Nonparametric Bayesian in Image Processing and Applications

October 24- Nov 23, 2011
Title: Shared Segmentation of Natural Scenes using Dependent Pitman-Yor Processes

Aim: Unsupervised Natural Scenes segmentation using Pitman-Yor Processes.

- Automatic discovery of visual categories (e.g., foliage, mountains, buildings, oceans) from a collection of (natural scene) images.

- Pitman-Yor processes are introduced as priors for the number of object categories and instances as well as the sizes of the image segments.

- Spatial coherence and dependence of image segments are modeled using a thresholded Gaussian Processes.
An empirical validation of the Power-law behavior of the Pitman-Yor Priors in natural image segmentation is given in the paper.
Model

Computationally, the paper studies a kind of hierarchical clustering problem

1. An intra-image clustering problem that gives the image segments.
2. An inter-image clustering problem that defines the categories of segments.
3. The input data are features computed from superpixels of the images.
Model: Hierarchical Pitman-Yor Processes

Features $x_{ji} = \{x_{ji}^t, x_{ji}^c\}$ computed from each superpixel:

- The texture feature, $x_{ji}^t$, band-pass filter responses quantized to 128 bins.
- The color feature, $x_{ji}^c$, (HSV) color histogram quantized to 120 bins.

Model Components:

- Object frequency $\varphi_k$ is modeled according to $\varphi \sim \text{GEM}(\gamma_a, \gamma_b)$.
- Area $\pi_{jt}$ of region $t$ in image $j$ is modeled according to $\pi_j \sim \text{GEM}(\alpha_a, \alpha_b)$.
- Each region has an object category assignment $k_{jt} \sim \varphi$.
- Each category $k$ has its own appearance model $\theta_k = (\theta_k^t, \theta_k^c)$, with $\theta_k^t, \theta_k^c$ parameterize multinomial distributions on texture and color bins.
- $\theta_k^t \sim \text{Dir}(\rho^t), \theta_k^c \sim \text{Dir}(\rho^c)$.
Model: Hierarchical Pitman-Yor Processes

The Pitman-Yor (PY) process is denoted by \( \varphi \sim \text{GEM}(\gamma_a, \gamma_b) \), id defined by two hyperparameters satisfying \( 0 \leq \gamma_a < 1, \gamma_b > -\gamma_a \):

\[
\varphi_k = w_k \prod_{l=1}^{k-1} (1 - w_l), \quad w_k \sim \text{Beta}(1 - \gamma_a, \gamma_b + k \gamma_a).
\]

The likelihood is given by

\[
p(x_{ji}^t, x_{ji}^c | t_{ji}, k_j, \theta) = \text{Mult}(x_{ji}^t | \theta_{ji}^t) \text{Mult}(x_{ji}^c | \theta_{z_{ji}}^c), \quad z_{ji} = k_j t_{ji}.
\]

\[
\log p(x | \alpha, \gamma, \rho) \geq H(q) + \mathbb{E}_q[\log p(x, k, t, v, w, \theta | \alpha, \gamma, \rho)]
\]

\[
q(k, t, v, w, \theta) = \prod_{k=1}^{K} q(w_k | \omega_k) q(\theta_k | \eta_k) \cdot \prod_{j=1}^{J} \left[ \prod_{t=1}^{T} q(v_{ji} | \nu_{ji}) q(k_{ji} | \kappa_{ji}) \prod_{i=1}^{N_j} q(t_{ji} | \tau_{ji}) \right]
\]
Model: Hierarchical Pitman-Yor Processes

Hyperparameters
The hyperparameters are $\alpha, \gamma, \rho$:

- $\alpha, \gamma$ are the parameters for the PY-prior on segment sizes and category frequencies
- $\rho$ are the (Dirichlet) parameters for the appearance model.

Latent Variables
The latent variables are $t, k, v, w, \theta$:

- $t = t_{ji}, k = k_{jt_{ji}}$ are the segment and category indicators, respectively.
- $\theta = (\theta^t_k, \theta^c_k)$ are the appearance model of each category.
- $v = \{v_{ji}\}, w = \{w_i\}$ are the stick-breaking proportions for image segments and category frequencies.

Observations
The observations are the collection of feature vectors $(x^t_{ji}, x^c_{ji})$ extracted from image $j$ and superpixel $i$. 
4.1 Coupling Assignments using Thresholded Gaussian Processes

Consider a generative model which partitions data into two clusters via assignments \( z_i \in \{0, 1\} \) sampled such that \( \mathbb{P}[z_i = 1] = v \). One representation of this sampling process first generates a Gaussian auxiliary variable \( u_i \sim \mathcal{N}(0, 1) \), and then chooses \( z_i \) according to the following rule:

\[
z_i = \begin{cases} 
1 & \text{if } u_i < \Phi^{-1}(v) \\
0 & \text{otherwise}
\end{cases}
\]

\[
\Phi(u) \triangleq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-s^2/2} \, ds
\]

Here, \( \Phi(u) \) is the standard normal cumulative distribution function (CDF). Since \( \Phi(u_i) \) is uniformly distributed on \([0, 1]\), we immediately have \( \mathbb{P}[z_i = 1] = \mathbb{P}[u_i < \Phi^{-1}(v)] = \mathbb{P}[\Phi(u_i) < v] = v \).

We adapt this idea to PY processes using the stick-breaking representation of Eq. (1). In particular, we note that if \( z_i \sim \pi \) where \( \pi_k = \nu_k \prod_{\ell=1}^{k-1} (1 - \nu_\ell) \), a simple induction argument shows that \( \nu_k = \mathbb{P}[z_i = k \mid z_i \neq k - 1, \ldots, 1] \). The stick-breaking proportion \( \nu_k \) is thus the conditional probability of choosing cluster \( k \), given that clusters with indexes \( \ell < k \) have been rejected. Combining this insight with Eq. (4), we can generate samples \( z_i \sim \pi \) as follows:

\[
z_i = \min \left\{ k \mid u_{ki} < \Phi^{-1}(\nu_k) \right\} \quad \text{where } u_{ki} \sim \mathcal{N}(0, 1) \text{ and } u_{ki} \perp u_{ki}, k \neq \ell
\]

As illustrated in Fig. 3, each cluster \( k \) is now associated with a zero mean Gaussian process (GP) \( u_k \), and assignments are determined by the sequence of thresholds in Eq. (5). If the GPs have identity covariance functions, we recover the basic HPY model of Sec. 3.1. More general covariances can be used to encode the prior probability that each feature pair occupies the same segment. Intuitively, the ordering of segments underlying this dependent PY model is analogous to layered appearance models [23], in which foreground layers occlude those that are farther from the camera.

To retain the power law prior on segment sizes justified in Sec. 2.3, we transform priors on stick proportions \( \nu_k \sim \text{Beta}(1 - \alpha, \alpha_b + k\alpha_a) \) into corresponding random thresholds:

\[
p(\bar{\nu}_k \mid \alpha) = \mathcal{N}(\bar{\nu}_k \mid 0, 1) \cdot \text{Beta}(\Phi(\bar{\nu}_k) \mid 1 - \alpha, \alpha_b + k\alpha_a) \quad \bar{\nu}_k \triangleq \Phi^{-1}(\nu_k)
\]

Fig. 2 illustrates the threshold distributions corresponding to several different PY stick-breaking priors. As the number of features \( N \) becomes large relative to the GP covariance length-scale, the proportion assigned to segment \( k \) approaches \( \pi_k \), where \( \pi \sim \text{GEM}(\alpha_a, \alpha_b) \) as desired.
Figure 3: A nonparametric Bayesian approach to image segmentation in which thresholded Gaussian processes generate spatially dependent Pitman–Yor processes. Left: Directed graphical model in which expected segment areas $\pi \sim \text{GEM}(\alpha)$ are constructed from stick-breaking proportions $v_k \sim \text{Beta}(1 - \alpha_a, \alpha_b + k\alpha_a)$; zero mean Gaussian processes ($u_{ki} \sim \mathcal{N}(0, 1)$) are cut by thresholds $\Phi^{-1}(v_k)$ to produce segment assignments $z_i$, and thereby features $x_i$. Right: Three randomly sampled image partitions (columns), where assignments (bottom, color-coded) are determined by the first of the ordered Gaussian processes $u_k$ to cross $\Phi^{-1}(v_k)$. 