# 鈴木譲「ベイジアンネットワーク入門」(培風館)
# 3. 統計的学習
# 3.5 有限型 Bayesian ネットワークの学習 (1)
# 一般的な方法

鈴木譲

大阪大学

2009 年 12 月 24 日 (1)

# あらまし

有限型 Bayesian ネットワーク

経営状態の例

事後確率最大化と記述長最小化

# 有限型 Bayesian ネットワーク

$X^{(1)}, \cdots, X^{(N)}$: $|X^{(i)}(\Omega)| < \infty$ なる確率変数
$\mathbf{U}_k := \{X^{(1)}, \cdots, X^{(k)}\},\ k \leq N$

### 定理 2.7 (Markov 境界が一意)

$$I(\mathbf{U}_{k-1} - \mathbf{B}_k, \mathbf{B}_k, \{X^{(k)}\})$$

となる極小の $\mathbf{B}_k \subseteq \mathbf{U}_{k-1},\ k = 1, 2, \cdots, N$ が一意

### 定理 2.10 (有限型 Bayesian ネットワークの構成法)

各 $X^{(j)} \in \mathbf{U}_{k-1}$ について

$$(X^{(j)}, X^{(k)}) \in \vec{\mathbf{E}} \Longleftrightarrow X^{(j)} \in \mathbf{B}_k$$

# 有限型 Bayesian ネットワーク (続)

順序 $X^{(1)} < X^{(2)} < \cdots < X^{(N)}$ のもとで $\{\mathbf{B}_k\}_{k=1}^N$ を定義
$\mathbf{B}_k = \{X^{(j)} \,|\, j \in \pi^{(k)}\}$
$\pi := (\pi^{(1)}, \cdots, \pi^{(N)})$, $\pi^{(k)} \subseteq \{1, 2, \cdots, k-1\}$

$$
\begin{aligned}
P_{X^{(1)} \ldots X^{(N)}} &= \prod_{j=1}^{N} P_{X^{(j)} | X^{(1)} \ldots X^{(j-1)}} \\
&= \prod_{j=1}^{N} P_{X^{(j)} | X^{(h)}, h \in \pi^{(j)}}
\end{aligned}
\tag{39}
$$

# Bayesian ネットワークの候補

$N = 3$
$X^{(1)} < X^{(2)} < X^{(3)}$

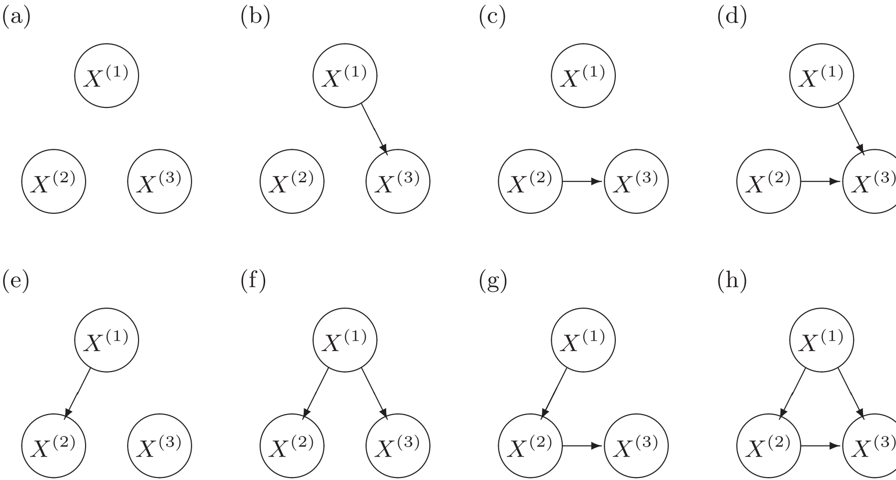
$$\begin{aligned} \pi &= (\pi^{(1)}, \pi^{(2)}, \pi^{(3)}) \\ &= (\{\}, \{\}, \{\}), (\{\}, \{\}, \{1\}), (\{\}, \{\}, \{2\}), (\{\}, \{\}, \{1, 2\}), \\ &\quad (\{\}, \{1\}, \{\}), (\{\}, \{1\}, \{1\}), (\{\}, \{1\}, \{2\}), (\{\}, \{1\}, \{1, 2\}) \end{aligned}$$

の $2^{(3-1)3/2} = 8$ 通りの非巡回有向グラフが存在
$\pi^{(1)} = \{\}$, $\pi^{(2)} = \{1\}$ の場合,
$\pi^{(3)} = \{1\}$ であれば (f),
$\pi^{(3)} = \{1, 2\}$ であれば (h) が対応

$$P_{X^{(1)}X^{(2)}X^{(3)}} = P_{X^{(1)}} P_{X^{(2)}|X^{(1)}} P_{X^{(3)}|X^{(1)}},$$

$$P_{X^{(1)}X^{(2)}X^{(3)}} = P_{X^{(1)}} P_{X^{(2)}|X^{(1)}} P_{X^{(3)}|X^{(1)}X^{(2)}}$$

# Bayesian ネットワークの候補 (続)



# 状態分割の推定に帰着

## $\pi^{(k)}$ の選択 $= \mathcal{S}^{(k)}$ の選択

$\mathcal{S}^{(k)} := \prod_{j \in \pi^{(k)}} X^{(j)}(\Omega)$ おくと,

$$P_{X^{(k)}|S^{(k)}} = P_{X^{(k)}|X^{(j)}, j \in \pi^{(k)}}$$

独立に発生した $n$ 個の $X^{(1)}(\Omega) \times \cdots \times X^{(N)}(\Omega)$ の要素の列

$$\begin{aligned} x_1 &:= (x_1^{(1)}, x_1^{(2)}, \cdots, x_1^{(N)}), \\ x_2 &:= (x_2^{(1)}, x_2^{(2)}, \cdots, x_2^{(N)}), \\ & \cdots\cdots \\ x_n &:= (x_n^{(1)}, x_n^{(2)}, \cdots, x_n^{(N)}) \end{aligned}$$

から、$P_{X^{(1)} \cdots X^{(N)}}$ の統計的推測

- $\pi$ の $2^{N(N-1)/2}$ 個の候補について、事後確率を比較
- 各 $P_{X^{(k)}|X^{(j)}, j \in \pi^{(k)}}$, $k = 1, \cdots, N$ を独立に推定

# 事後確率の最大化

$\alpha^{(k)} := |X^{(k)}(\Omega)|$

$\mathcal{S}^{(k)}$ が既知 ($\pi^{(k)}$ が既知) であり，$\theta_{y,s}$ を推定するだけであれば

$$\frac{c_{y,s} + a_{y,s}}{n_s + \sum_{y'} a_{y',s}},\ y \in X^{(k)}(\Omega),\ s \in \mathcal{S}^{(k)}$$

$$\frac{c_{y,s} + 1/2}{n_s + \alpha^{(k)}/2},\ y \in X^{(k)}(\Omega),\ s \in \mathcal{S}^{(k)}$$

$\mathcal{S}^{(k)}$ が未知 ($\pi^{(k)}$ が未知) であり，$\mathcal{S}^{(k)}$ を推定する必要があれば,

$$Q^{(\mathcal{S}^{(k)})}(z^n) := \prod_{s \in \mathcal{S}^{(k)}} \frac{\Gamma\left(\sum_{y \in X^{(k)}(\Omega)} a_{y,s}\right) \prod_{y \in X^{(k)}(\Omega)} \Gamma(c_{y,s} + 1/2)}{\left[\prod_{y \in X^{(k)}(\Omega)} \Gamma(a_{y,s})\right] \Gamma\left(n_s + \sum_{y' \in X^{(k)}(\Omega)} a_{y,s}\right)} \tag{40}$$

$$Q^{(\mathcal{S}^{(k)})}(z^n) := \prod_{s \in \mathcal{S}^{(k)}} \frac{\Gamma(\alpha^{(k)}/2) \prod_{y \in X^{(k)}(\Omega)} \Gamma(c_{y,s} + 1/2)}{(1/2)^{\alpha^{(k)}} \Gamma(n_s + \alpha^{(k)}/2)}$$

# 記述長の最小化

$$K[\pi^{(k)}] := (\alpha^{(k)} - 1) \prod_{j \in \pi^{(k)}} \alpha^{(j)},$$

$$\mathcal{H}[\pi^{(k)}](x^n) := \sum_{s \in \mathcal{S}^{(k)}} \sum_{y \in \mathcal{X}^{(k)}(\Omega)} c_{y,s} \log \frac{n_s}{c_{y,s}}$$

$$L[\pi^{(k)}](x^n) := \mathcal{H}[\pi^{(k)}](x^n) + \frac{K[\pi^{(k)}]}{2} \log n \tag{41}$$

# 経営状態の例

| 収益性 $X^{(1)}$ (問題なし 0) (あり 1) | 効率性 $X^{(2)}$ (問題なし 0) (あり 1) | 安全性 $X^{(3)}$ (問題なし 0) (あり 1) | 成長性 $X^{(4)}$ (問題なし 0) (あり 1) |
|---|---|---|---|
| 0 | 0 | 1 | 0 |
| 0 | 0 | 0 | 0 |
| 0 | 0 | 1 | 0 |
| 0 | 0 | 0 | 0 |
| 0 | 1 | 1 | 0 |
| 1 | 1 | 0 | 1 |
| 1 | 1 | 1 | 1 |
| 1 | 1 | 0 | 1 |
| 1 | 1 | 1 | 1 |
| 1 | 1 | 0 | 0 |

# 経営状態の例 (続)

$X^{(1)}(\Omega) = \{0,1\}$, $X^{(2)}(\Omega) = \{0,1\}$, $X^{(3)}(\Omega) = \{0,1\}$,
$X^{(4)}(\Omega) = \{0,1\}$
無条件に $\pi^{(1)} = \{\}$。効率性が収益性に依存するか否かで,

1. $\pi^{(2)} = \{\}$: $c_0 = 5$, $c_1 = 5$, $n = 10$,

$$L^{(2)}[\{\}](x^n) = -5\log\left(\frac{5}{10}\right) - 5\log\left(\frac{5}{10}\right) + \frac{1}{2}\log 10 = 11.661$$

2. $\pi^{(2)} = \{1\}$: $c_{0,0} = 4$, $c_{1,0} = 1$, $n_0 = 5$,
$c_{0,1} = 0$, $c_{1,1} = 5$, $n_1 = 5$,

$$\begin{aligned} & L^{(1)}(\{1\})(x^n) \\ = \ & -4\log\left(\frac{4}{5}\right) - 1\log\left(\frac{4}{5}\right) - 0\log\left(\frac{0}{5}\right) - 5\log\left(\frac{5}{5}\right) + \frac{2}{2}\log 10 \\ = \ & 4.932 \end{aligned}$$

となり, $\pi^{(2)} = \{1\}$ が選択。

## 経営状態の例 (続)

安全性が (収益性, 効率性) に依存するか否かで,

1. $\pi^{(3)} = \{\}$: $c_0 = 5,\ c_1 = 5,\ n = 10,$

$$L^{(3)}[\{\}](x^n) = -5\log\left(\frac{5}{10}\right) - 5\log\left(\frac{5}{10}\right) + \frac{1}{2}\log 10 = 11.661$$

2. $\pi^{(3)} = \{1\}$: $c_{0,0} = 2,\ c_{1,0} = 3,\ n_0 = 5,\ \ c_{0,1} = 3,$
$c_{1,1} = 2,\ n_1 = 5,$

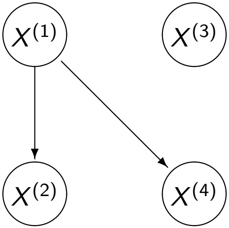
$$L^{(3)}[\{1\}](x^n) = -2\log\left(\frac{2}{5}\right) - 3\log\left(\frac{3}{5}\right) - 3\log\left(\frac{3}{5}\right) - 2\log\left(\frac{2}{5}\right) + \frac{2}{2}\log 10 = 13.032$$

3. $\pi^{(3)} = \{2\}$: $c_{0,0} = 2,\ c_{1,0} = 2,\ n_0 = 4,$
$c_{0,1} = 3,\ c_{1,1} = 3,\ n_1 = 6,$

$$L^{(3)}[\{2\}](x^n) = -2\log\left(\frac{2}{4}\right) - 2\log\left(\frac{2}{4}\right) + -3\log\left(\frac{3}{6}\right) - 3\log\left(\frac{3}{6}\right) + \frac{2}{2}\log 10 = 13.322$$

# 経営状態の例 (続)

成長性が (収益性，効率性，安全性) に依存するか否かで，
$L^{(4)}[\pi^{(4)}](x^n)$，$\pi^{(4)} \subseteq \{1,2,3\}$ を比較すると，$\pi^{(4)} = \{1\}$ が選択。



# 事後確率最大化と記述長最小化

記述長の最小化は，事後確率の最大化の近似にすぎないが、

- 例の適合度と構造の簡潔さのバランスを $\{d_n\}$ で調整可能 (3.4 節)。
- 2 項の和で表現され，計算量の低減のために分岐限定法 (3.5.2 項) や Chow-Liu アルゴリズム (3.5.3 項) が適用可能。