確率的グラフィカルモデルによる統計的機械学習の数理
Statistical Machine Learning by Probabilistic Graphical
and Expansion to Data Sciences

田中和之
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

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

- \( V \): Set of all the nodes (vertices) in graph \( G \)
- \( E \): Set of all the links (edges) in graph \( G \)

November, 2017
IEICE Technical Committee of Nurocomputing
Pairwise Markov Random Fields and Loopy Belief Propagation

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

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

\[ V \]

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

\[ E \]

Message

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

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

\[ a = (a_1, a_2, \ldots, a_V) \] State Vector

\[ a_i \]
Pairwise Markov Random Fields
and Bayes Inference

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

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

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

\( V \): Set of all the nodes (vertices)
\( E \): Set of all the links (edges)

Simulated Annealing
Graph Cut

Loopy Belief Propagation
Markov Chain Monte Carlo

\( d = (d_1, d_2, \cdots, d_{|V|}) \): Data Point
\( a = (a_1, a_2, \cdots, a_{|V|}) \): Parameter Vector

November, 2017
IEICE Technical Committee of Nurocomputing
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 \leftarrow 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

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

(Data Generative Model)

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

\[ E: \text{Set of all the links (edges)} \]

November, 2017

IEICE Technical Committee of Nurocomputing
In Bayesian image modeling, natural images are often assumed to be generated by according to the following pairwise Markov random fields:

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

Hamiltonian of Classical Spin System

$$f_i \in \{0,1,2,...,255\}$$

$i$-th pixel

State Variable of Light Intensity at $i$-pixel

$$f = (f_1, f_2, \ldots, f_{12})^T$$ 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
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:

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

\[ Q(f) \propto \frac{1}{|D|} \sum_{l \in D} \left( \prod_{i \in V} \delta(f_i, d_i(l)) \right) \]

Empirical Distribution

30 Standard Images \( d(l) = (d_1(l), d_2(l), \ldots, d_{|V|}(l)) \) \( (l=1,2,\ldots,30) \)
Supervised Learning Scheme by Loopy Belief Propagation in Pairwise Markov random fields:

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

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

\[
K_{m,0} = K_{0,m} = -3 \sum_{f_i=0}^{q-1} \Phi_m(f_i) \ln P^{(1)}(f_i)
\]

\[
+4 \sum_{f_i=0}^{q-1} \sum_{f_j=0}^{q-1} \Phi_m(f_i) \Phi_0(f_j) P^{(2)}(|f_i - f_j|)
\]

\[
K_{m,n} = K_{n,m} = \sum_{f_i=0}^{q-1} \sum_{f_j=0}^{q-1} \Phi_m(f_i) \Phi_n(f_j) P^{(2)}(|f_i - f_j|)
\]

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

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

\(q=256\)

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

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

$i$-th pixel

$f_i$ State Variable of Light Intensity at $i$-pixel

$f = (f_1, f_2, \ldots, f_{12})^T$ 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
Degradation Process in Bayesian Image Modeling

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

\[
\Pr\{\text{Original Image} \mid \text{Degraded Image}\} \\
\propto \Pr\{\text{Degraded Image} \mid \text{Original Image}\} \\
\times \Pr\{\text{Original Image}\}
\]

\[
\Pr\{G = g \mid F = f, \sigma\} \\
\propto \prod_{i \in \mathbb{I}} \exp\left(-\frac{1}{2\sigma^2} (g_i - f_i)^2\right)
\]

\(\sigma > 0\)

\(f = (f_1, f_2, \ldots, f_{12})^T\)

\(g = (g_1, g_2, \ldots, g_{12})^T\)
Deterministic Equations of $\alpha$ and $\beta$

\[
\frac{1}{|E|} \sum_{\{i,j\} \in E} \sum_{z_i} \sum_{z_j} |z_i - z_j|^p P_{ij}(z_i, z_j | g, p, \alpha, \beta) = \frac{1}{|E|} \sum_{\{i,j\} \in E} \sum_{z_i} \sum_{z_j} |z_i - z_j|^p P_{ij}(z_i, z_j | p, \alpha)
\]

\[
\frac{1}{|V|} \sum_{i \in V} \sum_{z_i} (z_i - g_i)^2 P_i(z_i | g, p, \alpha, \beta) = \frac{1}{\beta}
\]

Repeat E step and M step until $\alpha(t)$ and $\sigma(t)$ converge

[E Step] Compute $u(t + 1) \leftarrow \frac{1}{|E|} \sum_{\{i,j\} \in E} \sum_{z_i} \sum_{z_j} |z_i - z_j|^p P_{ij}(z_i, z_j | g, p, \alpha(t), \beta(t))$

and $\beta(t + 1) \leftarrow \left( \frac{1}{|V|} \sum_{i \in V} \sum_{z_i} (z_i - g_i)^2 P_i(z_i | g, p, \alpha(t), \beta(t)) \right)^{-1}$

[M Step] Solve $\frac{1}{|E|} \sum_{\{i,j\} \in E} \sum_{z_i} \sum_{z_j} |z_i - z_j|^p P_{ij}(z_i, z_j | p, \alpha(t + 1)) = u(t + 1)$ w.r.t. $\alpha(t + 1)$

[MPM Step] Update $t \leftarrow t + 1$ and $\hat{f}_i(g, p) \leftarrow \text{arg max}_{z_i} P_i(z_i | g, p, \alpha(t), \beta(t))$ (\forall i \in V)
Hyperparameter Estimation of Pairwise Markov Random Fields

\[ d = (d_1, d_2, \cdots, d_{|V|}) : \text{Data} \]

\[ a = (a_1, a_2, \cdots, a_{|V|}) : \text{Parameter} \]

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 \( \beta \)

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

Marginal Likelihood

Maximization of Marginal Likelihood

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

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

\[ E: \text{Set of all the links (edges)} \]
Noise Reductions by Generalized Sparse Priors and Loopy Belief Propagation

Original Image  Degraded Image  Restored Image

$p=0.5$  $p=2.0$

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

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

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

Potts Prior

Data Generative Model

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

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

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

IEICE Technical Committee of Nurocomputing
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 Maximum Likelihood Framework and Maximization of Posterior Marginal (MPM) Estimation with Loopy Belief Propagation for Observed Color Image \( d \)


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

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

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

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

Diagram representation of the process with labeled steps and values.
Coarse Graining

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

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

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

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

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

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

\[
q = 8
\]

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

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

Segmentation by Belief Propagation for Original Image

20 x 30 Labeled Image

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

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

321 x 481

November, 2017
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 Maximum Likelihood Framework with Belief Propagation for Original Image

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

481 x 321

Hyperparameter

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/

30 x 20

Coarse Graining Procedures (r = 8)

30 x 20

Labeled Image

Segmentation by Belief Propagation for Original Image

MPM with LBP

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

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

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

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

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

\[ q = 8 \]

\[ K = 2.9396 \]

1728Sec
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.

Labeled Image by our proposed algorithm

Observed Image

Ground Truth

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