Indian Buffet Processes with Power-law Behavior
Yee Whye Teh & Dilan Görür, Gatsby Unit

IBPs:
Bayesian nonparametric latent variable models with an infinite number of latent variables.

This paper:
A three parameter generalization of IBPs with power-law behavior.

Power-law:
Many natural phenomena exhibit power-law properties.

IBP with parameters $\alpha > 0$, $c > -\sigma$, $0 \leq \sigma < 1$.
- Customer 1 tries Poisson($\alpha$) dishes;
- Customer $n+1$:
  - tries dish $k$ with probability $\frac{m_k - \sigma}{n + c}$, $m_k = \# $ customers who tried dish $k$.
  - tries Poisson($\alpha \frac{\Gamma(1+c)\Gamma(n+c+\sigma)}{\Gamma(n+1+c)\Gamma(c+\sigma)}$) new dishes.

Infinitely exchangeable. The de Finetti measure is a completely random measure called stable-beta process.


Applied to modelling word occurrences in documents.

\[ \alpha=100, c=1, \sigma=0.5 \]
Submodular Maximization under a cardinality constraint

- Appears in many ML problems (feature selection, etc.)
- Main difficulty: Exponentially large number of locally optimal solutions

Submodularity-Cut Algorithm

- Converges to an exact solution in a finite number of iterations
- Can calculate the upper and lower bounds at each iteration
  $\Rightarrow \varepsilon$-optimal solution
Statistical Models of Linear and Non-linear Contextual Interactions in Early Visual Processing

Ruben Coen-Cagli
AECOM

Peter Dayan
GCNU, UCL

Odelia Schwartz
AECOM

CORTICAL NEUROPHYSIOLOGY SIMULATIONS

SALIENCY AND COLLINEAR FACILITATION

GENERATIVE MODEL OF IMAGE STATISTICS

An extension of the GSM that learns

a) non-homogeneity
b) linear correlations

in a database of natural images
STDP enables spiking neurons to detect hidden causes of their inputs.

We prove analytically: STDP in a soft-WTA circuit on population-coded inputs generates a probabilistic clustering of spike patterns.

Expectation Maximization (EM)
Probabilistic Inference of hidden causes
3D Object Recognition with Deep Belief Nets

Vinod Nair and Geoffrey E. Hinton, University of Toronto

\[ h^n \]
\[ h^{n-1} \]
\[ h^1 \]

New things
- Third-order Boltzmann machine model at the top.
- Hybrid gen. + disc. learning algorithm for the top-level model avoids poor mixing.

Results on NORB
- 6.5% error (conv. NN = 5.9%, SVM=11.6%).
- 5.2% error with lots of unlabelled data + semi-supervised learning.
Continuous-time Markov chains

Discrete state
Transition probabilities
Exponential dwell times

Maximum likelihood trajectory inference

From state observed at two times:

From series of noisy/partial observations:

Efficient dynamic programming solutions!

Optimal solutions have surprising properties!
Le Song, Mladen Kolar and Eric P. Xing

Problem: Structurally-varying networks needed for modeling evolving relations between non-stationary multivariate time series.

New Formalism: Time-varying DBNs $X^t = A^t \cdot X^{t-1} + \epsilon$, $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$

Structure Learning: Kernel reweighted $\ell_1$-regularized autoregression

$$\hat{A}_{i.}^t = \arg\min_{A_{i.}^t \in \mathbb{R}^{1 \times n}} \sum_{t=1}^{T} w^*(t) (x^t_i - A_{i.}^t x^{t-1})^2 + \lambda |A_{i.}^t|_1, \ \forall t^* \in [0, 1]$$

Properties: Computational efficiency and provable asymptotic sparsistency

$$\mathbb{P} \left[ \text{supp}(\hat{A}^t) = \text{supp}(A^t) \right] \to 1, \quad T \to \infty, \quad \forall t^* \in [0, 1].$$

Applications: Yeast cell cycle and brain-computer interface data

Time-varying gene regulatory networks

Time-varying brain connectivity
Construction of Nonparametric Bayesian Models from Parametric Bayes Equations

Peter Orbanz

Motivation

- Nonparametric Bayesian (NPB) models = Bayesian models on $\infty$-dimensional space
- Dirichlet process, Gaussian process:
  - Posteriors are conjugate
  - Finite-dimensional discretizations are in exponential family

Question

Which NPB models are conjugate?

Results

- NPB conjugacy $\Leftrightarrow$ Finite-dimensional conjugacy
- NPB model inherits exponential family properties
- “Recipe” for construction from exponential families
- Construction example: Infinite rankings
When local search is used for MAP...

**Error reductions** over likelihood approximations: 75% (mh+cd1) 43% (piecewise)
... and margin methods: 29% (perceptron s.r.)

...and still reach the *global* optimum!
1. Draw Fruits, Vegetable ranks independently

2. Draw interleaving to form full ranking

Riffled Independence Assumptions can:
- Hold in real data
- Be more appropriate than full independence for rankings
- Be exploited for efficient inference, low sample complexity
A Smoothed Approximate LP (Desai, Farias, Moallemi)

**Approximate LP**
- New ‘Relaxed’ LP for Approximate DP
  - Tractable Programs with *substantially stronger guarantees*
- Theory to Practice: *10x improvements* for Tetris over LP Approach

**SMOOTHED LP**
Randomized Pruning
Alexandre Bouchard-Côté, Slav Petrov, and Dan Klein

- Recurrent problem: computation of expectations over large-scale combinatorial spaces
  - High-degree dynamic programs
- Standard solution: a single pruning mask
  - Introduces bias
  - Requires tuning thresholds
- New solution: sequence of masks
  - Auxiliary variable construction
  - Easy to trade-off time and quality
  - Outperforms standard pruning heuristics
Subject independent EEG-based BCI decoding

Siamac Fazli\textsuperscript{a,b}, Cristian Grozea\textsuperscript{a}, Florin Popescu\textsuperscript{a}, Márton Danóczy\textsuperscript{a,b}, Benjamin Blankertz\textsuperscript{a,b}, Klaus-Robert Müller\textsuperscript{b}

\textsuperscript{a}Fraunhofer First, Kekuléstr. 7, 12489 Berlin, Germany
\textsuperscript{b}Technical University, Franklinstr. 28/29, 10587 Berlin, Germany

83 sessions of movement imaginations

\[ \downarrow \]

ensemble generation with subject-dependent filters

\[ \downarrow \]

leave one subject out

sparse regression on subject-dependent filters

\[ \downarrow \]

subject-independent classifier with almost no loss of performance vs. subject-tuned classifiers

\[
\begin{array}{|c|c|c|c|}
\hline
\text{method} & \text{LSR-}\ell_1 & \text{Lap} & \text{BP} \\
\hline
#<25\% & 36 & 24 & 11 \\
25\%-tile & 16.0 & 22.0 & 31.3 \\
median & 29.3 & 34.7 & 38.7 \\
75\%-tile & 40.7 & 45.3 & 45.3 \\
\hline
\end{array}
\]