Spatial Bayesian Nonparametrics for Natural Image Segmentation

Erik Sudderth
Brown University

Joint work with
Michael Jordan
University of California

Soumya Ghosh
Brown University
Parsing Visual Scenes

- sky
- skyscraper
- sky
- dome
- temple
- bell
- buildings
- trees
Region Classification with Markov Field Aspect Models

Verbeek & Triggs, CVPR 2007
Human Image Segmentation
Berkeley Segmentation Database & Boundary Detection Benchmark
BNP Image Segmentation

Segmentation as Partitioning

• How many regions does this image contain?
• What are the sizes of these regions?

Why Bayesian Nonparametrics?

• Huge variability in segmentations across images
• Want multiple interpretations, ranked by probability
The Infinite Hype

- Infinite Gaussian Mixture Models
- Infinite Hidden Markov Models
- Infinite Mixtures of Gaussian Process Experts
- Infinite Latent Feature Models
- Infinite Independent Components Analysis
- Infinite Hidden Markov Trees
- Infinite Markov Models
- Infinite Switching Linear Dynamical Systems
- Infinite Factorial Hidden Markov Models
- Infinite Probabilistic Context Free Grammars
- Infinite Hierarchical Hidden Markov Models
- Infinite Partially Observable Markov Decision Processes
- …
Some Hope: BNP Segmentation

Model
- Dependent *Pitman-Yor processes*
- Spatial coupling via *Gaussian processes*

Inference
- Stochastic search & *expectation propagation*

Learning
- Conditional covariance calibration

Results
- Multiple segmentations of natural images
Pitman-Yor Processes

The **Pitman-Yor process** defines a distribution on infinite discrete measures, or **partitions**

\[ \pi_1 = \nu_1 \]

\[ \pi_2 = \nu_2(1 - \nu_1) \]

\[ \pi_3 = \nu_3(1 - \nu_2)(1 - \nu_1) \]

\[ \pi_k = \nu_k \left( 1 - \sum_{\ell=1}^{k-1} \pi_\ell \right) = \nu_k \prod_{\ell=1}^{k-1} (1 - \nu_\ell) \]

\[ \nu_k \sim \text{Beta}(1 - a, b + ka) \]

*Dirichlet process:*

\[ a = 0 \]
Pitman-Yor Stick-Breaking

\[ v_k \sim \text{Beta}(1 - a, b + ka) \]

\[ E[v_k] = \frac{1 - a}{1 - a + b + ka} \]

- \( a = 0.1, b = 3 \)
- \( k = 1 \) blue
- \( k = 10 \) red
- \( k = 20 \) green

- \( a = 0.5, b = 7 \)

- \( k = 1 \) blue
- \( k = 10 \) red
- \( k = 20 \) green
Human Image Segmentations

Labels for more than **29,000 segments in 2,688 images of natural scenes**
Statistics of Human Segments

How many objects are in this image?

Object sizes follow a power law

Labels for more than 29,000 segments in 2,688 images of natural scenes
### Why Pitman-Yor?

**Generalizing the Dirichlet Process**

- Distribution on partitions leads to a generalized *Chinese restaurant process*
- Special cases of interest in probability: Markov chains, Brownian motion, ...

### Power Law Distributions

| Number of unique clusters in $N$ observations | DP $\mathcal{O}(b \log N)$ | PY $\mathcal{O}(bN^a)$ |
| Size of sorted cluster weight $k$ | $\mathcal{O}\left(\alpha_b \left(\frac{1+b}{b}\right)^{-k}\right)$ | $\mathcal{O}\left(\alpha_{ab} k^{-\frac{1}{a}}\right)$ |

*Heaps’ Law:*

*Zipf’s Law:*

| Natural Language Statistics | Goldwater, Griffiths, & Johnson, 2005 |
| Teh, 2006 |
Feature Extraction

- Partition image into ~1,000 superpixels
- Compute texture and color features:
  - Texton Histograms (VQ 13-channel filter bank)
  - Hue-Saturation-Value (HSV) Color Histograms
- Around 100 bins for each histogram
Pitman-Yor Mixture Model

PY segment size prior

$$\pi_k = v_k \prod_{\ell=1}^{k-1} (1 - v_\ell)$$

$$v_k \sim \text{Beta}(1 - a, b + ka)$$

Assign features to segments

$$z_i \sim \text{Mult}(\pi)$$

Observed features (color & texture)

$$x_i^c \sim \text{Mult}(\theta^c_{z_i})$$
$$x_i^s \sim \text{Mult}(\theta^s_{z_i})$$

Visual segment appearance model

Color:

$$\theta^c_k \sim \text{Dir}(\rho^c)$$

Texture:

$$\theta^s_k \sim \text{Dir}(\rho^s)$$
Dependent DP&PY Mixtures

Some dependent prior with DP/PY “like” marginals

Assign features to segments

Observed features (color & texture)

Kernel/logistic/probit stick-breaking process, order-based DDP, ...

Visual segment appearance model

Color: \( \theta^c_k \sim \text{Dir}(\rho^c) \)

Texture: \( \theta^s_k \sim \text{Dir}(\rho^s) \)
Example: Logistic of Gaussians

• Pass set of Gaussian processes through softmax to get *probabilities of independent* segment assignments

Fernandez & Green, 2002
Figueiredo et. al., 2005, 2007
Woolrich & Behrens, 2006
Blei & Lafferty, 2006

• Nonparametric analogs have similar properties
Discrete Markov Random Fields

Ising and Potts Models

\[ p(z) = \frac{1}{Z(\beta)} \prod_{(s,t) \in E} \psi_{st}(z_s, z_t) \]

\[ \log \psi_{st}(z_s, z_t) = \begin{cases} 
\beta_{st} > 0 & z_s = z_t \\
0 & \text{otherwise}
\end{cases} \]

Previous Applications

- Interactive foreground segmentation
- Supervised training for known categories

…but learning is challenging, and little success at unsupervised segmentation.

GrabCut: Rother, Kolmogorov, & Blake 2004

Verbeek & Triggs, 2007
Phase Transitions in Action

Potts samples, 10 states sorted by size: largest in blue, smallest in red
Product of Potts and DP?

Orbanz & Buhmann 2006

\[ p(z) = \frac{1}{Z(\beta, \pi)} \prod_{(s,t) \in E} \psi_{st}(z_s, z_t) \prod_{s \in V} \pi(z_s) \]

- **Potts Potentials**
- **DP Bias:** \( \pi \sim \text{Stick}(\alpha) \)
Spatially Dependent Pitman-Yor

- Cut random surfaces (samples from a GP) with thresholds (as in Level Set Methods)
- Assign each pixel to the first surface which exceeds threshold (as in Layered Models)

Duan, Guindani, & Gelfand, *Generalized Spatial DP*, 2007
Spatially Dependent Pitman-Yor

- Cut random *surfaces* (samples from a GP) with *thresholds* (as in Level Set Methods)

- Assign each pixel to the *first* surface which exceeds threshold (as in Layered Models)

Duan, Guindani, & Gelfand, *Generalized Spatial DP*, 2007
Spatially Dependent Pitman-Yor

- Cut random *surfaces* (samples from a GP) with *thresholds* (as in Level Set Methods)

- Assign each pixel to the *first* surface which exceeds threshold (as in Layered Models)

- Retains *Pitman-Yor marginals* while jointly modeling rich *spatial dependencies* (as in Copula Models)
Spatially Dependent Pitman-Yor

Non-Markov Gaussian Processes:
\[ u_{ki} \sim \mathcal{N}(0, 1) \]
\[ u_{ki} \perp u_{\ell i} \]

PY prior:
\[ v_k \sim \text{Beta}(1 - a, b + ka) \]

Feature Assignments
\[ z_i = \min\{ k \mid u_{ki} < \Phi^{-1}(v_k) \} \]
\[ x_i \sim \text{Mult}(\theta_{z_i}) \]
Samples from PY Spatial Prior

Comparison: Potts Markov Random Field
Outline

Model
- Dependent Pitman-Yor processes
- Spatial coupling via Gaussian processes

Inference
- Stochastic search & expectation propagation

Learning
- Conditional covariance calibration

Results
- Multiple segmentations of natural images
Mean Field for Dependent PY

**Factorized Gaussian Posteriors**

\[ q(u) = \prod_{k=1}^{K} \prod_{i=1}^{N} \mathcal{N}(u_{ki} \mid \mu_{ki}, \lambda_{ki}) \]

\[ q(\bar{v}) = \prod_{k=1}^{K} \mathcal{N}(\bar{v}_k \mid \nu_k, \delta_k) \]

**Sufficient Statistics**

\[ z_i = \min\{k \mid u_{ik} < \bar{v}_k\} \]

Allows closed form update of \( q(\theta_k) \) via

\[ \mathbb{P}_q[u_{ki} < \bar{v}_k] = \Phi \left( \frac{\nu_k - \mu_{ki}}{\sqrt{\delta_k + \lambda_{ki}}} \right) \]

\[ \log p(x \mid \alpha, \rho) \geq H(q) + \mathbb{E}_q[\log p(u, \bar{v}, \theta \mid \alpha, \rho)] \]
Robustness and Initialization

Log-likelihood bounds versus iteration, for many random initializations of mean field variational inference on a single image.
Alternative: Inference by Search

Consider hard assignments of superpixels to layers (partitions)

Integrate likelihood parameters analytically (conjugacy)

Marginalize layer support functions via expectation propagation (EP): approximate but very accurate

No need for a finite, conservative model truncation!
Discrete Search Moves

Stochastic proposals, accepted if and only if they improve our EP estimate of marginal likelihood:

- **Merge**: Combine a pair of regions into a single region
- **Split**: Break a single region into a pair of regions (for diversity, a few proposals)
- **Shift**: Sequentially move single superpixels to the most probable region
- **Permute**: Swap the position of two layers in the order

Marginalization of continuous variables simplifies these moves…
Inference Across Initializations

Mean Field Variational

EP Stochastic Search

Best

Worst

Best

Worst
BSDS: Spatial PY Inference

Spatial PY (EP)

Spatial PY (MF)
Outline

Model
- Dependent Pitman-Yor processes
- Spatial coupling via Gaussian processes

Inference
- Stochastic search & expectation propagation

Learning
- Conditional covariance calibration

Results
- Multiple segmentations of natural images
Covariance Kernels

• Thresholds determine segment *size*: Pitman-Yor
• Covariance determines segment *shape*:

\[ C(y_i, y_j) \quad \text{probability that features at locations} \ (y_i, y_j) \ \text{are in the same segment} \]

Roughly Independent Image Cues:

- Color and texture histograms within each region: Model generatively via multinomial likelihood (Dirichlet prior)
- Pixel locations and *intervening contour* cues: Model conditionally via GP covariance function
Data unavailable to learn models of all the categories we’re interested in: We want to discover new categories!

Use logistic regression, and basis expansion of image cues, to learn binary “are we in the same segment” predictors:

- **Generative**: Distance only
- **Conditional**: Distance, intervening contours, …
From Probability to Correlation

There is an injective mapping between covariance and the probability that two superpixels are in the same segment.

\[ q_k^-(\alpha, \rho) = \int_{-\infty}^{\delta_k} \int_{-\infty}^{\delta_k} \int_{-\infty}^{\delta_k} \mathcal{N} \left( \begin{bmatrix} u_i \\ u_j \end{bmatrix} \parallel \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right) p(\delta_k | \alpha) du_i du_j d\delta_k \]

\[ q_k^+(\alpha, \rho) = \int_{-\infty}^{\delta_k} \int_{-\infty}^{\delta_k} \int_{-\infty}^{\delta_k} \mathcal{N} \left( \begin{bmatrix} u_i \\ u_j \end{bmatrix} \parallel \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right) p(\delta_k | \alpha) du_i du_j d\delta_k \]

\[ p_{ij} = q_1^-(\alpha, \rho) + q_2^-(\alpha, \rho) q_1^+(\alpha, \rho) + q_3^-(\alpha, \rho) q_1^+(\alpha, \rho) q_2^+(\alpha, \rho) + \ldots \]
Low-Rank Covariance Projection

- The pseudo-covariance constructed by considering each superpixel pair independently may not be positive definite.
- Projected gradient method finds low rank (factor analysis), unit diagonal covariance close to target estimates.
Prediction of Test Partitions

Heuristic versus Learned Image Partition Probabilities

Learned Probability versus Rand index measure of partition overlap
Comparing Spatial PY Models

Image | PY Learned | PY Heuristic

- Image
- PY Learned
- PY Heuristic
Outline

Model
- Dependent *Pitman-Yor processes*
- Spatial coupling via *Gaussian processes*

Inference
- Stochastic search & *expectation propagation*

Learning
- Conditional covariance calibration

Results
- Multiple segmentations of natural images
Other Segmentation Methods

- FH Graph
- Mean Shift
- NCuts
- gPb+UCM
- Spatial PY
Quantitative Comparisons

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>PRI</th>
<th>VI</th>
<th>SegCover</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ncuts</td>
<td>0.74</td>
<td>2.5</td>
<td>0.38</td>
</tr>
<tr>
<td>MS</td>
<td>0.77</td>
<td>2.5</td>
<td>0.44</td>
</tr>
<tr>
<td>FH</td>
<td>0.77</td>
<td>2.1</td>
<td>0.52</td>
</tr>
<tr>
<td>gPb</td>
<td>0.81</td>
<td>2.0</td>
<td>0.58</td>
</tr>
<tr>
<td>PYdist</td>
<td>0.72</td>
<td>2.1</td>
<td>0.51</td>
</tr>
<tr>
<td>PYall</td>
<td>0.76</td>
<td>2.1</td>
<td>0.52</td>
</tr>
</tbody>
</table>

Berkeley Segmentation

- On BSDS, similar or better than all methods except gPb
- On LabelMe, performance of Spatial PY is better than gPb

Room for Improvement:

- Implementation efficiency and search run-time
- Histogram likelihoods discard too much information
- Most probable segmentation does not minimize Bayes risk
Multiple Spatial PY Modes

Most Probable
Multiple Spatial PY Modes

Most Probable
Spatial PY Segmentations
Conclusions

**Successful BNP modeling** requires…

- careful study of how model assumptions match data statistics & *model comparisons*
- reliable, consistent (general-purpose?) *inference* algorithms, carefully validated
- methods for *learning* hyperparameters from data, often with partial supervision