ベイジアンネットワークと確率伝搬法の計算アルゴリズム

Bayesian Network and Computational Algorithm of Belief Propagation

東北大学大学院情報科学研究科応用情報科学専攻

田中和之

(Kazuyuki Tanaka, GSIS, Tohoku University)

kazu@smapip.is.tohoku.ac.jp

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

田中和之: ベイジアンネットワークの統計的推論の数理, コロナ社, 2009.
Contents

1. Introduction
2. Bayesian Network
3. Belief Propagation
4. Interpretation of Belief Propagation in Information Theory
5. Other Practical Applications
6. Summary
Probabilistic Model and Belief Propagation

Bayes Formulas

Probabilistic Information Processing

Belief Propagation

Probabilistic Models

Bayesian Networks

Contents

1. Introduction
2. Bayesian Network
3. Belief Propagation
4. Interpretation of Belief Propagation in Information Theory
5. Other Practical Applications
6. Summary
Joint Probability and Conditional Probability

Probability of Event $A$ \[ \Pr\{A\} \]

Joint Probability of Events $A$ and $B$ \[ \Pr\{A, B\} = \Pr\{A \cap B\} \]

Conditional Probability of Event $A$ when Event $B$ has happened.

\[ \Pr\{B|A\} \equiv \frac{\Pr\{A, B\}}{\Pr\{A\}} \]

\[ \Rightarrow \Pr\{A, B\} = \Pr\{B|A\} \Pr\{A\} \]
Fundamental Probabilistic Theory for Image Processing

Marginal Probability of Event $B$

\[
\Pr\{B\} = \sum_A \sum_C \sum_D \Pr\{A, B, C, D\}
\]

Marginalization
Causal Independence

\[ \Pr\{C \mid A, B\} \equiv \frac{\Pr\{A, B, C\}}{\Pr\{A, B\}} \]

\[ \Pr\{C \mid A = T, B\} = \Pr\{C \mid A = F, B\} \]

\[ \Pr\{C \mid A, B\} = \Pr\{C \mid B\} \]
Fundamental Probabilistic Theory for Image Processing

Causal Independence

\[ \Pr\{D \mid A, B, C\} \equiv \frac{\Pr\{A, B, C, D\}}{\Pr\{A, B, C\}} \]

\[ \Pr\{D \mid A = T, B = T, C\} = \Pr\{D \mid A = T, B = F, C\} = \Pr\{D \mid A = F, B = T, C\} = \Pr\{D \mid A = F, B = F, C\} \]

\[ \Pr\{D \mid A, B, C\} = \Pr\{D \mid C\} \]
\[ \Pr\{A, B, C\} = \Pr\{C|A, B\}\Pr\{A, B\} \]
\[ = \Pr\{C|A, B\}\Pr\{B|A\}\Pr\{A\} \]
\[ = \Pr\{C|B\}\Pr\{B|A\}\Pr\{A\} \]

\[ \Pr\{A|C\} = \frac{\Pr\{A, C\}}{\Pr\{C\}} \sum_{B} \Pr\{A, B, C\} \]
\[ = \frac{B}{\sum_{A, B} \Pr\{A, B, C\}} \]

Causal Independence
Simple Example of Bayesian Networks

Pr\{A_R = T | A_W = T\} > Pr\{A_S = T | A_W = T\} < 

\[ \begin{align*}
\text{Cloudy} & \quad A_C=\text{True or False} \\
\text{Sprinkler} & \quad A_S=\text{True or False} \\
\text{Rain} & \quad A_R=\text{True or False} \\
\text{Wet Grass} & \quad A_W=\text{True or False}
\end{align*} \]
Simple Example of Bayesian Networks

\[
\Pr\{A_C, A_S, A_R, A_W\} \\
= \Pr\{A_W | A_C, A_S, A_R\} \Pr\{A_C, A_S, A_R\} \\
= \Pr\{A_W | A_C, A_S, A_R\} \Pr\{A_R | A_C, A_S\} \Pr\{A_C, A_S\} \\
= \Pr\{A_W | A_C, A_S, A_R\} \Pr\{A_R | A_C, A_S\} \Pr\{A_S | A_C\} \Pr\{A_C\} \\
= \Pr\{A_W | A_S, A_R\} \Pr\{A_R | A_C\} \Pr\{A_S | A_C\} \Pr\{A_C\}
\]
Simple Example of Bayesian Networks

\[ \Pr\{A_R = T | A_W = T\} = \frac{\Pr\{A_R = T, A_W = T\}}{\Pr\{A_W = T\}} \]

\[ \Pr\{A_S = T | A_W = T\} = \frac{\Pr\{A_S = T, A_W = T\}}{\Pr\{A_W = T\}} \]
Simple Example of Bayesian Networks

\[
\Pr\{A_R = T, A_W = T\} = \sum_{A_C = T, F} \sum_{A_S = T, F} \Pr\{A_C, A_S, A_R = T, A_W = T\} = 0.4581
\]

\[
\Pr\{A_S = T, A_W = T\} = \sum_{A_C = T, F} \sum_{A_R = T, F} \Pr\{A_C, A_S = T, A_R, A_W = T\} = 0.2781
\]

\[
\Pr\{A_W = T\} = \sum_{A_C = T, F} \sum_{A_S = T, F} \sum_{A_R = T, F} \Pr\{A_C, A_S, A_R, A_W = T\} = 0.6471
\]
Simple Example of Bayesian Networks

\[ \Pr\{A_R = T|A_W = T\} = \frac{\Pr\{A_R = T, A_W = T\}}{\Pr\{A_W = T\}} = \frac{0.4581}{0.6471} = 0.7079 \]

\[ \Pr\{A_S = T|A_W = T\} = \frac{\Pr\{A_S = T, A_W = T\}}{\Pr\{A_W = T\}} = \frac{0.2781}{0.6471} = 0.4298 \]
Contents

1. Introduction
2. Bayesian Network
3. Belief Propagation
4. Interpretation of Belief Propagation in Information Theory
5. Other Practical Applications
6. Summary
Bayesian Network and Graphical Model

Probability distribution with one random variable is assigned to a graph with one node.

\[ P(X_1) = f_1(X_1) \]

1 Node

Probability distribution with two random variables is assigned to an edge.

\[ P(X_1, X_2) = f_{\{1,2\}}(X_1, X_2) \]

1 2 Edge

\[ P(f_1, f_2, f_3) = f_{\{1,2,3\}}(X_1, X_2, X_3) \]

1 2 Hyper-edge

3

19 July, 2011
Yokohama National University
Bayesian network is expressed in terms of a product of functions and is assigned to chain, tree, cycle or hypergraph representation.

\[
P(X_1, X_2, X_3) = f_{\{1,2\}}(X_1, X_2)f_{\{2,3\}}(X_2, X_3)
\]

\[
P(X_1, X_2, X_3, X_4)
= f_{\{1,2\}}(X_1, X_2)f_{\{1,3\}}(X_1, X_3)f_{\{1,4\}}(X_1, X_4)
\]

\[
P(X_1, X_2, X_3)
= f_{\{1,2\}}(X_1, X_2)f_{\{2,3\}}(X_2, X_3)f_{\{3,1\}}(X_3, X_1)
\]

\[
P(X_1, X_2, X_3, X_4, X_5)
= f_{\{1,2,3\}}(X_1, X_2, X_3)f_{\{3,4,5\}}(X_3, X_4, X_5)
\]
Graphical Representations of Tractable Probabilistic Models

\[ W_{AB}(A, B)W_{BC}(B, C)W_{CD}(C, D)W_{DE}(D, E) \]

\[ = \]

\[ A \quad B \quad \times \quad B \quad C \quad \times \quad C \quad D \quad \times \quad D \quad E \]

\[ W_{AB}(A, B) \quad W_{BC}(B, C) \quad W_{CD}(C, D) \quad W_{DE}(D, E) \]

\[ = \]

\[ A \quad B \quad C \quad D \quad E \]

\[ W_{AB}(A, B) \quad W_{BC}(B, C) \quad W_{CD}(C, D) \quad W_{DE}(D, E) \]
Graphical Representations of Tractable Probabilistic Models

\[ \sum_{A} \sum_{B} \sum_{C} \sum_{D} \sum_{E} \]

\[ = \sum_{A} \sum_{B} \sum_{C} \sum_{D} \sum_{E} \]

19 July, 2011

Yokohama National University
Graphical Representations of Tractable Probabilistic Models

\[ \sum_{A} \sum_{B} \sum_{C} \sum_{D} \sum_{E} \]

\[ = \sum_{A} \sum_{B} \sum_{C} \sum_{D} \sum_{E} \]

\[ = \sum_{B} \sum_{C} \sum_{D} \sum_{E} \left( \sum_{A} \right) \]
Graphical Representations of Tractable Probabilistic Models
Graphical Representations of Tractable Probabilistic Models

\[ \sum \sum \sum \sum \sum \]

\[ \text{A} \quad \text{B} \quad \text{C} \quad \text{D} \quad \text{E} \]

\[ = \sum \sum \sum \sum \sum \]

\[ \text{A} \quad \text{B} \quad \text{X} \quad \text{B} \quad \text{C} \quad \text{D} \quad \text{E} \]

\[ = \sum \sum \sum \sum \sum \left( \sum \right) \]

\[ \text{A} \quad \text{B} \quad \text{B} \quad \text{C} \quad \text{D} \quad \text{E} \]

\[ \text{A} \rightarrow \text{B} \]

\[ = \sum \sum \sum \sum \sum \]

\[ \text{A} \quad \text{B} \quad \text{C} \quad \text{D} \quad \text{E} \]

19 July, 2011
Graphical Representations of Tractable Probabilistic Models
Graphical Representations of Tractable Probabilistic Models

\[
\sum_{B} \sum_{C} \sum_{D} \sum_{E} A \rightarrow B \quad C \quad D \quad E
\]

= \sum_{B} \sum_{C} \sum_{D} \sum_{E} \sum_{A} A \rightarrow B \quad C \quad X \quad C \quad D \quad E

19 July, 2011
Graphical Representations of Tractable Probabilistic Models

\[ \sum \sum \sum \sum \sum \quad \Rightarrow \quad B \quad C \quad D \quad E \]

\[ = \sum \sum \sum \sum \sum \quad \Rightarrow \quad B \quad C \quad X \quad C \quad D \quad E \]

\[ = \sum \sum \sum \sum \left( \sum \sum \right) \quad \Rightarrow \quad B \quad C \quad \left( \sum \sum \right) \quad X \quad C \quad D \quad E \]
Graphical Representations of Tractable Probabilistic Models

\[ \sum_{B} \sum_{C} \sum_{D} \sum_{E} \sum_{A} \rightarrow B \rightarrow C \rightarrow D \rightarrow E \]

\[ = \sum_{B} \sum_{C} \sum_{D} \sum_{E} \sum_{A} \rightarrow B \rightarrow C \rightarrow X \rightarrow C \rightarrow D \rightarrow E \]

\[ = \sum_{C} \sum_{D} \sum_{E} \sum_{B} \left( \sum_{A} \rightarrow B \rightarrow C \right) \rightarrow X \rightarrow C \rightarrow D \rightarrow E \]

\[ \rightarrow B \rightarrow C \]
Graphical Representations of Tractable Probabilistic Models

\[
\sum_{B} \sum_{C} \sum_{D} \sum_{E} A \rightarrow B \rightarrow C \rightarrow D \rightarrow E
\]

\[
= \sum_{B} \sum_{C} \sum_{D} \sum_{E} \sum_{B} A \rightarrow B \rightarrow C \rightarrow \left( \sum_{B} \sum_{C} \sum_{D} \sum_{E} \right) \rightarrow \left( \sum_{B} \sum_{C} \sum_{D} \sum_{E} \right) \rightarrow \sum_{B} A \rightarrow B \rightarrow C \rightarrow \sum_{B} C \rightarrow D \rightarrow \sum_{B} E
\]

\[
= \sum_{B} \sum_{C} \sum_{D} \sum_{E} B \rightarrow C \rightarrow D \rightarrow E
\]
Graphical Representations of Tractable Probabilistic Models

\[ \sum \sum \sum \sum \sum \]

A B C D E

19 July, 2011
Graphical Representations of Tractable Probabilistic Models

\[ \sum \sum \sum \sum \sum \sum_A B C D E \]

\[ = \sum \sum \sum \sum \sum_B C D E \]

19 July, 2011
Graphical Representations of Tractable Probabilistic Models

\[
\sum_{A} \sum_{B} \sum_{C} \sum_{D} \sum_{E} = \sum_{B} \sum_{C} \sum_{D} \sum_{E} \sum_{A}
\]

\[
\sum_{A} \sum_{B} \sum_{C} \sum_{D} \sum_{E} = \sum_{B} \sum_{C} \sum_{D} \sum_{E} \sum_{A}
\]

19 July, 2011
Graphical Representations of Tractable Probabilistic Models

\[
\sum \sum \sum \sum \sum_{\text{A} \text{ B} \text{ C} \text{ D} \text{ E}}
\]

\[
= \sum \sum \sum \sum \sum_{\text{B} \text{ C} \text{ D} \text{ E}}
\]

\[
= \sum \sum \sum \sum_{\text{C} \text{ D} \text{ E}}
\]

\[
= \sum \sum_{\text{D} \text{ E}}
\]

19 July, 2011

Yokohama National University
Graphical Representations of Tractable Probabilistic Models

\[
\sum_{A} \sum_{B} \sum_{C} \sum_{D} \sum_{E} A \sum_{B} \sum_{C} \sum_{D} \sum_{E} B \sum_{C} \sum_{D} \sum_{E} C \sum_{D} \sum_{E} D \sum_{E} E
\]
Graphical Representations of Tractable Probabilistic Models

\[ W_{ABC}(A,B,C)W_{CDE}(C,D,E)W_{DFG}(D,F,G)W_{EHI}(E,H,I) \]

\[ = W_{ABC}(A,B,C) \times W_{CDE}(C,D,E) \times W_{DFG}(D,F,G) \times W_{EHI}(E,H,I) \]

\[ = W_{ABC}(A,B,C) \times W_{CDE}(C,D,E) \times W_{DFG}(D,F,G) \times W_{EHI}(E,H,I) \]

19 July, 2011
Graphical Representations of Tractable Probabilistic Models

\[
\Pr\{C, D, F\} = \sum_{A} \sum_{B} \sum_{F} \sum_{G} \sum_{H} \sum_{I} \quad (1)
\]

\[
= \left( \sum_{A} \sum_{B} \right) \left( \sum_{F} \sum_{G} \right) \left( \sum_{H} \sum_{I} \right)
\]

19 July, 2011
Graphical Representations of Tractable Probabilistic Models

\[ \Pr\{C\} = \sum_{A} \sum_{B} \sum_{D} \sum_{E} \sum_{F} \sum_{G} \sum_{H} \sum_{I} \]

\[
\left( \sum_{A} \sum_{B} \right) \times \left( \sum_{D} \sum_{E} \right) \times \left( \sum_{F} \sum_{G} \right) \times \left( \sum_{H} \sum_{I} \right)
\]

19 July, 2011
Graphical Representations of Tractable Probabilistic Models

\[ \Pr\{C, D, F\} = \] 

\[ \Pr\{C\} = \sum_{D} \sum_{E} \Pr\{C, D, E\} \]

\[ \Pr\{C\} = \]
Belief Propagation on Hypergraph Representations in terms of Cactus Tree

Update Flow of Messages in computing the marginal probability $\Pr\{C\}$

19 July, 2011
Belief propagation cannot give us exact computations in Bayesian networks on cycle graphs.

Applications of belief propagation to Bayesian networks on cycle graphs provide us many powerful approximate computational models and practical algorithms for probabilistic information processing.
Simple Example of Bayesian Networks

Visit to Asia: \( X_1 \)
- Visit: \( X_1 = 1 \)
- No Visit: \( X_1 = 0 \)

Smoking: \( X_2 \)
- Smoker: \( X_2 = 1 \)
- Non Smoker: \( X_2 = 0 \)

Tuberculosis: \( X_3 \)
- Present: \( X_3 = 1 \)
- Absent: \( X_3 = 0 \)

Lung Cancer: \( X_4 \)
- Present: \( X_4 = 1 \)
- Absent: \( X_4 = 0 \)

Bronchitis: \( X_5 \)
- Present: \( X_5 = 1 \)
- Absent: \( X_5 = 0 \)

X-ray Result: \( X_7 \)
- Abnormal: \( X_7 = 1 \)
- Normal: \( X_7 = 0 \)

Dyspnea: \( X_8 \)
- Present: \( X_8 = 1 \)
- Absent: \( X_8 = 0 \)

Tuberculosis or Cancer: \( X_6 \)
- True: \( X_6 = 1 \)
- False: \( X_6 = 0 \)

\[
\Pr\{X_1 = x_1, X_2 = x_2, \ldots, X_8 = x_8\} = \Pr\{X_8 \mid X_5, X_6\}\Pr\{X_7 \mid X_6\}\Pr\{X_6 \mid X_3, X_4\} \times \Pr\{X_5 \mid X_2\}\Pr\{X_4 \mid X_2\}\Pr\{X_3 \mid X_1\} \times \Pr\{X_1\}\Pr\{X_2\}
\]

Undirected Hypergraph

Directed Graph

Undirected Hypergraph

Directed Graph

\[
\Pr\{X_1 = x_1, X_2 = x_2, \ldots, X_8 = x_8\} = W_{568}(x_5, x_6, x_8)W_{346}(x_3, x_4, x_5)W_{67}(x_6, x_7) \times W_{25}(x_2, x_5)W_{24}(x_2, x_4)W_{13}(x_1, x_3)
\]
Marginal Probability Distributions

\[ P_i(X_i) \equiv \sum_{X \setminus \{X_i\}} \text{Pr}\{X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8\} \]

\[ P_{ij}(X_i, X_j) \equiv \sum_{X \setminus \{X_i, X_j\}} \text{Pr}\{X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8\} \]

\[ P_{ijk}(X_i, X_j, X_k) \equiv \sum_{X \setminus \{X_i, X_j, X_k\}} \text{Pr}\{X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8\} \]
Approximate Representations of Marginal Probability Distributions in terms of Messages

\[ P_{346}(X_3, X_4, X_6) \cong \sum \sum \sum \left( W_{346}(X_3, X_4, X_6) M_{3\leftarrow 13}(X_3) M_{4\leftarrow 24}(X_4) \right) \times M_{6\leftarrow 67}(X_6) M_{6\leftarrow 568}(X_6) \]
Approximate Representations of Marginal Probability Distributions in terms of Messages

\[ P_6(x_6) \approx \frac{M_{6 \leftarrow 346}(x_6)M_{6 \leftarrow 568}(x_6)M_{6 \leftarrow 67}(x_6)}{\sum_{x_6} M_{6 \leftarrow 346}(x_6)M_{6 \leftarrow 568}(x_6)M_{6 \leftarrow 67}(x_6)} \]
Basic Strategies of Belief Propagations in Probabilistic Model with Graphical Representation including Cycles

$$P_6(x_6) = \sum_{x_3} \sum_{x_4} P_{346}(x_3, x_4, x_6)$$

Approximate Expressions of Marginal Probabilities

$$P_6(x_6) \approx \frac{1}{Z_6} M_{6\leftarrow 346}(x_6) \times M_{6\leftarrow 568}(x_6) \times M_{6\leftarrow 67}(x_6)$$

$$P_{346}(x_3, x_4, x_6) \approx \frac{1}{Z_{346}} W_{346}(x_3, x_4, x_6) \times M_{3\leftarrow 13}(x_3) M_{4\leftarrow 24}(x_4) \times M_{6\leftarrow 568}(x_6) M_{6\leftarrow 67}(x_6)$$

19 July, 2011
Simultaneous Fixed Point Equations for Belief Propagations in Hypergraph Representations

\[
M_{6 \leftarrow 346}(x_6) = \frac{\sum_{x_3} \sum_{x_4} \sum_{x_6} (W_{346}(x_3, x_4, x_6)M_{3 \leftarrow 13}(x_3)M_{4 \leftarrow 24}(x_4))}{\sum_{x_3} \sum_{x_4} \sum_{x_6} (W_{346}(x_3, x_4, x_6)M_{3 \leftarrow 13}(x_3)M_{4 \leftarrow 24}(x_4))}
\]

Belief Propagation Algorithm
Fixed Point Equation and Iterative Method

**Fixed Point Equation**

\[
\vec{M}^* = \Phi(\vec{M}^*)
\]

**Iterative Method**

\[
\begin{align*}
\vec{M}_1 & \leftarrow \Phi(\vec{M}_0) \\
\vec{M}_2 & \leftarrow \Phi(\vec{M}_1) \\
\vec{M}_3 & \leftarrow \Phi(\vec{M}_2) \\
\vdots &
\end{align*}
\]
Belief propagation can be applied to Bayesian networks also on hypergraphs as powerful approximate algorithms.
Numerical Experiments

\[ P_8(1) = 0.5607 \quad P_8(0) = 0.4393 \]
\[ P_{58}(1,1) = 0.4736 \quad P_{58}(1,0) = 0.0764 \]
\[ P_{58}(0,1) = 0.0871 \quad P_{58}(0,0) = 0.3629 \]

\[ P_8(x_8) \equiv \sum_{x \setminus \{x_8\}} P(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8) \]
\[ P_{58}(x_5, x_8) \equiv \sum_{x \setminus \{x_5, x_8\}} P(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8) \]
Numerical Experiments

**Belief Propagation**

\[
\Pr\{X_{\text{Bronchitis}} = \text{Present} \mid X_{\text{Dyspnea}} = \text{Present}\} = \frac{\Pr\{X_{\text{Bronchitis}} = \text{Present}, X_{\text{Dyspnea}} = \text{Present}\}}{\Pr\{X_{\text{Dyspnea}} = \text{Present}\}} = \frac{0.3629}{0.4393} = 0.8261
\]

Visit to Asia: \(X_1\)
- Visit: \(X_1 = 1\)
- No Visit: \(X_1 = 0\)

Smoking: \(X_2\)
- Smoker: \(X_2 = 1\)
- Non Smoker: \(X_2 = 0\)

Tuberculosis: \(X_3\)
- Present: \(X_3 = 1\)
- Absent: \(X_3 = 0\)

Lung Cancer: \(X_4\)
- Present: \(X_4 = 1\)
- Absent: \(X_4 = 0\)

Bronchitis: \(X_5\)
- Present: \(X_5 = 1\)
- Absent: \(X_5 = 0\)

Tuberculosis or Cancer: \(X_6\)
- True: \(X_6 = 1\)
- False: \(X_6 = 0\)

X-ray Result: \(X_7\)
- Abnormal: \(X_7 = 1\)
- Normal: \(X_7 = 0\)

Dyspnea: \(X_8\)
- Present: \(X_8 = 1\)
- Absent: \(X_8 = 0\)
Linear Response Theory

\[ P_{ij}(m, n) - P_i(m)P_j(n) \]

\[ = \lim_{h_i \to 0} \left( \frac{\Delta^{(i)} P_j(n)}{h_i} \right) \]

\[ x = (x_1, x_2, \ldots, x_{|V|}) \]

\[ \Delta^{(i)} P_j(n) = \tilde{P}_j(n) - P_j(n) = \sum_{x, n, x_j} \delta_{n, x_j} \tilde{P}(x) - \sum_{x} \delta_{n, x_j} P(x) \]

\[ \text{Deviation of Average at Node } j \text{ with respect to External Field at Node } i \]

\[ \tilde{P}(x) \equiv \frac{1}{Z} P(x) \exp \left(h_i \delta_{x_i, m} \right) \]

Numerical Experiments

\[
\Pr\{X_{\text{Tuberculosis}} = \text{Present} \mid X_{\text{Dyspnea}} = \text{Present}\} = \frac{\Pr\{X_{\text{Tuberculosis}} = \text{Present}, X_{\text{Dyspnea}} = \text{Present}\}}{\Pr\{X_{\text{Dyspnea}} = \text{Present}\}} = \frac{0.0082}{0.4393} = 0.0187
\]

Contents

1. Introduction
2. Bayesian Network
3. Belief Propagation
4. Interpretation of Belief Propagation in Information Theory
5. Other Practical Applications
6. Summary
Factor Graph Representations of Bayesian Networks and Belief Propagations
We consider hypergraphs which satisfy

\[ \forall \gamma \in E, \forall \gamma' \in E : \gamma \cap \gamma' \in V \text{ or } \gamma \cap \gamma' = \emptyset \]

Cactus Tree

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

\[ E: \text{Set of all the hyperedges} \]

Hypergraph
Interpretation of Belief Propagation based on Information Theory

**Kullback-Leibler Divergence**

\[ D[Q|P] = \sum_x Q(x) \ln \left( \frac{Q(x)}{P(x)} \right) \geq 0 \]

\[ Q(x) = P(x) \Rightarrow D[Q|P] = 0 \]

\[ D[Q|P] = \sum_x Q(x) \sum_{\gamma \in E} \ln W_\gamma(x_\gamma) + \sum_x Q(x) \ln Q(x) + \ln Z \]

\[ = F[Q] + \ln Z \quad \text{Free Energy} \]

\[ Z \equiv \sum_x \prod_{\gamma \in E} W_\gamma(x_\gamma) \]

\[ \min_Q \left\{ F[Q] \left| \sum_x Q(x) = 1 \right. \right\} = F[P] = -\ln Z \]

\[ \therefore Q(x) \geq 0, \quad \sum_x Q(x) = 1 \]

\[ P(x_1, x_2, \ldots, x_{|V|}) = \frac{1}{Z} \prod_{\gamma \in E} W_\gamma(x_\gamma) \]
Interpretation of Belief Propagation based on Information Theory

\[ D[Q|P] = \sum_x Q(x) \ln \left( \frac{Q(x)}{P(x)} \right) \geq 0 \]

Free Energy

\[ P(x) = \frac{1}{Z} \prod_{\gamma \in E} W_\gamma (x_\gamma) \]

KL Divergence

\[ D[Q|P] = F[Q] + \ln(Z) \]

\[ F[Q] = \sum_x Q(x) \sum_{\gamma \in E} \ln W_\gamma (x_\gamma) + \sum_x Q(x) \ln Q(x) \]

\[ Q_\gamma (x_\gamma) = \sum_{x \setminus x_\gamma} Q(x) \]

19 July, 2011
Interpretation of Belief Propagation based on Information Theory

KL Divergence

\[ D[Q|P] = F[Q] + \ln(Z) \]

Free Energy

\[ F[Q] = \sum_{\gamma \in E} \sum_{x_\gamma} Q_\gamma(x_\gamma) \ln W_\gamma(x_\gamma) + \sum_{x} Q(x) \ln Q(x) \]

\[ \approx \sum_{\gamma \in E} \sum_{x_\gamma} Q_\gamma(x_\gamma) \ln W_\gamma(x_\gamma) \]

\[ + \sum_{i \in V} \sum_{\xi} Q_i(\xi) \ln Q_i(\xi) \]

\[ + \sum_{\gamma \in E} \left( \sum_{x_\gamma} Q_\gamma(x_\gamma) \ln Q_\gamma(x_\gamma) - \sum_{i \in \gamma} \sum_{x_i} Q_i(x_i) \ln Q_i(x_i) \right) \]

Free Energy

\[ P(x) = \frac{1}{Z} \prod_{\gamma \in E} W_\gamma(x_\gamma) \]

\[ Q_i(x_i) = \sum_{x \setminus x_i} Q(x) \]

\[ Q_\gamma(x_\gamma) = \sum_{x \setminus x_\gamma} Q(x) \]

19 July, 2011

Yokohama National University
Interpretation of Belief Propagation based on Information Theory

\[
D[Q|P] \approx F_{\text{Bethe}}(\{Q_i, Q_{ij}\}) + \ln Z
\]

\[
F_{\text{Bethe}}(\{Q_i, Q_{\gamma}\}) \equiv \sum_{\gamma \in E} \sum_{x_{\gamma}} Q_{\gamma}(x_{\gamma}) \ln W_{\gamma}(x_{\gamma}) + \sum_{i \in V} \sum_{\xi} Q_i(\xi) \ln Q_i(\xi)
\]

\[
+ \sum_{\gamma \in E} \left( \sum_{x_{\gamma}} Q_{\gamma}(x_{\gamma}) \ln Q_{\gamma}(x_{\gamma}) - \sum_{i \in \gamma} \sum_{x_i} Q_i(x_i) \ln Q_i(x_i) \right)
\]

\[
\arg \min_Q D[Q|P] \Rightarrow \arg \min_{\{Q_i, Q_{\gamma}\}} F_{\text{Bethe}}(\{Q_i, Q_{\gamma}\})
\]

\[
\sum_{x_i} Q_i(x_i) = \sum_{x_{\gamma}} Q_{\gamma}(x_{\gamma}) = 1
\]

\[
Q_i(x_i) = \sum_{x_{\gamma} \setminus x_i} Q_{\gamma}(x_{\gamma})
\]
Interpretation of Belief Propagation based on Information Theory

\[
\arg \min_{\{Q_i, Q_\gamma\}} \left\{ F_{\text{Bethe}} \left[ \{Q_i, Q_\gamma\} \right] \middle| \begin{array}{l}
Q_i(x_i) = \sum_{x_\gamma \setminus x_i} Q_\gamma(x_\gamma), \\
\sum_{Q_i(x_i)} = \sum_{Q_\gamma(x_\gamma)} = 1
\end{array} \right\}
\]

Lagrange Multipliers to ensure the constraints

\[
L_{\text{Bethe}} \left[ \{Q_i, Q_\gamma\} \right] \equiv F_{\text{Bethe}} \left[ \{Q_i, Q_\gamma\} \right]
\]

\[
- \sum_{i \in V} \sum_{\gamma \in \partial i} \sum_{x_\gamma \setminus x_i} \lambda_{i, \gamma}(\xi) \left( Q_i(x_i) - \sum_{\gamma} Q_\gamma(x_\gamma) \right)
\]

\[
- \sum_{i \in V} \nu_i \left( \sum_{x_i} Q_i(x_i) - 1 \right) - \sum_{\gamma \in E} \nu_\gamma \left( \sum_{x_\gamma} Q_\gamma(x_\gamma) - 1 \right)
\]
Interpretation of Belief Propagation based on Information Theory

\[
L_{\text{Bethe}}[\{Q_i, Q_{ij}\}] = F_{\text{Bethe}}[\{Q_i, Q_{ij}\}] - \sum_{i \in V} \sum_{\gamma \in \partial i} \lambda_{i,\gamma}(x_i) \left( Q_i(x_i) - \sum_{x_{\gamma} \setminus x_i} Q_{\gamma}(x_{\gamma}) \right)

- \sum_{i \in V} \nu_i \left( \sum_{x_i} Q_i(x_i) - 1 \right) - \sum_{\gamma \in E} \nu_{\gamma} \left( \sum_{x_{\gamma}} Q_{\gamma}(x_{\gamma}) - 1 \right)

= \sum_{\gamma \in E} \sum_{x_{\gamma}} Q_{\gamma}(x_{\gamma}) \ln W_{\gamma}(x_{\gamma}) + \sum_{i \in V} \sum_{x_i} Q_i(x_i) \ln Q_i(x_i)

+ \sum_{\gamma \in E} \left( \sum_{x_{\gamma}} Q_{\gamma}(x_{\gamma}) \ln Q_{\gamma}(x_{\gamma}) - \sum_{i \in V} \sum_{x_i} Q_i(\bar{x}_i) \ln Q_i(\bar{x}_i) \right)

- \sum_{i \in V} \sum_{\gamma \in \partial i} \lambda_{i,\gamma}(x_i) \left( Q_i(x_i) - \sum_{x_{\gamma} \setminus x_i} Q_{\gamma}(x_{\gamma}) \right) - \sum_{i \in V} \nu_i \left( \sum_{x_i} Q_i(x_i) - 1 \right) - \sum_{\gamma \in E} \nu_{ij} \left( \sum_{x_{\gamma}} Q_{\gamma}(x_{\gamma}) - 1 \right)

\]

\textbf{Extremum Condition}

\[
\frac{\partial}{\partial Q_i(x_i)} L_{\text{Bethe}}[\{Q_i, Q_{ij}\}] = 0
\]

\[
\frac{\partial}{\partial Q_{\gamma}(x_{\gamma})} L_{\text{Bethe}}[\{Q_i, Q_{ij}\}] = 0
\]
Interpretation of Belief Propagation based on Information Theory

\[
\frac{\partial}{\partial Q_i(x_i)} L_{\text{Bethe}}[\{Q_i, Q_{ij}\}] = 0
\]

\[
\frac{\partial}{\partial Q_{ij}(x_i, x_j)} L_{\text{Bethe}}[\{Q_i, Q_{ij}\}] = 0
\]

Extremum Condition

\[
Q_i(x_i) = \frac{1}{Z_i} \prod_{\gamma \in \partial i} M_{\gamma \rightarrow i}(x_i)
\]

\[
Q_{\gamma}(x_{\gamma}) = \frac{1}{Z_{\gamma}} W_{\gamma}(x_{\gamma}) \left( \prod_{j \in \gamma} \prod_{\gamma' \in \partial_j \setminus \gamma} M_{\gamma' \rightarrow j}(x_j) \right)
\]

\[
Q_i(x_i) = \sum_{x_{\gamma \setminus x_i}} Q_{\gamma}(x_{\gamma})
\]

\[
M_{\gamma \rightarrow i}(x_i) = \frac{Z_i}{Z_{\gamma}} W_{\gamma}(x_{\gamma}) \left( \prod_{j \in \gamma \setminus i} \prod_{\gamma' \in \partial_j \setminus \gamma} M_{\gamma' \rightarrow j}(x_j) \right)
\]

19 July, 2011
Contents

1. Introduction
2. Bayesian Network
3. Belief Propagation
4. Interpretation of Belief Propagation in Information Theory
5. Other Practical Applications
6. Summary
Image Restoration by Probabilistic Model

Assumption 1: The degraded image is randomly generated from the original image by according to the degradation process.
Assumption 2: The original image is randomly generated by according to the prior probability.

\[
\Pr\{\text{Original Image} \mid \text{Degraded Image}\} = \frac{\Pr\{\text{Degraded Image} \mid \text{Original Image}\} \Pr\{\text{Original Image}\}}{\Pr\{\text{Degraded Image}\}}
\]

Bayes Formula

19 July, 2011
Prior Probability of Image Processing

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

Digital Images
\( (f_i = 0, 1, \ldots, q-1) \)

\[ Z_{PR}(\alpha) \equiv \sum_{z_1 = 0}^{q-1} \sum_{z_1 = 0}^{q-1} \cdots \sum_{z_{|V|} = 0}^{q-1} \exp\left(-\frac{1}{2} \alpha \sum_{\{i,j\} \in E} (z_i - z_j)^2\right) \]

**Sampling of Markov Chain Monte Carlo Method**

**\( E \): Set of all the nearest-neighbour pairs of pixels**

**\( V \): Set of all the pixels**

\( \alpha = 1 \) **Paramagnetic**

\( \alpha = 2 \)

\( \alpha = 4 \) **Ferromagnetic**

Near the critical point \( \alpha = 1.76 \ldots \)
Degradation Process is assumed to be the additive white Gaussian noise.

\[
\Pr\{G = g | F = f, \sigma\} = \prod_{i \in V} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (f_i - g_i)^2\right)
\]

V: Set of all the pixels

Degraded image is adding a white Gaussian noise to the original image.

\[
f_i \sim N(0, \sigma^2)
\]
Bayesian Network and Posterior Probability of Image Processing

Posterior Probability

\[
\Pr\{ F = f | G = g \} = \frac{\Pr\{ G = g | F = f \} \Pr\{ F = f \}}{\Pr\{ G = g \}}
\]

Probabilistic Model expressed in terms of Graphical Representations with Cycles

\[
\Pr\{ F = f | G = g \} = \frac{1}{Z_{\text{Posterior}}} \prod_{\{i, j\} \in E} \Phi_{i, j}(f_i, f_j)
\]

\[
\Phi_{i, j}(f_i, f_j) = \exp \left( -\frac{1}{8\sigma^2} (f_i - g_i)^2 - \frac{1}{8\sigma^2} (f_j - g_j)^2 - \frac{1}{2} \alpha (f_i - f_j)^2 \right)
\]

\[
\hat{f}_i \leftarrow \Pr\{ F_i = f_i | G = g \} = \sum_{f \neq f_i} \Pr\{ F = f | G = g \}
\]

\( V: \) Set of all the pixels

\( E: \) Set of all the nearest neighbour pairs of pixels
Belief Propagation Algorithm for Image Processing

Step 1: Solve the simultaneous fixed point equations for messages by using iterative method

\[ M_{i \rightarrow j}(f_j) = \frac{\sum_{z_i=0}^{q-1} \Phi_{\{i,j\}}(z_i, f_j) \prod_{k \in \partial i} M_{k \rightarrow i}(z_i)}{\sum_{z_i=0}^{q-1} \sum_{z_j=0}^{q-1} \Phi_{\{i,j\}}(z_i, z_j) \prod_{k \in \partial i} M_{k \rightarrow i}(z_i)} \]

(\(\forall j \in \partial i, \forall i \in V, f_i \in \{0,1,2,\cdots,q-1\}\))

Step 2: Substitute the messages to approximate expressions of marginal probabilities for each pixel and compute estimated image so as to maximize the approximate marginal probability at each pixel.

\[ \Pr\{F_i = f_i | G = g\} \leftarrow \frac{\prod_{k \in \partial i} M_{k \rightarrow i}(f_i)}{\sum_{z_i=0}^{q-1} \prod_{k \in \partial i} M_{k \rightarrow i}(f_i)} \quad (i \in V, f_i \in \{0,1,2,\cdots,q-1\}) \]

\[ \hat{f}_i = \arg \max_{z_i=0,1,\cdots,q-1} P_i(z_i) \]
Image Restorations by Exact Solution and Loopy Belief Propagation of Gaussian Graphical Model

\[
(\alpha(t + 1), \sigma(t + 1)) \leftarrow \arg \max_{(\alpha, \sigma)} Q(\alpha, \sigma | \alpha(t), \sigma(t), g).
\]

\[
\Pr\{F = f | G = g\} = \frac{1}{Z_{\text{Posterior}}} \prod_{\{i, j\} \in E} \Phi_{\{i, j\}}(f_i, f_j)
\]

\[
\hat{\sigma}_{\text{Exact}} = 37.624
\]

\[
\hat{\alpha}_{\text{Exact}} = 0.000713
\]

\[
\Phi_{\{i, j\}}(f_i, f_j) = \exp\left(-\frac{1}{8\sigma^2}(f_i - g_i)^2 - \frac{1}{8\sigma^2}(f_j - g_j)^2 - \frac{1}{2}\alpha(f_i - f_j)^2\right)
\]

\[
\hat{\sigma}_{\text{LBP}} = 36.335
\]

\[
\hat{\alpha}_{\text{LBP}} = 0.000600
\]

\[
f_i \in (-\infty, +\infty)
\]

Exact

\[
\hat{f} = h(\hat{\alpha}, \hat{\sigma}, g)
\]

Loopy Belief Propagation
Error Correcting Codes and Belief Propagation

\[ X_7 = X_1 + X_2 + X_3 \pmod{2} \]
\[ X_8 = X_2 + X_3 + X_5 \pmod{2} \]
\[ X_9 = X_3 + X_4 + X_6 \pmod{2} \]

\[
\begin{pmatrix}
  x_1 \\
  x_2 \\
  x_3 \\
  x_4 \\
  x_5 \\
  x_6
\end{pmatrix}
= 
\begin{pmatrix}
  1 \\
  1 \\
  0 \\
  0 \\
  1 \\
  1
\end{pmatrix}
Error Correcting Codes and Belief Propagation

\[ X_7 = X_1 + X_2 + X_3 \pmod{2} \]
\[ X_8 = X_2 + X_3 + X_5 \pmod{2} \]
\[ X_9 = X_3 + X_4 + X_6 \pmod{2} \]

\[ \begin{pmatrix}
  x_1 \\
  x_2 \\
  x_3 \\
  x_4 \\
  x_5 \\
  x_6 \\
\end{pmatrix} = \begin{pmatrix}
  1 \\
  1 \\
  0 \\
  0 \\
  1 \\
  1
\end{pmatrix} \]

\[ X_7 = 1 + 1 + 0 \pmod{2} = 0 \]
\[ X_8 = 1 + 0 + 1 \pmod{2} = 0 \]
\[ X_9 = 0 + 0 + 1 \pmod{2} = 1 \]
Error Correcting Codes and Belief Propagation

\[ X_7 = X_1 + X_2 + X_3 \pmod 2 \]
\[ X_8 = X_2 + X_3 + X_5 \pmod 2 \]
\[ X_9 = X_3 + X_4 + X_6 \pmod 2 \]

\[
\begin{bmatrix}
  x_1 \\
  x_2 \\
  x_3 \\
  x_4 \\
  x_5 \\
  x_6 \\
\end{bmatrix} =
\begin{bmatrix}
  1 \\
  1 \\
  0 \\
  0 \\
  1 \\
  1 \\
\end{bmatrix}
\]

\[
\begin{bmatrix}
  x_1 \\
  x_2 \\
  x_3 \\
  x_4 \\
  x_5 \\
  x_6 \\
  x_7 \\
  x_8 \\
  x_9 \\
\end{bmatrix} =
\begin{bmatrix}
  1 \\
  1 \\
  0 \\
  0 \\
  1 \\
  1 \\
  0 \\
  0 \\
  1 \\
\end{bmatrix}
\]

\[ X_7 = 1 + 1 + 0 \pmod 2 = 0 \]
\[ X_8 = 1 + 0 + 1 \pmod 2 = 0 \]
\[ X_9 = 0 + 0 + 1 \pmod 2 = 1 \]
Error Correcting Codes and Belief Propagation

\[
\begin{align*}
X_7 &= X_1 + X_2 + X_3 \pmod{2} \\
X_8 &= X_2 + X_3 + X_5 \pmod{2} \\
X_9 &= X_3 + X_4 + X_6 \pmod{2}
\end{align*}
\]

\[
\begin{bmatrix}
    x_1 \\
x_2 \\
x_3 \\
x_4 \\
x_5 \\
x_6
\end{bmatrix}
= 
\begin{bmatrix}
    1 \\
    1 \\
    0 \\
    0 \\
    1 \\
    1
\end{bmatrix}
\]

\[
\begin{align*}
X_7 &= 1 + 1 + 0 \pmod{2} = 0 \\
X_8 &= 1 + 0 + 1 \pmod{2} = 0 \\
X_9 &= 0 + 0 + 1 \pmod{2} = 1
\end{align*}
\]

Received Word

\[
\begin{bmatrix}
y_1 \\
y_2 \\
y_3 \\
y_4 \\
y_5 \\
y_6 \\
y_7 \\
y_8 \\
y_9
\end{bmatrix}
= 
\begin{bmatrix}
    1 \\
    1 \\
    1 \\
    0 \\
    0 \\
    1 \\
    0 \\
    1 \\
    1
\end{bmatrix}
\]

Binary Symmetric Channel

\[
\begin{align*}
l &\xrightarrow{1-p} l \\
l &\xrightarrow{p} 0 \\
0 &\xrightarrow{p} l \\
0 &\xrightarrow{1-p} 0
\end{align*}
\]
Error Correcting Codes and Belief Propagation

\[ X_7 = X_1 + X_2 + X_3 \pmod{2} \]
\[ X_8 = X_2 + X_3 + X_5 \pmod{2} \]
\[ X_9 = X_3 + X_4 + X_6 \pmod{2} \]
Error Correcting Codes and Belief Propagation

\[ X_7 = X_1 + X_2 + X_3 \pmod{2} \]
\[ X_8 = X_2 + X_3 + X_5 \pmod{2} \]
\[ X_9 = X_3 + X_4 + X_6 \pmod{2} \]
Error Correcting Codes and Belief Propagation

\[ X_7 = X_1 + X_2 + X_3 \pmod{2} \]
\[ X_8 = X_2 + X_3 + X_5 \pmod{2} \]
\[ X_9 = X_3 + X_4 + X_6 \pmod{2} \]
Error Correcting Codes and Belief Propagation

\[ X_7 = X_1 + X_2 + X_3 \pmod{2} \]
\[ X_8 = X_2 + X_3 + X_5 \pmod{2} \]
\[ X_9 = X_3 + X_4 + X_6 \pmod{2} \]
Satisfactory Problem (3-SAT)

\[
(X_1 \lor (\neg X_3) \lor (\neg X_4)) \land (X_3 \lor (\neg X_5) \lor X_6) \land (X_2 \lor X_4 \lor X_6)
\land (X_2 \lor X_7 \lor (\neg X_8)) \land ((\neg X_5) \lor X_8 \lor X_9)
\]

\[
\text{Pr}\{X_1 = x_1, \ldots, X_6 = x_6\} = \frac{1}{Z} \prod_{i=1}^{3} f_{i,3,4}(x_1, x_3, x_4) f_{3,5,6}(x_3, x_5, x_6) f_{2,4,6}(x_2, x_4, x_6)
\times f_{2,7,8}(x_2, x_7, x_8) f_{5,8,9}(x_5, x_8, x_9)
\]
Contents

1. Introduction
2. Bayesian Network
3. Belief Propagation
4. Interpretation of Belief Propagation in Information Theory
5. Other Practical Applications
6. Summary
Belief Propagation and Advanced Mean Field Method in Statistical Physics

In graphical models with tree graphical structures, Bethe approximation is equivalent to Transfer Matrix Method in Statistical Physics and give us exact results for computations of statistical quantities.

In Graphical Models with Cycles, Belief Propagation is equivalent to Bethe approximation or Cluster Variation Method.
Summary

- Bayesian Network for Probabilistic Inference
- Belief Propagation for Bayesian Networks
- Some Practical Applications