

一般化確率伝播というよりは、一般化因数分解

鈴木 譲

2010年11月30日

概要

- 1 確率的推論
- 2 Junction Tree
- 3 領域グラフ
- 4 一般化因数分解の応用

確率的推論

確率空間 (Ω, \mathcal{F}, P)

確率論の公理

- $P(\{\}) = 0, P(\Omega) = 1$
- $0 \leq P(A) \leq 1, A \in \mathcal{F}$
- $A \cap B = \{\} \implies P(A \cup B) = P(A) + P(B)$

$\{X_i\}_{i=1}^N: X_i(\Omega) < \infty$ なる確率変数 $X_i: \Omega \rightarrow \mathbb{R}$ の列
 $P_{X_1 \dots X_N}(x_1, \dots, x_N) := P(X_1 = x_1, \dots, X_N = x_N),$
 $x_1 \in X_1(\Omega), \dots, x_N \in X_N(\Omega)$

確率的推論: N とともに指数的な計算が必要

$$P_{X_i}(x_i) := \sum_{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_N} P_{X_1 \dots X_N}(x_1, \dots, x_N)$$

確率分布の因子グラフによる表現

\mathcal{M} : 有限集合 (因子頂点の集合)

$\mathcal{N} := \{1, \dots, N\}$ (変数頂点の集合)

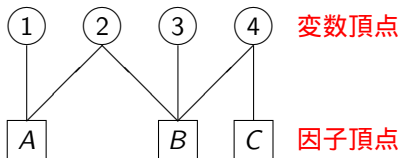
$\mathcal{N}(a) \subseteq \mathcal{N}$: (因子頂点 $a \in \mathcal{M}$ に隣接する変数頂点の集合)

$$X_a(\Omega) := \prod_{j \in \mathcal{N}(a)} X_j(\Omega), \quad f_a : X_a(\Omega) \rightarrow \mathbb{R}_{\geq 0}, \quad a \in \mathcal{M}$$

$$P_{X_1 \dots X_N}(x_1, \dots, x_N) = \frac{1}{Z} \prod_{a \in \mathcal{M}} f_a(\mathbf{x}_a), \quad \mathbf{x}_i \in X_i(\Omega), \quad \mathbf{x}_a \in X_a(\Omega)$$

例: $\mathcal{N} = \{1, 2, 3, 4\}$, $\mathcal{M} = \{A, B, C\}$

$$P_{X_1, X_2, X_3, X_4}(x_1, x_2, x_3, x_4) = \frac{1}{Z} f_A(x_1, x_2) f_B(x_2, x_3, x_4) f_C(x_4)$$



確率伝播

$\mathcal{M}(i) \subseteq \mathcal{M}$: $i \in \mathcal{N}$ に隣接する因子頂点の集合

$i \in \mathcal{N}$, $a \in \mathcal{M}$

$n_{i \rightarrow a}$, $m_{a \rightarrow i}$: $X_i(\Omega) \rightarrow \mathbb{R}_{\geq 0}$

$$n_{i \rightarrow a}(x_i) := 1, \quad m_{a \rightarrow i}(x_i) := 1, \quad x_i \in X_i(\Omega)$$

$$n_{i \rightarrow a}(x_i) := \prod_{c \in \mathcal{M}(i) \setminus \{a\}} m_{c \rightarrow i}(x_i)$$

$$m_{a \rightarrow i}(x_i) := \sum_{\mathbf{x}_{a,i} \in X_{a,i}(\Omega)} f_a(\mathbf{x}_{a,i}, x_i) \prod_{j \in \mathcal{N}(a) \setminus \{i\}} n_{j \rightarrow a}(x_j)$$

$$X_{a,i}(\Omega) := \prod_{j \in \mathcal{N}(a) \setminus \{i\}} X_j(\Omega)$$

$$b_i(x_i) := \frac{1}{Z_i} \prod_{a \in \mathcal{N}(i)} m_{a \rightarrow i}(x_i)$$

因子グラフがループをもつとき、もたないとき

ループをもたないとき

- N に比例する時間で確率伝播が収束
- $\{b_i\}_{i \in \mathcal{N}} = \{P_{X_i}\}_{i \in \mathcal{N}}$

ループをもつとき

- 一般には、NP 困難
- 一般には、 $\{b_i\}_{i \in \mathcal{N}} \neq \{P_{X_i}\}_{i \in \mathcal{N}}$

確率的推論は NP 困難

確率的推論の決定問題版 PRD

$P_{X_i}(x_i) > 0$ か否か

3SAT から PRD へ多項式時間還元

- リテラル $L := U \cup \bar{U}$, $U = \{u_1, \dots, u_N\}$, $\bar{U} = \{\bar{u}_1, \dots, \bar{u}_N\}$,
節 $c_j \subseteq L$, $|L| = 3$, $j = 1, \dots, M$
充足可能 (各 c_j が真のリテラルを含む $U \rightarrow \{\text{真}, \text{偽}\}$ が存在
ただし、 u が真なら \bar{u} は偽)
- $\mathcal{M} = \{1, \dots, M\}$, $\mathcal{N} = \{0, 1, \dots, N\}$ の因子グラフの確率分布
 $P_{X_0}(1) > 0$

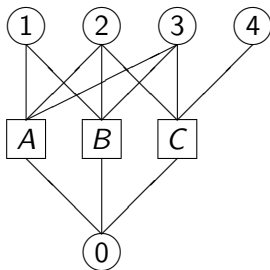
PRD: NP 完全 \implies 確率的推論 PR: NP 困難

確率的推論は NP 困難 (続)

例: $c_A = \{u_1, u_2, u_3\}$, $c_B = \{\bar{u}_1, \bar{u}_2, u_3\}$, $c_C = \{u_2, \bar{u}_3, u_4\}$

$$\begin{aligned}
 & P_{X_0 X_1 X_2 X_3 X_4}(x_0, x_1, x_2, x_3, x_4) \\
 &= \frac{1}{Z} f_A(x_0, x_1, x_2, x_3) f_B(x_0, x_1, x_2, x_3) f_C(x_0, x_2, x_3, x_4)
 \end{aligned}$$

- $i = 1, 2, 3, 4$ で、 $x_i = 1 \iff u_i$ が真
- $x_0 = 1$ のとき、 $a = A, B, C$ で、 $f_a(x_0, x_1, x_2, x_3) = 0 \iff c_a$ が偽



Junction Tree

$$P_{X_1 \dots X_N}(x_1, \dots, x_N) = \prod_{i=1}^N P(X_i = x_i | X_k = x_k, k \in \pi(i))$$

$$\pi(i) \subseteq \{1, \dots, i-1\}, \quad i = 1, \dots, N$$

確率分布 $P_{X_1 \dots X_N}$ から定義される無向グラフ $G = (V, E)$

$$V := \{1, \dots, N\}$$

$$E := \{\{i, k\} | i \in V, k \in \pi(i)\}$$

G の極大クリーク木 $T = (V_G, E_G)$

V_G : G の極大クリーク (V の部分集合であらわす) の集合

$$E_G := \{\{U, U'\} | U \cap U' \neq \{\}, U, U' \in V_G\}$$

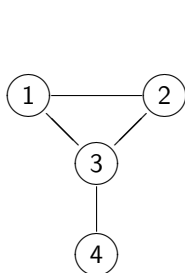
G の極大クリーク木 $T = (V_G, E_G)$ が **Junction Tree**

G の頂点が T の連結された頂点に含まれる (RIP)

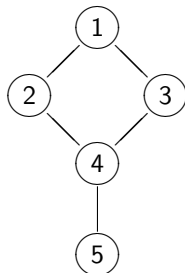
Junction Tree (続)

任意の無向グラフ G について、以下の各条件は同値

- 1 G が chordal
- 2 G が分解可能
- 3 G が再帰的に単純
- 4 G の接合木が存在



条件を満たす



条件を満たさない

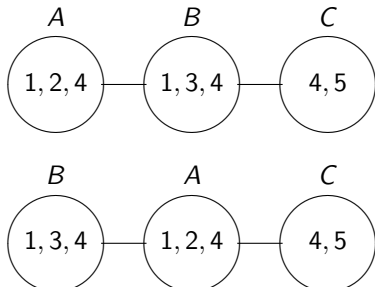
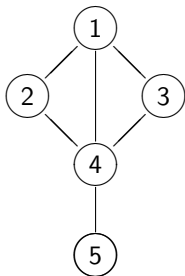
Junction Tree の構成方法

G の接合木 $T = (V_G, E_G)$ が存在しないとき、 G に辺を加えて chordal にする

Kruscal のアルゴリズムの適用

$U, U' \in V_G$ 間の重みを $|U \cap U'|$ とし、ループができない限り、重みの大きいものから辺を結ぶ

(例) $A = \{1, 2, 4\}$, $B = \{1, 3, 4\}$, $C = \{4, 5\}$



一般化確率伝播というよりは、一般化因数分解

Junction Tree を用いた確率推論の効率化

- ① G の接合木 $T = (V_G, E_G)$ が存在しないとき、chordal 化
- ② Junction Tree $T = (V_G, E_G)$ を構成における確率伝播

効率化・最適化をはかりたい

- G に辺を加えても確率的推論の解は変わらないが、計算量が増える
- T の各頂点 (極大クリーク) のサイズ (含んでいる G の頂点の数) が小さくなるように chordal 化をはかる

Partial K-Tree (NP 完全)

$G = (V, E)$ および正整数 K に対して、 $E \subseteq E'$ となる chordal な $G' = (V, E')$ で、 G' の各極大クリークのサイズが高々 K 以下のものが存在するか。

因子グラフから領域グラフへ: シナリオ

\mathcal{R} : 領域 $R(\subseteq \mathcal{N} \cup \mathcal{M})$ の集合 (領域の定義は、次のスライド)

C_R : 領域 R の領域定数

$$C_R = 1 - \sum_{R' \in A(R)} C_{R'}$$

$$\mathbf{x}_R \in X_R(\Omega) := \prod_{i \in \mathcal{N} \cap R} X_i(\Omega)$$

$$\prod_{a \in \mathcal{M}} f_a(\mathbf{x}_a) = \prod_{R \in \mathcal{R}} F_R(\mathbf{x}_R)^{C_R}$$

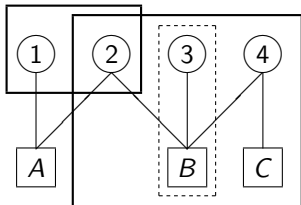
例)

$$P_{X_1 X_2}(x_1, x_2) P_{X_3 | X_2}(x_3 | x_2) = \frac{P_{X_1 X_2}(x_1, x_2) P_{X_2 X_3}(x_2, x_3)}{P_{X_2}(x_2)}$$

因子グラフから領域の集合を構成

領域 R

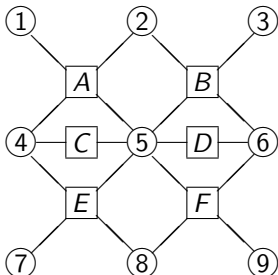
$$a \in \mathcal{M} \cap R \implies \mathcal{N}(a) \in R$$

を満足する $R \subseteq \mathcal{M} \cup \mathcal{N}$ □ 領域、 領域でない

領域グラフ: 領域を頂点とする DAG

領域グラフ: DAG (V, E)

- $V = \{R \in \mathcal{R}\}$, $(R, R') \in E \implies R \supsetneq R'$
- $\mathcal{N} \cup \mathcal{M}$ の同じ要素を含む R が連結



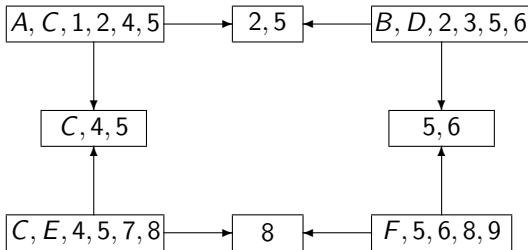
$$f_A(x_1, x_2, x_4, x_5) f_B(x_2, x_3, x_5, x_6) f_C(x_4, x_5) f_D(x_5, x_6) \\ \cdot f_E(x_4, x_5, x_7, x_8) f_F(x_5, x_6, x_8, x_9)$$

領域グラフ (1): Junction Graph

- 有向経路の長さが高々1 ($\mathcal{R} = \mathcal{R}_L \cup \mathcal{R}_S$, 親 \mathcal{R}_L と子 \mathcal{R}_S)
- $c_R = 1 - d_R$, $R \in \mathcal{R}_S$ (d_R : R の親の頂点の数)

$$\mathcal{R}_L = \{\{A, C, 1, 2, 4, 5\}, \{B, D, 2, 3, 5, 6\}, \{C, E, 4, 5, 7, 8\}, \{F, 5, 6, 8, 9\}\}$$

$$\mathcal{R}_S = \{\{2, 5\}, \{C, 4, 5\}, \{5, 6\}, \{8\}\}$$

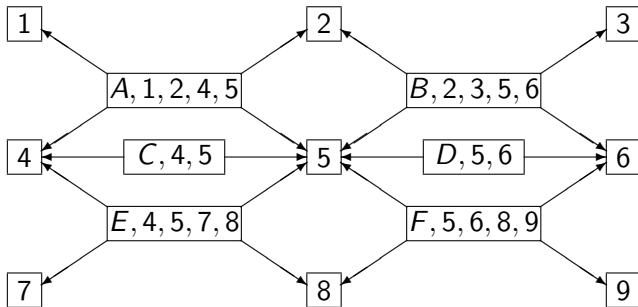


$$F_{AC1245}(x_1, x_2, x_4, x_5) F_{CE4578}(x_4, x_5, x_7, x_8) F_{BD2356}(x_2, x_3, x_5, x_6) \\ \cdot F_{F5689}(x_5, x_6, x_8, x_9) \cdot [F_{C45}(x_4, x_5) F_{25}(x_2, x_5) F_{56}(x_5, x_6) F_8(x_8)]^{-1}$$

領域グラフ (2): Bethe 近似

Junction Graph の特殊ケース

$$\mathcal{R}_S = \{\{i\} | i \in \mathcal{N}\}, \mathcal{R}_L = \{\{a\} \cup \{i \in \mathcal{N}(a)\} | a \in \mathcal{M}\}$$



$$F_{A1245}(x_1, x_2, x_4, x_5) F_{B2356}(x_2, x_3, x_5, x_6) F_{E4578}(x_4, x_5, x_7, x_8)$$

$$\cdot F_{F5689}(x_5, x_6, x_8, x_9) F_{C45}(x_4, x_5) F_{D56}(x_5, x_6)$$

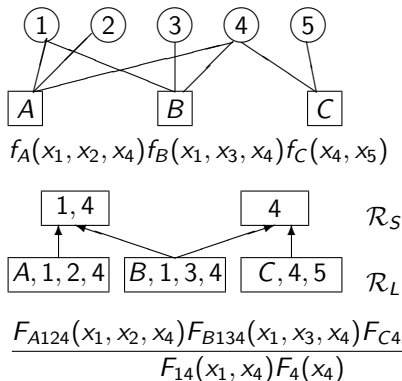
$$\cdot [F_2(x_1) F_4(x_4)^2 F_5(x_5)^5 F_6(x_6)^2 F_8(x_8)]^{-1}$$

領域グラフ (3): Junction Tree

Junction Graph の特殊ケース

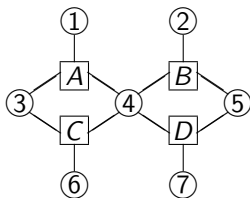
$R \in \mathcal{R}_S$ は、 $d_R = 2$ で、 $a \in \mathcal{M}$ を含まない

- $R \in \mathcal{R}_S$ が接続木の辺
- \mathcal{M} の部分集合が、接続木の頂点としての役割を果たす

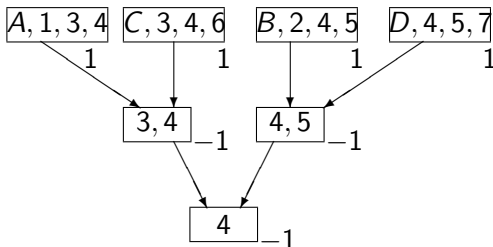


一般化確率伝播というよりは、一般化因数分解

領域グラフ (4): Kikuchi 近似 (世代数が 2 以上になる)



$$f_A(x_1, x_3, x_4) f_B(x_2, x_4, x_5) f_C(x_3, x_4, x_6) f_D(x_4, x_5, x_7)$$



$$\frac{F_{A134}(x_1, x_3, x_4) F_{C346}(x_3, x_4, x_6) F_{B245}(x_2, x_4, x_5) F_{D457}(x_4, x_5, x_7)}{F_{34}(x_3, x_4) F_{45}(x_4, x_5) F_4(x_4)}$$

一般化確率伝播というよりは、一般化因数分解

領域グラフにおける確率伝播 (一般化確率伝播)

$$R \in \mathcal{R}, P \in \mathcal{P}(R), C \in \mathcal{C}(R), F_R(\mathbf{x}_R) := \left(\prod_{a \in R \cap \mathcal{M}} f_a(\mathbf{x}_a) \right)^{c_R}$$

$$n_{R \rightarrow P}(\mathbf{x}_R) = F_R(\mathbf{x}_R) \prod_{P' \in \mathcal{P}(R) \setminus \{P\}} M_{P' \rightarrow R}(\mathbf{x}_R) \prod_{C \in \mathcal{C}(R)} N_{C \rightarrow R}(\mathbf{x}_C)$$

$$m_{R \rightarrow C}(\mathbf{x}_C) = \sum_{\mathbf{x}_R \setminus \mathbf{x}_C} F_R(\mathbf{x}_R) \prod_{P \in \mathcal{P}(R)} M_{P \rightarrow R}(\mathbf{x}_R) \prod_{C' \in \mathcal{C}(R) \setminus \{C\}} N_{C' \rightarrow R}(\mathbf{x}_{C'})$$

$$N_{R \rightarrow P}(\mathbf{x}_R) := n_{R \rightarrow P}(\mathbf{x}_R)^{\beta_R} m_{P \rightarrow R}(\mathbf{x}_R)^{\beta_R - 1}$$

$$M_{R \rightarrow C}(\mathbf{x}_C) := n_{C \rightarrow R}(\mathbf{x}_C)^{\beta_C - 1} m_{R \rightarrow C}(\mathbf{x}_C)^{\beta_C}$$

$$p_R: R \in \mathcal{R} \text{ の親の頂点の個数}, q_R := \begin{cases} 1 & p_R = 0 \\ (1 - c_R)/p_R & p_R \neq 0 \end{cases},$$

$$\beta_R := (2 - q_R)^{-1}$$

領域グラフが Bethe 近似のとき、通常確率伝播に一致

(Yielda, et. al, 2005)

一般的には NP 困難でも、因数分解でしのぎたい

理論:

- 無向グラフ → Junction Tree
- 因子グラフ → 領域グラフ

$$f(x_1, \dots, x_N) = \prod_{a \in \mathcal{M}} f_a(\mathbf{x}_a) = \prod_{R \in \mathcal{R}} F_R(\mathbf{x}_R)^{C_R}$$

$$N^* := \max_{R \in \mathcal{R}} |\mathcal{N} \cap R| \rightarrow \min$$

応用:

- 確率伝播
- 論理回路設計
- 遺伝的アルゴリズム (Mühlenbein, 2006)

論理回路設計への応用

論理回路設計

$f : \{0, 1\}^n \rightarrow \{0, 1\}$ を少ない項数で記述したい

$\pi(i) \subseteq \{1, \dots, i-1\}$ が決まり、条件付確率の表

$$\{\theta_i(x_i | (x_k)_{k \in \pi(i)})\}, x_i \in X_i(\Omega), (x_k)_{k \in \pi(i)} \in \prod_{k \in \pi(i)} X_k(\Omega)$$

から MLF 式 (N の指数) を求めて、 $\lambda_i(x_i) := 0, 1, i = 1, \dots, N$

$$\sum_{x_1} \cdots \sum_{x_N} \prod_{i=1}^N \lambda_i(x_i) \theta_i(x_i | (x_k)_{k \in \pi(i)})$$

計算時間を低減するには

因子頂点ごとではなく、領域ごとに条件付確率の表を用意

まとめ

- Junction Tree は、領域グラフに一般化される
- 確率伝播というよりは、因数分解が重要
- NP 困難でもしのげるかもしれない