Graphical Models, Bayesian Method, Sampling, and Variational Inference
With Application in Function MRI Analysis and Other Imaging Problems

Wei Liu

Scientific Computing and Imaging Institute
University of Utah

September 19, 2013
Table of Contents

1 Motivation

2 Graphical Model, Markov Random Field, Sampling and Inference.
   - Markov Random Field
   - Sampling
   - Variational Inference

3 Applications
   - A Warm-Up: Pairwise Connectivity with Six Dimensional MRF
   - Hierarchical Model
   - Traumatic Brain Injury Image Segmentation with Active Learning
Outline for section 1

1 Motivation

2 Graphical Model, Markov Random Field, Sampling and Inference.
   - Markov Random Field
   - Sampling
   - Variational Inference

3 Applications
   - A Warm-Up: Pairwise Connectivity with Six Dimensional MRF
   - Hierarchical Model
   - Traumatic Brain Injury Image Segmentation with Active Learning
Why Resting-State fMRI

- Large energy consumption.
- Matches existing neuro-anatomical systems.
- Reflect increased and decreased activity in task.
- Predict task response as a priori hypothesis.
fMRI Data Acquisition

- fMRI is 4D. Many consecutive 3D volumes.
- BOLD signal.
- Spatial dependency.
- Temporal correlation.
- fMRI is noisy.
- Experiment stimulus signal.
- Subjects undertake cognitive tasks.
- General linear model is used for multi-regression analysis between stimulus and BOLD signal of a voxel.

- No experiment paradigm signal.
- Subject stay in scanner. Eyes closed/open to a fixation cross.
- Correlation analysis between two voxels.

Figure: M. Fox, Nat. Rev., Neuroscience
Current Group Analysis Methods.

- Arbitrary spatial blurring to enforce spatial dependency.
- Lack of methods for jointly estimate group and subjects.
- Variability analysis.
State-of-Art Group Analysis Approach

**Bottom-up Approach (Heuvel, 2008; Craddock, HBM, 2011)**
- Estimate Subject network first.
- Estimate group network from subjects.

**Top-down Approach (Calhoun, HBM, 2001)**
- Estimate group network from all subjects.
- Back-reconstruct subject network maps.
Outline for section 2

1 Motivation

2 Graphical Model, Markov Random Field, Sampling and Inference.
   • Markov Random Field
   • Sampling
   • Variational Inference

3 Applications
   • A Warm-Up: Pairwise Connectivity with Six Dimensional MRF
   • Hierarchical Model
   • Traumatic Brain Injury Image Segmentation with Active Learning
Definition

\( \mathcal{G} = (\mathcal{V}, \mathcal{E}) \): undirected graph.

\( s \in \mathcal{V} \): a node site in \( \mathcal{V} \).

\( X = \{x_1, \ldots, x_s, \ldots\} \): a collection of random variables defined on graph \( \mathcal{G} \).

\( \mathcal{N}_s \): the set of sites neighboring \( s \).

\( (r, s) \in \mathcal{E} \iff r \in \mathcal{N}_s \).

Definition

A Markov Random Field is a collection of variables \( X \) defined on graph \( \mathcal{G} = (\mathcal{V}, \mathcal{E}) \) if for all \( s \in \mathcal{V} \)

\[
P(X_s|X_{\mathcal{V} - s}) = P(X_s|X_{\mathcal{N}_s})
\]
The definition of MRF is a local property.

Theorem (Hammersley-Clifford, 1971)

$X$ is an MRF on $G$ if and only if $X$ obeys Gibbs distribution in the following form

$$P(X) = \frac{1}{Z} \exp \left( -\frac{1}{T} U(X) \right),$$

$$U(X) = \sum_{c \in C} V_c(X).$$

Gibbs distribution gives a global property that can be used as a prior distribution.
The observed time series $Y$ can be seen as *generated* from the hidden variables $X$.

- $X$ is MRF to guarantee smoothness.
- Inverse problem: Given $Y$, estimate $X$.

- other forms exist: conditional random fields.
- No Bayesian interpretation.
Metropolis sampling: choose a candidate, accept the candidate based on the energy change.
Simulation

(a) $\beta = 0.8$

(b) $\beta = 0.88$

(c) $\beta = 1.0$

(d) $\beta = 1.5$

(e) $\beta = 2.0$

(f) $\beta = 0.88$ details
Variation Inference

- Assumption of factorization: $q(X) = \prod_s q_s(x_s)$
- Estimate the posterior form a constrained search space.
- Iteratively update each factor:
  \[
  \log q_s(x_s) = \mathbb{E}_{r \neq s}[\log p(X, Y)] + \text{const} \text{ until convergence.}
  \]
- Also apply to Bayesian settings.
Outline for section 3

1 Motivation

2 Graphical Model, Markov Random Field, Sampling and Inference.
   - Markov Random Field
   - Sampling
   - Variational Inference

3 Applications
   - A Warm-Up: Pairwise Connectivity with Six Dimensional MRF
   - Hierarchical Model
   - Traumatic Brain Injury Image Segmentation with Active Learning
The Goal

- The Connectivity between each pair of voxels.
- No Seed region needed.
- Spatial smoothness as a regularization, without blurring.
- Learn the strength of the smoothness from the data.
Solution

- Define a 6 dimension graph $\mathcal{G}$. Define pairwise connectivity variable on each node. Add an edge $(x_{ij}, x_{st})$ if any voxels between $i, j$ and $s, t$ are neighbors.
- Likelihood: Gaussian $\mathcal{N}(\mu, \sigma^2)$.
- Two class segmentation: no connectivity and connectivity.
- Gibbs sampling and mean field approximation to compute posterior mean of $X$. 

Voxel correlations $y$ in 2d dimensional space

Connectivity map $x$ in 2d dimensional MRF

Original d dimensional image space
Figure: Correlation map without smoothing; With smoothing; Posterior from MRF
Some Computation Issues

- GPU (CUDA) programming.
- Checkerboard effects.

![Diagram showing checkerboard effects and an image of a flipped pattern.]
- A hierarchical structure including group and subject.
- Estimate both group and subject jointly.
- Model Inter-subject variation and within-subject coherence.
- Use Markov random field with a single graph.
- Bayesian method. Parameter estimation.
Group network map inform subjects as a prior.
Subject network maps feedback into group estimation.
Jointly estimate both levels.
Spatial coherence is again modeled by MRF.
A new MRF including both levels

Put group and all subject network label variables in a single graph.
Definition

\[ \mathcal{G}_G = (\mathcal{V}_G, \mathcal{E}_G) : \text{Graph that represents group map.} \]

\[ \mathcal{G}_H^j = (\mathcal{V}_H^j, \mathcal{E}_H^j), \forall j = 1, \ldots, J : \text{subject map.} \]

\[ \mathcal{G} = (\mathcal{V}, \mathcal{E}) : \text{new graph that includes both group and subject level.} \]

\[ \mathcal{V} = (\mathcal{V}_H^1, \mathcal{V}_H^2, \ldots, \mathcal{V}_H^J), \]

\[ \mathcal{E} = \{(r, s) | (r, s) \in \mathcal{E}_G \} \cup \{(r, s) | r \in \mathcal{V}_G, s \in \mathcal{V}_H^j, r \simeq s \} \cup \{(r, s) | (r, s) \in \mathcal{E}_H^j \}. \]
Energy Function

$$U(X) = \sum_{(s,r) \in \mathcal{E}_H} \beta \psi(x_s, x_r)$$

$$+ \sum_{j=1}^{J} \left( \sum_{s \in \mathcal{V}_G, \tilde{s} \in \mathcal{V}_H^j} \alpha \psi(x_s, x_{\tilde{s}}) \right)$$

$$+ \sum_{(s,r) \in \mathcal{E}_H^j} \beta \psi(x_s, x_r) \right).$$
y_s: normalized time series in p-sphere.

\[
P(y_s \mid X) = C_p(\kappa_\ell) \exp(\kappa_\ell \mu_\ell^\top y_s), \, y_s \in S^{p-1}.
\]

\[
\log P(Y \mid X) = \sum_{s \in H} \log p(y_s \mid x_s)
\]
Monte-Carlo Sampling used to approximate $\mathbb{E}_{X|Y}[\log P(X, Y; \theta)]$

Gibbs sampling also in a multi-level fashion.
The Algorithm

**Data:** Normalized fMRI, initial group label map

**Result:** Group and subjects label map $X$, parameters

$$\{\alpha, \beta, \mu, \sigma\}$$

**while** $\mathbb{E}[\log p(X, Y)]$ **not converge** **do**

**repeat**

**foreach** $s \in V_G$ **do**

Draw consecutive samples of $x_s$;

**foreach** $j = 1 \ldots J$ **do**

**foreach** $s \in V_H^j$ **do**

Draw consecutive samples of $x_s$;

end

Save sample $Y^m$ after $B$ burn-ins;

**until** $B + M$ **times**;

**foreach** $l = 1 \ldots L$ **do**

Estimate $\{\mu_l, \kappa_l\}$ by maximizing

$$\frac{1}{M} \sum_{m=1}^{M} \log p(Y|X^m);$$

Estimate $\{\alpha, \beta\}$ by maximizing

$$\frac{1}{M} \sum_{m=1}^{M} \log p(X^m);$$

end
Synthetic Data Experiment

Truth | K-Means | N-Cuts | groupmrf
--- | --- | --- | ---
sub 1 | | | |
K-means | 92.9 | 87.0 | 0.67
N-Cuts | 85.4 | 87.1 | 0.58
groupmrf | 95.7 | 97.5 | 0.59
Estimated Maps without Hierarchical Structure

Subject 1

Subject 2

Subject 3
Real rs-fMRI Data

K-Means

N-Cuts

groupmrf

$z = 26$  $z = 54$

$z = 26$  $z = 54$

$z = 26$  $z = 54$
No prior knowledge of lesion.
Multi-modality, longitudinal data. Complex patterns.
Existing algorithm: high false-positive and false-negative.
A slight user involvement significantly improves result.
Computer should be active, user will be passive. (less burden).
- Semi-supervised approach based on graph-cuts.
- Tight control of false positive rate.
- Using MRF to represent soft-constraints: within normal, lesion, and the boundary.
- Compute query score: posterior ratio (logistic ratio)
Active Learn TBI Images.

User initialization with a bounding box.

Active learn good candidate objects, followed by self-training/active-learning.
Thank you.
This ends the talk.