Applied Bayesian Nonparametrics

5. Spatial Models via Gaussian Processes, not MRFs

Tutorial at CVPR 2012

Erik Sudderth
Brown University

NIPS 2008: E. Sudderth & M. Jordan, Shared Segmentation of Natural Scenes using Dependent Pitman-Yor Processes.
Human Image Segmentation
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
BNP Image 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
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)

\[ x_i^c \sim \text{Mult}(\theta_z^c) \]
\[ x_i^s \sim \text{Mult}(\theta_z^s) \]

Visual segment appearance model

Color:
\[ \theta_k^c \sim \text{Dir}(\rho^c) \]

Texture:
\[ \theta_k^s \sim \text{Dir}(\rho^s) \]
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, …

Some dependent prior with DP/PY “like” marginals

Assign features to segments

Observed features (color & texture)

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  

- Nonparametric analogs have similar properties

  Woolrich & Behrens, 2006  
  Blei & Lafferty, 2006
Discrete Markov Random Fields

Ising and Potts Models

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

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

Previous Applications

- Interactive foreground segmentation
- Supervised training for known categories

…but learning is challenging, and little success at unsupervised segmentation.
Region Classification with Markov Field Aspect Models

Verbeek & Triggs, CVPR 2007
10-State Potts Samples

States sorted by size: largest in blue, smallest in red
The Ising/Potts model is not well suited to segmentation tasks.

\[ E[N(z)] \]

\[ N(z) \rightarrow \text{number of edges on which states take same value} \]

\[ \beta \rightarrow \text{edge strength} \]

Even within the phase transition region, samples lack the size distribution and spatial coherence of real image segments.
Geman & Geman, 1984

128 x128 grid
8 nearest neighbor edges
K = 5 states
Potts potentials: \( \beta = \frac{2}{3} \)

200 Iterations

10,000 Iterations
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)
Stick-Breaking Revisited

\[ \pi_k = v_k \prod_{\ell=1}^{k-1} (1 - v_\ell) \quad v_k \sim \text{Beta}(1 - a, b + ka) \]

Multinomial Sampler:
- \( u_i \sim \text{Unif}(0, 1) \)
- \( z_i = \text{CDF}_\pi^{-1}(u_i) \)

Sequential Binary Sampler:
- \( b_{ki} \sim \text{Bernoulli}(v_k) \)
- \( z_i = \min\{k \mid b_{ki} = 1\} \)
PY Gaussian Thresholds

Normal CDF

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

\[ \mathbb{P}[\Phi(u_{ki}) < v_k] = v_k \]

because

\[ \Phi(u_{ki}) \sim \text{Unif}(0, 1) \]

Gaussian Sampler:

\[ u_{ki} \sim \mathcal{N}(0, 1) \]

\[ z_i = \min\{k \mid u_{ki} < \Phi^{-1}(v_k)\} \]

Sequential Binary Sampler:

\[ b_{ki} \sim \text{Bernoulli}(v_k) \]

\[ z_i = \min\{k \mid b_{ki} = 1\} \]
PY Gaussian Thresholds

Gaussian Sampler:

$$u_{ki} \sim \mathcal{N}(0, 1)$$

$$z_i = \min\{k \mid u_{ki} < \Phi^{-1}(v_k)\}$$

Sequential Binary Sampler:

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

$$b_{ki} \sim \text{Bernoulli}(v_k)$$

$$z_i = \min\{k \mid b_{ki} = 1\}$$
Spatially Dependent Pitman-Yor Processes:

PY prior:

Feature Assignments

Normal CDF

Non-Markov Gaussian Processes:

\[ u_{k_i} \sim \mathcal{N}(0, 1) \]

\[ u_{k_i} \perp u_{k_i} \]

Segment size

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

Feature Assignments

\[ z_i = \min\{k \mid u_{k_i} < \Phi^{-1}(v_k)\} \]

\[ x_i \sim \text{Mult}(\theta_{z_i}) \]
Preservation of PY Marginals

Why Ordered Layer Assignments?

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

\[ \nu_k = \mathbb{P}(z_i = k \mid z_i \neq k - 1, \ldots, 1) \]

Stick Size Prior \xrightarrow{} Random Thresholds

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

\[ \bar{\nu}_k = \Phi^{-1}(\nu_k) \]

\[ \mathcal{N}(0, 1) \]
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

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

\[ \log p(x \mid \alpha, \rho) \geq H(q) + \mathbb{E}_q[\log p(u, \bar{v}, \theta \mid \alpha, \rho)] \]
Mean Field for Dependent PY

Updating Layered Partitions

Evaluation of beta normalization constants:
\[ \mathbb{E}_q[\log \Phi(\tilde{v}_k)] \leq \log \mathbb{E}_q[\Phi(\tilde{v}_k)] \]
\[ = \log \Phi \left( \frac{\nu_k}{\sqrt{1 + \delta_k}} \right) \]

Jointly optimize each layer’s threshold and Gaussian assignment surface, fixing all other layers, via backtracking conjugate gradient with line search

Reducing Local Optima

Place factorized posterior on eigenfunctions of Gaussian process, not single features

\[ \log p(x | \alpha, \rho) \geq H(q) + \mathbb{E}_q[\log p(u, \tilde{v}, \theta | \alpha, \rho)] \]
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!
Maximization Expectation

**EM Algorithm**
- E-step: Marginalize latent variables (approximate)
- M-step: Maximize likelihood bound given model parameters

**ME Algorithm**  
Kurihara & Welling, 2009  
- M-step: Maximize likelihood given latent assignments  
- E-step: Marginalize random parameters (exact)

**Why Maximization-Expectation?**
- Parameter marginalization allows Bayesian “model selection”
- Hard assignments allow efficient algorithms, data structures
- Hard assignments consistent with clustering objectives
- *No need for finite truncation of nonparametric models*
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…*
Inferring Ordered Layers

Order A: Front, Middle, Back
Order B: Front, Middle, Back

Which is preferred by a diagonal covariance? Order B
Which is preferred by a spatial covariance? Order A
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

*Berkeley Pb (probability of boundary) detector*
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, …
There is an injective mapping between covariance and the probability that two superpixels are in the same segment.
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 of snowy landscape]
- [Image of mountainous terrain]
- [Image of hiking trail]
- [Image of tiger]

Legend:
- Image: Original image
- PY Learned: Spatial PY model learned
- PY Heuristic: Spatial PY model with a heuristic approach
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>

<table>
<thead>
<tr>
<th></th>
<th>PRI</th>
<th>VI</th>
<th>SegCover</th>
</tr>
</thead>
<tbody>
<tr>
<td>gPb</td>
<td>0.74</td>
<td>2.1</td>
<td>0.53</td>
</tr>
<tr>
<td>PYall</td>
<td>0.73</td>
<td>1.9</td>
<td>0.55</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

*Spatial Pitman-Yor Processes* allow...

- efficient variational *parsing* of scenes into unknown numbers of segments
- empirically justified *power law* priors
- accurate learning of non-local spatial statistics of natural scenes
- promise in other application domains...
Conclusions

…but bravery is required

- Conventional MCMC & variational learning prone to local optima, hard to scale to large datasets.
  
  *But better methods on the way!*

- Literature remains fairly technical.
  
  *But growing number of tutorials!*