Statistical Machine Learning in Markov Random Field
and Expansion to Data Sciences

Kazuyuki Tanaka

GSIS, Tohoku University, Sendai, Japan

http://www.smapip.is.tohoku.ac.jp/~kazu/

Collaborators
- Muneki Yasuda (Yamagata University, Japan)
- Masayuki Ohzeki (Tohoku University, Japan)
- Shun Kataoka (Tohoku University, Japan)
- Candy Hsu (National Tsing Hua University, Taiwan)
- Mike Titterington (University of Glasgow, UK)
Outline

1. Introduction
2. Probabilistic Graphical Models and Loopy Belief Propagation
3. Statistical Machine Learning in Markov Random Fields
4. Generalized Sparse Prior for Image Modeling
5. Probabilistic Noise Reduction
6. Probabilistic Image Segmentation
7. Statistical Machine Learning by Inverse Renormalization Group Transformation
8. Related Works
9. Concluding Remarks
Pairwise Markov Random Fields and Loopy Belief Propagation

Probabilistic Information Processing

Bayes Formulas
Maximum Likelihood
KL Divergence

Markov Random Fields
Probabilistic Models and Statistical Machine Learning

Loopy Belief Propagation

\[
P(a) \propto \prod_{\{i,j\} \in E} W_{ij}(a_i, a_j)
\]

\(V\): Set of all the nodes (vertices) in graph \(G\)
\(E\): Set of all the links (edges) in graph \(G\)
Pairwise Markov Random Fields and Loopy Belief Propagation

\( a = (a_1, a_2, \ldots, a_{|V|}) \): State Vector

\[
P(a) = \frac{1}{Z} \prod_{\{i, j\} \in E} W_{ij}(a_i, a_j)
\]

\[
P_i(a_i) \equiv \sum_{a \setminus \{a_i\}} P(a) \equiv P_{ij}(a_i, a_j)
\]

\[
P_{ij}(a_i, a_j) = \sum_{a \setminus \{a_i, a_j\}} P(a) \equiv P_i(a_i)
\]

\[
\sum_{a_j} P_{ij}(a_i, a_j) = P_i(a_i)
\]

\( V \): Set of all the nodes (vertices) in graph \( G \)

\( E \): Set of all the links (edges) in graph \( G \)

\( G = (V, E) \)
Pairwise Markov Random Fields and Bayes Inference

Data Generative Model

\[ P(d|a) \propto \left( \prod_{i \in V} W_i(a_i, d_i) \right) \]

Prior Probability ⇐ Pairwise MRF

\[ P(a) \propto \left( \prod_{\{i,j\} \in E} W_{ij}(a_i, a_j) \right) \]

Posterior Probability ⇐ Pairwise MRF

\[ P(a|d) = \frac{P(a, d)}{P(d)} \]

Maximum A Posteriori (MAP) Estimation

\[ \hat{a} = \arg \max_a P(a|d) \]

Maximum Posterior Marginal (MPM) Estimation

\[ \hat{a}_i = \arg \max_{a_i} P_i(a_i|d) \quad (\forall i \in V) \]

\[ V: \text{Set of all the nodes (vertices)} \]

\[ E: \text{Set of all the links (edges)} \]
Hyperparameter Estimation of Pairwise Markov Random Fields

Data Generative Model

\[ P(d|a, \beta) \propto \left( \prod_{i \in V} W_i(a_i, d_i|\beta) \right) \]

Prior Probability \& Pairwise MRF

\[ P(a|\alpha) \propto \left( \prod_{\{i,j\} \in E} W_{ij}(a_i, a_j|\alpha) \right) \]

Joint Probability

\[ P(a, d|\alpha, \beta) = P(d|a, \beta)P(a|\alpha) \]

Probability of Data d \rightarrow Likelihood of \alpha and \beta

\[ P(d|\alpha, \beta) = \sum_a P(a, d|\alpha, \beta) \]

Maximization of Marginal Likelihood

\[ (\hat{\alpha}, \hat{\beta}) = \arg \max_{(\alpha, \beta)} P(d|\alpha, \beta) \]

\( V \): Set of all the nodes (vertices)

\( E \): Set of all the links (edges)

December, 2016
Yokohama National University
In Bayesian image modeling, natural images are often assumed to be generated by according to the following pairwise Markov random fields:

\[
\Pr\{F = f\} = P(f) \propto \prod_{\{i,j\} \in E} U(f_i - f_j) = \exp\left(-\sum_{\{i,j\} \in E} (-\ln(U(f_i - f_j)))\right)
\]

In this model, each pixel \( f_i \) is a state variable of light intensity at the \( i \)-th pixel. The set of all the pixels is \( V = \{1,2,3,4,5,6,7,8,9,10,11,12\} \), and the set of all the nearest neighbour pairs of pixels is:

\[
E = \{\{1,2\}, \{2,3\}, \{3,4\}, \{5,6\}, \{6,7\}, \{7,8\}, \{9,10\}, \{10,11\}, \{11,12\}, \{1,5\}, \{2,6\}, \{3,7\}, \{4,8\}, \{5,9\}, \{6,10\}, \{7,11\}, \{8,12\}\}
\]
Conventional Pairwise Markov Random Fields

\[ \Pr \{ F = f \} = P(f) \propto \prod_{\{i,j\} \in E} U(|f_i - f_j|) = \exp \left( - \sum_{\{i,j\} \in E} \ln(U(|f_i - f_j|)) \right) \]

Gaussian Prior
\[- \ln(U(x)) \propto x^2\]

Huber Prior
\[- \ln(U(x)) \propto \begin{cases} 
  |x|^2 & (x < c) \\
  2c|x| - c^2 & (x > c) 
\end{cases}\]

Saturated Quadratic Prior
\[- \ln(U(x)) \propto \frac{x^2}{x^2 + c^2}\]

Convex Functions near \(x=0\)
Supervised Learning of Pairwise Markov Random Fields

Supervised Learning Scheme by Loopy Belief Propagation in Pairwise Markov random fields:

\[ P(f) \propto \exp \left( - \sum_{\{i,j\} \in E} \left( - \ln U(f_i - f_j) \right) \right) \]

30 Standard Images

Supervised Learning

December, 2016

Yokohama National University
Supervised Learning of Pairwise Markov Random Fields by Loopy Belief Propagation:

**Supervised Learning Scheme by Loopy Belief Propagation in Pairwise Markov random fields:**

\[
P(f) \propto \prod_{\{i,j\} \in E} U(|f_i - f_j|), \quad \ln U(|f_i - f_j|) = \frac{1}{2} \sum_{m=0}^{255} \sum_{n=0}^{255} K_{m,n} \Phi_m(f_i) \Phi_n(f_j)
\]

\[
\sum_{f_i=0}^{255} \sum_{f_j=0}^{255} \Phi_m(f_i) \Phi_n(f_i) = \delta_{m,n}, \quad \Phi_0(f_i) \equiv \frac{1}{\sqrt{256}}
\]

\[
K_{m,0} = K_{0,m} = -3 \sum_{f_i=0}^{255} \Phi_m(f_i) \ln P(f_i)
\]

\[
K_{m,n} = K_{n,m} = \sum_{f_i=0}^{255} \sum_{f_j=0}^{255} \Phi_m(f_i) \Phi_0(f_j) P(f_i, f_j)
\]

*Histogram from Supervised Data*

Supervised Learning Scheme by Loopy Belief Propagation in Pairwise Markov random fields:

\[
P(f) \propto \exp\left(-\sum_{\{i,j\} \in E} (-\ln U(|f_i - f_j|))\right) \sim \exp\left(-2.35 \times \sum_{\{i,j\} \in E} |f_i - f_j|^{0.32}\right)
\]

\(q=256\)

\(f_i \in \{0, 1, 2, \ldots, 255\}\)

Bayesian Image Modeling by Generalized Sparse Prior

Assumption: Prior Probability is given as the following Gibbs distribution with the interaction $\alpha$ between every nearest neighbour pair of pixels:

$$\text{Pr}\{F = f \mid p, \alpha\} = P(f|p,\alpha) = \frac{1}{Z(p,\alpha)} \exp\left(-\frac{1}{2} \alpha \sum_{\{i,j\} \in E} |f_i - f_j|^p\right)$$

- $i$-th pixel
- $f_i$ State Variable of Light Intensity at $i$-pixel
- $f = (f_1, f_2, \ldots, f_{12})^\top$ State Vector of Original Image
- $V = \{1,2,3,4,5,6,7,8,9,10,11,12\}$ Set of All the pixels
- $E = \{\{1,2\}, \{2,3\}, \{3,4\}, \{5,6\}, \{6,7\}, \{7,8\}, \{9,10\}, \{10,11\}, \{11,12\}, \{1,5\}, \{2,6\}, \{3,7\}, \{4,8\}, \{5,9\}, \{6,10\}, \{7,11\}, \{8,12\}\}$ Set of All the Nearest Neighbour Pairs of Pixels

$p=0$: Potts model
$p=2$: Discrete Gaussian Graphical Model
Prior Analysis by LBP and Conditional Maximization of Entropy in Generalized Sparse Prior

Log-Likelihood for $p$ and $\alpha$ when the original image $f^*$ is given

\[
\left( p^*, \alpha^* \right) = \arg \max_{(p, \alpha)} \ln \Pr \{ F = f^* \mid p, \alpha \}
\]

\[
\Pr \{ F = f \mid p, \alpha \} \propto \exp \left( -\frac{1}{2} \alpha \sum_{\{i,j\} \in E} |f_i - f_j|^p \right)
\]

Degradation Process in Bayesian Image Modeling

Assumption: Degraded image is generated from the original image by Additive White Gaussian Noise.

Pr\{Original Image | Degraded Image\} \propto Pr\{Degraded Image | Original Image\} \times Pr\{Original Image\}

Pr\{G = g | F = f, \sigma\} \propto \prod_{i \in \Omega} \exp\left(-\frac{1}{2\sigma^2} (g_i - f_i)^2\right)

\sigma > 0

● : Pixel of Original Image
\[ f = (f_1, f_2, \ldots, f_{12})^T \]

■ : Pixel of Degraded Image
\[ g = (g_1, g_2, \ldots, g_{12})^T \]
Hyperparameter Estimation of Pairwise Markov Random Fields

\( d = (d_1, d_2, \ldots, d_{|V|}) \): Data

\( a = (a_1, a_2, \ldots, a_{|V|}) \): Parameter

**Data Generative Model**

\[
P(d | a, \beta) \propto \left( \prod_{i \in V} W_i(a_i, d_i | \beta) \right)
\]

**Prior Probability ↔ Pairwise MRF**

\[
P(a | \alpha) \propto \left( \prod_{\{i, j\} \in E} W_{ij}(a_i, a_j | \alpha) \right)
\]

**Joint Probability**

\[
P(a, d | \alpha, \beta) = P(d | a, \beta)P(a | \alpha)
\]

**Probability of Data \( d \) → Likelihood of \( \alpha \) and \( \beta \)**

\[
P(d | \alpha, \beta) = \sum_a P(a, d | \alpha, \beta)
\]

**Maximization of Marginal Likelihood**

\[
(\hat{\alpha}, \hat{\beta}) = \arg \max_{(\alpha, \beta)} P(d | \alpha, \beta)
\]

\( V \): Set of all the nodes (vertices)

\( E \): Set of all the links (edges)

**Expectation Maximization (EM) Algorithm**
Noise Reductions by Generalized Sparse Priors and Loopy Belief Propagation

Original Image

Degraded Image

Restored Image

$p=0.5$

$p=2.0$

Continuous Gaussian Graphical Model

$p=2.0$


Bayesian Modeling for Image Segmentation

Image Segmentation using MRF

Segment image data into some regions using belief propagation and EM algorithm

\[
d_i = \begin{pmatrix} R_i \\ G_i \\ B_i \end{pmatrix}
\]

Parameter \( a_i = 0, 1, \ldots, q-1 \)

\[
P(a|d, K, m(0), m(1), \ldots, m(q-1), C(0), C(1), \ldots, C(q-1))
\]

Posterior Probability Distribution

\[
\alpha \left( \prod_{j \in V} \frac{1}{\det(2\pi C(a_i))} \exp \left( -\frac{1}{2} (d_i - m(a_i))C^{-1}(a_i)(d_i - m(a_i)) \right) \right) \prod_{\{i, j\} \in E} \exp \left( \frac{1}{2} K \delta(a_i, a_j) \right)
\]

Potts Prior

\[
\hat{K}, \hat{m}(0), \hat{m}(1), \ldots, \hat{m}(q-1), \hat{C}(0), \hat{C}(1), \ldots, \hat{C}(q-1)
\]

\( \leftarrow \arg \max_{K, \{m(l), C(l)|l=0,1,\ldots,q-1\}} P(d|K, m(0), m(1), \ldots, m(q-1), C(0), C(1), \ldots, C(q-1)) \)

Likelihood of Hyperparameter

\[
\hat{a}_i \leftarrow \arg \max_{a_i \in \{0,1,\ldots,q-1\}} P_i(a_i|d, \hat{K}, \hat{m}(0), \hat{m}(1), \ldots, \hat{m}(q-1), \hat{C}(0), \hat{C}(1), \ldots, \hat{C}(q-1)) (\forall i \in V)
\]

Maximization of Posterior Marginal (MPM)

Berkeley Segmentation Data Set 500 (BSDS500),
http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/

Yokohama National University
Bayesian Image Segmentation

$$P(a) \propto \prod_{\{i,j\} \in E} \exp\left(\frac{1}{2} K\delta(a_i, a_j)\right)$$

Hyperparameter

$$q = 8$$

Hyperparameter Estimation in Maximum Likelihood Framework and Maximization of Posterior Marginal (MPM) Estimation with Loopy Belief Propagation for Observed Color Image $$d$$


Intel® Core™ i7-4600U CPU with a memory of 8 GB

$$\hat{K} = 3.0301$$

1331Sec

Berkeley Segmentation Data Set 500 (BSDS500),
http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/
P. Arbelaez, M. Maire, C. Fowlkes and J. Malik:
Bayesian Image Segmentation
Coarse Graining

\[ P(a_1, a_2, \ldots, a_N) \propto \prod_{i=1}^{N} \exp\left(\frac{1}{2} K \delta(a_i, a_{i+1})\right) \]

\[ a_i \in \{0, 1, 2, \ldots, q - 1\} \]

\[ a_{N+1} = a_1 : \text{Periodic Boundary Condition} \]

Step 1: \( \exp\left(\frac{1}{2} K^{(1)} \delta(a_{i-1}, a_{i+1})\right) \propto \sum_{a_i = 0}^{q-1} \exp\left(\frac{1}{2} K \delta(a_{i-1}, a_i)\right) \exp\left(\frac{1}{2} K \delta(a_i, a_{i+1})\right), \ (i = 2, 4, 6, \ldots) \)

Step 2: \( (a_0, a_2, a_4, \ldots, a_N) \rightarrow (a_0, a_1, a_2, \ldots, a_{N/2}) \)

\[ P^{(1)}(a_1, a_2, \ldots, a_{N/2}) \propto \prod_{i=1}^{N/2} \exp\left(\frac{1}{2} K^{(1)} \delta(a_i, a_{i+1})\right) \]

\[ K^{(1)} = 2 \ln \left( \frac{q - 1 + e^K}{q - 2 + 2e^{K/2}} \right) \]
Coarse Graining

\[ P(a_1, a_2, \cdots, a_N) \propto \prod_{i=1}^{N} \exp \left( \frac{1}{2} K \delta(a_i, a_{i+1}) \right) \]

\[ P^{(2)}(a_1, a_2, \cdots, a_{N/4}) \propto \prod_{i=1}^{N/4} \exp \left( \frac{1}{2} K^{(2)} \delta(a_i, a_{i+1}) \right) \]

\[ K \to K^{(1)} = 2 \ln \left( \frac{q - 1 + e^K}{q - 2 + 2e^{K/2}} \right) \to K^{(2)} = 2 \ln \left( \frac{q - 1 + e^{K^{(1)}}}{q - 2 + 2e^{K^{(1)}/2}} \right) \to K^{(2)} \]

If \( K^{(2)} \) is given, the original value of \( K \) can be estimated by iterating

\[ K^{(2)} \to K^{(1)} = 2 \ln \left( e^{K^{(2)}/2} + \sqrt{e^{K^{(2)}/2} + q - 1}(e^{K^{(2)}/2} - 1) \right) \to K = 2 \ln \left( e^{K^{(1)}/2} + \sqrt{e^{K^{(1)}/2} + q - 1}(e^{K^{(1)}/2} - 1) \right) \to K \]
Coarse Graining

\[
K^{(r-1)} = 2 \ln \left( e^{K^{(r)} / 4} + \sqrt{ e^{K^{(r)} / 4} + q - 1 (e^{K^{(r)} / 4} - 1) } \right)
\]

\[
K^{(r)} = 4 \ln \left( \frac{q - 1 + e^{K^{(r-1)}}}{q - 2 + 2e^{K^{(r-1)} / 2}} \right)
\]

\[y = 4 \ln \left( \frac{q - 1 + e^x}{q - 2 + 2e^{x / 2}} \right)
\]
Bayesian Image Segmentation

\[ P(a) \propto \prod_{\{i,j\} \in E} \exp\left( \frac{1}{2} K \delta(a_i, a_j) \right) \]

Potts Prior

Hyperparameter Estimation in
Maximization of Marginal Likelihood
with Belief Propagation for Original Image

Hyperparameter Estimation in
Maximization of Marginal Likelihood
with Belief Propagation after Coarse Graining Procedures

Hyperparameter

\[ \hat{K} = 3.0301 \]
1331Sec

\[ \hat{K} = 3.6765, 128Sec \]

Coarse Graining Procedures \((r = 8)\)

20 x 30 Labeled Image

\[ \hat{K}^{(8)} = 2.5288 \]
\[ \hat{K}^{(6)} = 3.3833 \]
\[ \hat{K}^{(4)} = 3.6084 \]
\[ \hat{K}^{(2)} = 3.6634 \]
\[ \hat{K} = 3.6765 \]

20 x 30 Labeled Image

Segmentation by Belief Propagation for Original Image

December, 2016
Yokohama National University
Bayesian Image Segmentation

\[ P(a) \propto \prod_{\{i,j\} \in E} \exp \left( \frac{1}{2} K \delta(a_i, a_j) \right) \]

Hyperparameter Estimation in Maximum Likelihood Framework with Belief Propagation after Coarse Graining Procedures

Hyperparameter Estimation in Maximum Likelihood Framework with Belief Propagation for Original Image

Intel® Core™ i7-4600U CPU with a memory of 8 GB

Berkeley Segmentation Data Set 500 (BSDS500), http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/

\[ \hat{K}^{(8)} = 2.5068 \]

\[ \hat{K}^{(6)} = 3.3770 \]

\[ \hat{K}^{(4)} = 3.6069 \]

\[ \hat{K}^{(2)} = 3.6603 \]

\[ \hat{K} = 3.6765, 101\text{Sec} \]

\[ \hat{K} = 2.9396, 1728\text{Sec} \]
Our related Works

Image Inpainting

Reconstruction of Traffic Density


Summary

- Bayesian Image Modeling by Loopy Belief Propagation and EM algorithm.
- By introducing Inverse Real Space Renormalization Group Transformations, the computational time can be reduced.

It is expected that prior can be learned from data base set of ground truths.

Berkeley Segmentation Data Set 500 (BSDS500),
http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/
References


4. 田中和之 (分担執筆): 確率的グラフィカルモデル(鈴木譲, 植野真臣編), 第8章 マルコフ確率場と確率的画像処理, pp.195-228, 共立出版, July 2016