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

To cite this version:

HAL Id: hal-01950666
https://hal.archives-ouvertes.fr/hal-01950666
Submitted on 11 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Bayesian Nonparametric Priors for Hidden Markov Random Fields: Application to Image Segmentation
Hongliang Lu, Jylana Arbel, Florence Forbes
Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France

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:

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

where \( \theta^1 \sim G_0 \) and \( \tau_k \sim \text{B}(1, \alpha) \).

IV. DP-POTTS MIXTURE MODEL

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

- \( G \sim \text{DP}(\alpha, G_0) \)
- \( \theta \sim G, i = 1, \ldots, N \)
- \( y_i | \theta_i \sim f(y_i | \theta_i), i = 1, \ldots, N \)

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 \( z_i = z(\theta_i) \), so as to favor spatial aggregation [1]:

\[
M(\theta) \propto \exp \left( \beta \sum_{i=1}^{N} \delta_{z(i)=z(i)} \right)
\]

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

- \( G \sim \text{DP}(\alpha, G_0) \)
- \( \theta | G, \beta \sim M(\theta | G, \beta) \)
- \( y_i | \theta_i \sim f(y_i | \theta_i), i = 1, \ldots, N \)

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

- \( \theta_k \sim G_0 \) and \( \tau_k \sim \text{B}(1, \alpha), k = 1, 2, \ldots \)
- \( x_{ij}(\tau) = \tau_k \prod_{l=k+1}^{\infty} (1 - \tau_l) \tau_k \)
- \( z_{ij} \sim \text{Potts}(\theta_{ij}, \beta) \)
- \( y_i | \theta_i \sim f(y_i | \theta_i), i = 1, \ldots, N \)

V. VARIATIONAL BAYES

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

\[
q(\alpha, \tau, \theta^*, y | \phi) \geq q(\alpha)(y|\tau)(\tau|\alpha)q(\theta^*)
\]

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

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

which implies [2]

- E-steps: VE-\(\alpha\), VE-\(\tau\), VE-\(\theta^*\) and VE-\(\theta^*\)
- M-steps: \(\phi\) updating is straightforward except for \(\beta\).

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

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

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

\[
V(\alpha, \tau, \beta) = \sum_i \log \tau_i (\tau) + \beta \sum_i \delta_{z(i)=z(i)}
\]

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

REFERENCES