi \text{ updating is straightforward except for } \beta.
\end{align*}
\]

Here, the \( \text{M-\beta} \) step leads to the estimation of \( \beta \):

\[
\beta = \text{arg max} \mathbb{E}_{q(\alpha)} \left[ \log p(\alpha | \beta) \right]
\]

which involves \( p(\alpha | \beta) \sim K(\beta, \tau)^{-1} \exp \left( V(\alpha, \tau, \beta) \right) \) with the normalization constant \( K(\beta, \tau) \) and the potential function

\[
V(\alpha, \tau, \beta) = \sum_{i} \log \tau_i(v) + \beta \sum_{i, j} \delta_{z_i(z_j)}.
\]

To find the optimal value of \( \beta \), further approximations, such as the mean-field-like approximation [3] of \( \delta_k \) and replacing \( \tau \) with a fixed \( \bar{\tau} = \mathbb{E}_{q(\alpha)}[\tau] \), are required.

REFERENCES


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:

\[
\begin{align*}
\text{PRI} & = \text{arg max } \mathbb{E}_{q(\alpha)} \left[ \log p(\alpha | \beta) \right] \\
\text{DP} & = \text{DP-Potts mixture model} \\
\text{DP-Potts} & = \text{DP-Potts mixture model} \\
\text{MRF-PYP} & = \text{MRF-PYP mixture model} \\
\text{Graph Cuts} & = \text{Graph Cuts algorithm}
\end{align*}
\]

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

\[
\begin{align*}
\text{Proposed models} & \quad \text{Results given in [3]} \\
\text{PRI} & \quad 77.23 \quad 77.98 \quad 75.50 \quad 76.49 \quad 76.10 \\
\text{Median (\%)} & \quad 79.05 \quad 79.56 \quad 76.89 \quad 78.08 \quad 77.59
\end{align*}
\]

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 \( \beta \) on the segmentation results and presented an estimation procedure for \( \beta \).

In the sequel, we plan to survey how \( \beta \) affects the inferred number of components. Other types of priors (Pittman-Yor process, Gibbys-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.