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 (3)
コスト最小極大木アルゴリズムの適用

鈴木譲

大阪大学

2009 年 12 月 24 日 (2)

あらまし

木への近似

木への近似

$$P_{X^{(j)}|X^{(0)}}(\cdot|x^{(0)}) = P_{X^{(j)}}(\cdot)$$

Dendroid 分布

$$P_{X^{(1)}\dots X^{(N)}}(\mathbf{x}^{(1)},\cdots,\mathbf{x}^{(N)}),\ \ \mathbf{x}^{(1)} \in X^{(1)}(\Omega),\cdots,\mathbf{x}^{(N)} \in X^{(N)}(\Omega)$$

$$\widehat{P}_{X^{(1)}\dots X^{(N)}}(\mathbf{x}^{(1)},\cdots,\mathbf{x}^{(N)}) := \prod_{j\in N} P_{X^{(a_j)}|X^{(a_{\lambda[j]})}}(x^{(a_j)}|x^{(a_{\lambda[j]})}) \quad (43)$$

$1 \le \lambda[j] \le j-1,\ j \ne 1,\ \ \lambda[1] = 0,$

$\mathbf{U} := \{X^{(i)} \mid i = 1,\cdots N\}$: 頂点集合

補題 3.3

無向グラフ $G = (\mathbf{U}, \mathbf{E})$ が木であることと，
$\mathbf{E} = \{\{X^{(a_j)}, X^{(a_{\lambda[j]})}\} \mid j = 1,\cdots,N\}$, $1 \le \lambda[j] \le j-1,\ j \ne 1,$
$\lambda[1]=0$ なる $(1,\cdots,N)$ の順列 $(a_1,\cdots,a_N)$ が存在することは同値

Chow-Liu アルゴリズム

$$
\begin{aligned}
&D(P_{X^{(1)}\dots X^{(N)}}||\widehat{P}_{X^{(1)}\dots X^{(N)}}) \\
&\quad := \sum_{\mathbf{x}^{(1)}\in X^{(1)}(\Omega),\cdots,\mathbf{x}^{(N)}\in X^{(N)}(\Omega)} P_{X^{(1)}\dots X^{(N)}}(x^{(1)},\cdots,x^{(N)}) \\
&\qquad\quad \cdot \log \frac{P_{X^{(1)}\dots X^{(N)}}(x^{(1)},\cdots,x^{(N)})}{\prod_{j\in N} P_{X^{(a_j)}|X^{(a_{\lambda[j]})}}(x^{(a_j)}|x^{(a_{\lambda[j]})})}
\end{aligned} \tag{44}
$$

$$
I(i,j) := \sum_{x^{(i)}\in X^{(i)}(\Omega),x^{(j)}\in X^{(j)}(\Omega)} P_{X^{(i)}X^{(j)}}(x^{(i)},x^{(j)}) \log \frac{P_{X^{(i)}X^{(j)}}(x^{(i)},x^{(j)})}{P_{X^{(i)}}(x^{(i)})P_{X^{(j)}}(x^{(j)})}
$$

命題 (Chow-Liu)

$D(P_{X^{(1)}\dots X^{(N)}}||\widehat{P}_{X^{(1)}\dots X^{(N)}})$ 最小の $(a_1,\cdots,a_N)$, $\lambda[k]$, $2\le k\le N$ は，アルゴリズム 2.1 に $\mathbf{U} := \{X^{(i)} \mid i=1,\cdots N\}$, $w(\{X^{(i)},X^{(j)}\}) := I(i,j)$ を入力することで得られる。

Chow-Liu アルゴリズム

証明: $D(P_{X^{(1)}\dots X^{(N)}}||\widehat{P}_{X^{(1)}\dots X^{(N)}}) = \sum_{j=1}^{N} -I(a_j, a_{\lambda[j]})+(\text{定数})$

例から出発して木を生成する

n 個の例
$x^n = (x_1, \cdots, x_n),\ \ x_i \in X(\Omega) = \prod_{j=1}^{N} X^{(j)}(\Omega),\ \ i = 1, \cdots, n$

$$I_n(i,j) := \sum_{x \in X^{(i)}, y \in X^{(j)}} c_{i,j}[x,y] \log \frac{c_{i,j}[x,y]}{c_i[x]c_j[y]} \geq 0$$

$c_i[x]$: $x_1, \cdots, x_n$ における $X^{(i)} = x \in X^{(i)}(\Omega)$ の頻度
$c_{i,j}[x,y]$: $x_1, \cdots, x_n$ における
$X^{(i)} = x \in X^{(i)}(\Omega),\ \ X^{(j)} = y \in X^{(j)}(\Omega)$ の同時頻度

$$\sum_{x^{(1)}, \cdots, x^{(N)}} c_{1,\cdots,N}(x^{(1)}, \cdots, x^{(N)}) \cdot \log \prod_{j=1}^{N} \frac{c_{a_{\lambda[j]}}(x^{(a_{\lambda[j]})})}{c_{a_j, a_{\lambda[j]}}(x^{(a_j)}, x^{(a_{\lambda[j]})})}$$

($\mathcal{H}[\pi](x^n)$ とおく) の最小化をはかる
$c_{1,\cdots,N}(x^{(1)}, \cdots, x^{(N)})$: $X^{(1)} = x^{(1)} \in X^{(1)}(\Omega), \cdots,$
$X^{(N)} = x^{(N)} \in X^{(N)}(\Omega)$ が同時に成立する頻度

例から出発して木を生成する

アルゴリズム 3.1: Chow-Liu アルゴリズム

入力: n 個の例 x^n

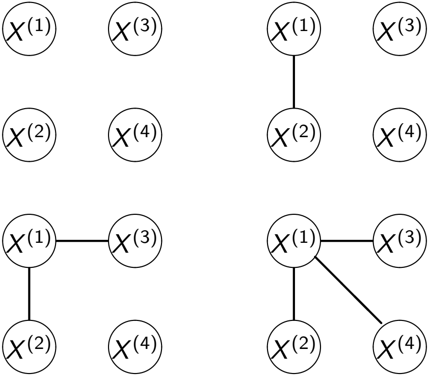
出力: $\mathcal{H}[\pi](x^n)$ を最小とする木 T

1. $\mathbf{E} \leftarrow \{\}$
2. $\mathcal{E} \neq \{\}$ である限り，以下を繰り返す。
   (1) $I_n(i,j)$ が最大の $\{X^{(i)}, X^{(j)}\} \in \mathcal{E}$ について，
   $$\mathcal{E} \leftarrow \mathcal{E} - \{\{X^{(i)}, X^{(j)}\}\}$$
   (2) $(\mathbf{U}, \mathbf{E} \cup \{\{X^{(i)}, X^{(j)}\}\})$ が巡回経路をもたないとき，
   $$\mathbf{E} \leftarrow \mathbf{E} \cup \{\{X^{(i)}, X^{(j)}\}\}$$
3. $T \leftarrow (\mathbf{U}, \mathbf{E})$

$\mathcal{H}[\pi](x^n)$(3.45) の値は，例 x^n の木への適合性とみなせる

実行例 (1)

i	j	$I_n(i,j)$
1	2	12
1	3	10
2	3	8
1	4	6
2	4	4
3	4	2



例から出発して森を生成する

$\mathcal{H}[\pi](x^n)$ の最小化ではなく，パラメータ数 $K[\pi]$ を加えた

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

の最小化

$$J_n(i,j) := I_n(i,j) - \frac{(\alpha^{(i)}-1)(\alpha^{(j)}-1)}{2}\log n$$

補題

無向グラフ $G = (\mathbf{U}, \mathbf{E})$ が森であることと，
$\mathbf{E} = \{\{X^{(a_j)}, X^{(a_{\lambda[j]})}\} \mid j = 1, \cdots, N\}$,
$0 \le \lambda[j] \le j-1,\ j = 1, \cdots, N,\ \lambda[1] = 0$ なる $(1, \cdots, N)$ の順列
$(a_1, \cdots, a_N)$ が存在することは同値。

木では $1 \le \lambda[j] \le j-1$，森では $0 \le \lambda[j] \le j-1$

例から出発して森を生成する (続)

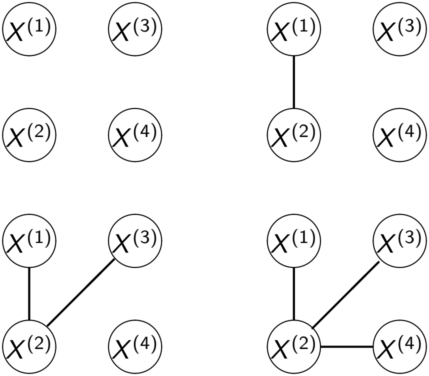
アルゴリズム 3.2: Suzuki, 1993

入力: n 個の例 x^n
出力: $L(x^n)[\pi]$ を最小とする森 F

1. $\mathbf{E} \leftarrow \{\}$
2. $\mathcal{E} \neq \{\}$ である限り，以下を繰り返す。
   (1) $J_n(i,j)$ が最大の $\{X^{(i)}, X^{(j)}\} \in \mathcal{E}$ について，
   $\mathcal{E} \leftarrow \mathcal{E} - \{\{X^{(i)}, X^{(j)}\}\}$
   (2) $(\mathbf{U}, \mathbf{E} \cup \{\{X^{(i)}, X^{(j)}\}\})$ が巡回経路をもたず，かつ
   $J_n(i,j) \geq 0$ のとき，$\mathbf{E} \leftarrow \mathbf{E} \cup \{\{X^{(i)}, X^{(j)}\}\}$
3. $F \leftarrow (\mathbf{U}, \mathbf{E})$

実行例 (2)

i	j	$I_n(i,j)$	$\alpha^{(i)}$	$\alpha^{(j)}$	$J_n(i,j)$
1	2	12	5	2	8
1	3	10	5	3	2
2	3	8	2	3	6
1	4	6	5	4	-6
2	4	4	2	4	1
3	4	2	3	4	-4



例から出発して森を生成する (続)

定理 3.10, Suzuki,1993

アルゴリズム 10 は，n 個の例 x^n から $L[\pi](x^n)$ を最小とする森 T を正しく出力する。

証明：

$$K[\pi] = \sum_{j=1}^{N} (\alpha^{(a_j)} - 1)\alpha^{(a_{\lambda[j]})}$$

$L[\pi](x^n)$ の最小化は，以下のの最小化と同値

$$\sum_{j=1}^{N} -I_n(a_j, a_{\lambda[j]}) + \frac{1}{2}\sum_{j=1}^{N}(\alpha^{(a_j)} - 1)(\alpha^{(a_{\lambda[j]})} - 1)\log n$$

$$= \quad -\sum_{j=1}^{N} J_n(a_j, a_{\lambda[j]})$$