鈴木譲「ベイジアンネットワーク入門」(培風館)
4. 確率的推論
4.2 確率的推論における NP 困難性

鈴木譲

大阪大学

2010 年 7 月 29 日 (木) (2)

あらまし

1. $\mathcal{P}$ と $\mathcal{NP}$

2. 確率的推論と最大事後確率設定 $\mathcal{NP}$ の困難性

問題の実行時間

計算機: 有限の 2 進列の入力に対して

- 停止して、有限の 2 進列を出力
- 停止しない

アルゴリズム

すべての入力に対して停止する対応 $f : \{0,1\}^* \to \{0,1\}^*$ の記述

アルゴリズム A の実行時間 (命令実行回数) T_A

A への入力 $\sigma \in \{0,1\}^*$ から実行時間 $T_A(\sigma)$ への対応

入力長が $n := |\sigma|$ のときのアルゴリズム A の実行時間 $T_{n,A}$

$$T_{n,A} := \max_{\sigma \in \{0,1\}^n} T_A(\sigma)$$

クラス $\mathcal{P}$

アルゴリズム A が多項式時間で解ける

有限個の n を除いて、$T_{n,A} \leq n^k$ となる k が存在

決定問題 $f : \{0,1\}^* \to \{0,1\}$

入力は

1. “1” が出力される 2 進列の集合 Σ
2. “0” が出力される 2 進列の集合 $\{0,1\}^* \backslash \Sigma$

に大別できる (言語 $\Sigma \subseteq \{0,1\}^*$ の認識)

$\mathcal{P}$

決定性 TM のアルゴリズムで、多項式時間で解ける決定問題の集合

クラス $\mathcal{NP}$

決定性 TM 命令の実行が毎回 1 個のみ実行

非決定性 TM 命令の実行が毎回複数同時に実行可 (仮想的)
1 個の決定性 TM で Σ の認識が完了すればよい

$\mathcal{NP}$

非決定性 TM のアルゴリズムで、多項式時間で解ける決定問題の集合

- $\mathcal{P} \subseteq \mathcal{NP}$
- $\mathcal{P} \neq \mathcal{NP}$ であることが**強く予想**されている

$\mathcal{NP}$ 完全

言語 Σ は言語 Σ' に多項式時間で還元 ($\Sigma \propto \Sigma'$)

$f : \{0,1\}^* \to \{0,1\}^*$ s.t.

- $f(\Sigma) \subseteq \Sigma'$
- $f(\{0,1\}^* \backslash \Sigma) \subseteq \{0,1\}^* \backslash \Sigma'$

を多項式時間で解く決定性 TM のアルゴリズムが存在

$\mathcal{NP}$ 完全な言語

任意の $\Sigma \in \mathcal{NP}$ に対して $\Sigma \propto \Sigma'$ となる言語 Σ'

$\mathcal{P} \neq \mathcal{NP}$ を仮定すると

命題 4.2

1. $\Sigma \propto \Sigma', \Sigma' \propto \Sigma'' \Longrightarrow \Sigma \propto \Sigma''$
2. $\Sigma \propto \Sigma', \Sigma' \in \mathcal{P} \Longrightarrow \Sigma \in \mathcal{P}$
3. Σ が $\mathcal{NP}$ 完全, $\Sigma \propto \Sigma' \Longrightarrow \Sigma'$ が $\mathcal{NP}$ 完全

$\Sigma \in \mathcal{NP}$ 完全を示せば、$\Sigma \notin \mathcal{P}$ の強い証拠になる

$\Sigma \in \mathcal{NP}$ 完全, $\Sigma \in \mathcal{P}$ なる Σ が存在 $\Longrightarrow \mathcal{P} = \mathcal{NP}$ (予想と矛盾)

充足可能性問題 SAT

1. 有限集合 $U = \{u_1, u_2, \cdots, u_n\}$ に対して、
$$t : U \to \{T, F\}$$
2. $t(u_i) = F \iff t(\bar{u}_i) = T$ なる $\bar{U} = \{\bar{u}_1, \bar{u}_2, \cdots, \bar{u}_n\}$ に対して、
$$t : U \cup \bar{U} \to \{T, F\}$$
3. $c \subseteq U \cup \bar{U}$ に対して、$t(c) = T \iff t(z) = T$ for $\exists z \in c$
4. U 上の節集合 C に対して、$t(C) = T \iff t(c) = T$ for $\forall c \in C$

リテラル　$U \cup \bar{U}$ の要素

節　$U \cup \bar{U}$ の部分集合 c ($|c|$: 節の大きさ、要素数)

節集合　節を要素とする集合

充足可能性問題 SAT (続)

事例：　有限集合 U, U 上の節集合 C

質問：　$t(C) = T$ となる $t : U \to \{T, F\}$ が存在するか。

1. $U = \{u_1, u_2\}$, $C = \{\{u_1, \bar{u}_2\}, \{\bar{u}_1, u_2\}\}$ 「存在する」
   $t(u_1) = t(u_2) = T$ なる t に対して $t(C) = T$。
2. $U = \{u_1, u_2\}$, $C = \{\{u_1, u_2\}, \{u_1, \bar{u}_2\}, \{\bar{u}_1\}\}$ 「存在しない」
   $t(C) = T$ なる t が存在しない。

符号化

決定問題:

入力: 事例 $I \in$ 問題 Π

出力: T または F

符号化 $\varphi : \Pi \to \{0, 1\}^*$

事例 $I \in \Pi$ を 2 進列に対応

問題 Π の $\mathcal{NP}$ 完全性

$\varphi(\Pi)$ についての $\mathcal{NP}$ 完全性

符号化の例

$\Psi := \{0, 1, -, [,], (,), ,\}$

1. $x \in \mathbb{Z}$: 非負の数は 2 進、負の数は前に $-$。
2. $[x]$: で数値ではなく、識別のための記号。
3. $(x_1, \cdots, x_m)$: $x_1, \cdots, x_m$ からなる列。

Ψ の要素	0	1	$-$	[	]	(	)	,
2 進列	000	001	010	011	100	101	110	111

1. $U = \{u_1, u_2\}$, $C = \{\{u_1, \bar{u}_2\}, \{\bar{u}_1, u_2\}\}$:
   `(([1],[10]),(([1],[-10]),([-1],[10])))`
   を 2 進列に直した $41 \times 3 = 123$ ビットが入力の長さ
2. $U = \{u_1, u_2\}$, $C = \{\{u_1, u_2\}, \{u_1, \bar{u}_2\}, \{\bar{u}_1\}\}$:
   `([1],[10]),(([1],[10]),([1],[-10]),([-1]))`
   を 2 進列に直した $48 \times 3 = 144$ ビットが入力の長さ

Cook の定理

Cook の定理

SAT は、$\mathcal{NP}$ 完全問題である。

(証明略)

Π, Π': 問題

問題 Π が問題 Π' に多項式時間で還元 $\Pi \propto \Pi'$

言語の多項式時間還元と同様に定義

1. $\Pi \propto \Pi', \Pi' \propto \Pi'' \Longrightarrow \Pi \propto \Pi''$
2. $\Pi \propto \Pi', \Pi' \in \mathcal{P} \Longrightarrow \Pi \in \mathcal{P}$
3. Π が $\mathcal{NP}$ 完全, $\Pi \propto \Pi' \Longrightarrow \Pi'$ が $\mathcal{NP}$ 完全

3-充足可能性問題 3SAT

事例： 有限集合 U, U 上の大きさ 3 の節からなる節集合 C
質問： $t(C) = T$ となる $t : U \to \{T, F\}$ が存在するか。

命題 4.3

3SAT は、$\mathcal{NP}$ 完全である。

$U = \{u_1, \cdots, u_n\}$, $C = \{c_1, \cdots, c_m\}$: SAT の任意の事例
C と C' の充足可能性が一致するように 3SAT の事例
$U' := U \cup (\cup_{j=1}^m U_j')$, $C' := \cup_{j=1}^m C_j'$, $c_j = \{z_1, \cdots, z_k\}$

k	U_j'	C_j'
1	$\{y_j^1, y_j^2\}$	$\{\{z_1, y_j^1, y_j^2\}, \{z_1, y_j^1, \bar{y}_j{}^2\}, \{z_1, \bar{y}_j{}^1, y_j^2\}, \{z_1, \bar{y}_j{}^1, \bar{y}_j{}^2\}\}$
2	$\{y_j^1\}$	$\{\{z_1, z_2, y_j^1\}, \{z_1, z_2, \bar{y}_j^1\}\}$
3	ϕ	$\{c_j\}$
≥ 4	$\{y_j^i \mid 1 \leq i$	$\{\{z_1, z_2, y_j^1\}, \{\bar{y}_j^1, z_3, y_j^2\}, \{\bar{y}_j^2, z_4, y_j^3\}, \cdots,$
	$\leq k-3\}$	$\{\bar{y}_j^{k-5}, z_{k-3}, y_j^{k-4}\}, \{\bar{y}_j^{k-4}, z_{k-2}, y_j^{k-3}\}, \{\bar{y}_j^{k-3}, z_{k-1}, z_k\}\}$

3-充足可能性問題 3SAT

$k \geq 4$:

$$t(z_1) = F, \cdots, t(z_{l-1}) = F, t(z_l) = T$$

l	i	$t'(y_j^i)$
$l = 1, 2$	$1 \leq i \leq k-3$	F
$3 \leq l \leq k-2$	$1 \leq i \leq l-2$	T
	$l-1 \leq i \leq k-3$	F
$l = k-1, k$	$1 \leq i \leq k-3$	T

逆に、$t' : U' \to \{T, F\}$ が C' を充足させれば、t' の定義域 U' を U に制限して、C を充足させる $t : U \to \{T, F\}$ が得られる。

SAT の事例 U, C が mn の多項式の長さで記述でき、決定的 TM を用いて，その長さの多項式の時間で 3SAT の事例が計算

$\mathcal{NP}$ 困難

$\mathcal{NP}$ 完全 決定問題に対して定義

$\mathcal{NP}$ 困難 決定問題ではなくともよいが、その問題が決定性 TM で多項式時間で解かれると、$\mathcal{P} = \mathcal{NP}$ を意味する問題

例:

事例： 有限集合 U, U 上の節集合 C

質問： $t(C) = T$ となる $t : U \to \{T, F\}$ が存在するとき、その 1 個を出力せよ

確率的推論と最大事後確率設定 $\mathcal{NP}$ の困難性

PR: Probablistic Reasoning (確率的推論)
MAP: Maximum A Posteriori (最大事後確率)

確率的推論 (PR)

事例： $P_{X_1 \cdots X_N}(x_1, \cdots, x_N)$, $x_1 \in \mathcal{X}_1, \cdots, x_N \in \mathcal{X}_N$, $i \in \mathcal{N}$
質問： $P_{X_i}(x_i)$, $x_i \in \mathcal{X}_i$

最大事後確率設定 (MAP)

事例： $P_{X_1 \cdots X_N}(x_1, \cdots, x_N)$, $x_1 \in \mathcal{X}_1, \cdots, x_N \in \mathcal{X}_N$
質問： $argmax_{x_1, \cdots, x_N} P_{X_1 \cdots X_N}(x_1, \cdots, x_N)$

確率的推論と最大事後確率設定 $\mathcal{NP}$ の困難性

確率的推論の決定問題版 (PRD)

事例： $P_{X_1 \cdots X_N}(x_1, \cdots, x_N)$, $x_1 \in \mathcal{X}_1, \cdots, x_N \in \mathcal{X}_N$, $i \in \mathcal{N}$, $x_i \in \mathcal{X}_i$

質問： $P_{X_i}(x_i) > 0$ が成立するか

最大事後確率設定の決定問題版 (MAPD)

事例： $P_{X_1 \cdots X_N}(x_1, \cdots, x_N)$, $x_1 \in \mathcal{X}_1, \cdots, x_N \in \mathcal{X}_N$, 有理数 $0 < p < 1$

質問： $P_{X_1 \cdots X_N}(x_1, \cdots, x_N) > p$ なる $x_1 \in \mathcal{X}_1, \cdots, x_N \in \mathcal{X}_N$ は存在するか。

PR は $\mathcal{NP}$ 困難

定理 4.3

PRD は $\mathcal{NP}$ 完全、すなわち PR は $\mathcal{NP}$ 困難

証明: 3SAT からの多項式時間還元性のみを示す。
$U = \{u_1, \cdots, u_n\}$, $C = \{c_1, \cdots, c_m\}$: 3SAT の事例

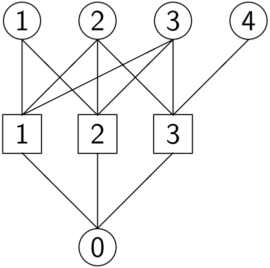
$$C \text{ が } U \text{ 上充足可能} \iff \text{ある } i \in \mathcal{N} \text{ と } x_i \in \mathcal{X}_i \text{ について } P_{X_i}(x_i) > 0$$

なる $P_{X_1 \cdots X_N}(x_1, \cdots, x_N)$ を構成

PR は $\mathcal{NP}$ 困難 (続)

$\mathcal{N} = \{0, 1, \cdots, n\}$ (変数頂点集合)
$\mathcal{M} := \{1, \cdots, m\}$ (因数頂点集合)
$\mathcal{N}(j)$: c_j, $j \in \mathcal{M}$ のリテラル u_i の 3 個の添え字 $i \in \mathcal{N}$ と 0 からなる集合
$j \in \mathcal{M}$ と $i \in \mathcal{N}(j)$ を辺として結んで因数グラフを構成
$\mathcal{M}(i)$, $i \in \mathcal{N}$ が定義

例 4.10 $m = 3$, $n = 4$



PR は $\mathcal{NP}$ 困難 (続)

$$f_j(u_0, c_j) = \begin{cases} 1, & u_0 = \mathrm{F},\ \mathrm{or},\ c_j = \mathrm{T} \\ 0, & \mathrm{otherwise} \end{cases}$$

$$P_{U_0 \cdots U_n}(u_0, \cdots, u_n) = K \prod_{j \in \mathcal{M}} f_j(u_0, c_j) \tag{21}$$

$$K := \left[\sum_{u_0, \cdots, u_n} \prod_{j \in \mathcal{M}} f_j(u_0, c_j) \right]^{-1}$$

PR は $\mathcal{NP}$ 困難 (続)

$$
\begin{aligned}
& C \text{ が } U \text{ 上充足可能} \\
\Longrightarrow \quad & c_1 = \cdots = c_m = \mathrm{T} \\
& \text{なる } u_1, \cdots, u_n \in \{\mathrm{T}, \mathrm{F}\} \text{ が存在} \\
\Longrightarrow \quad & \text{各 } j \in \mathcal{M} \text{ について，} \ f_j(u_0 = \mathrm{T}, c_j) = 1 \\
& \text{なる } u_i \in \{\mathrm{T}, \mathrm{F}\},\ i \in \mathcal{N}(j) \backslash \{0\} \text{ が存在} \\
\Longrightarrow \quad & P_{U_0 \cdots U_n}(u_0 = \mathrm{T}, u_1, \cdots, u_n) \propto \textstyle\prod_{j \in \mathcal{M}} f_j(u_0 = \mathrm{T}, c_j) = 1 \\
& \text{なる } u_i \in \{\mathrm{T}, \mathrm{F}\}, i \in \mathcal{N} \backslash \{0\} \text{ が存在} \\
\Longrightarrow \quad & P_{U_0}(u_0 = \mathrm{T}) > 0
\end{aligned}
$$

PR は $\mathcal{NP}$ 困難 (続)

$$
\begin{array}{ll}
 & C \text{ が } U \text{ 上充足可能でない} \\
\Longrightarrow & c_1 = \cdots = c_m = \mathrm{T} \text{ となる } u_1, \cdots, u_n \in \{\mathrm{T}, \mathrm{F}\} \text{ が存在しない} \\
\Longrightarrow & \text{任意の } u_i \in \{\mathrm{T}, \mathrm{F}\},\ i \in \mathcal{N}(j) \backslash \{0\} \text{ について} \\
 & f_j(u_0 = \mathrm{T}, c_j) = 0, j \in \mathcal{M} \text{ が存在} \\
\Longrightarrow & \text{任意の } u_i \in \{\mathrm{T}, \mathrm{F}\},\ i \in \mathcal{N} \backslash \{0\} \text{ について} \\
 & P_{U_0 \cdots U_n}(u_0 = \mathrm{T}, u_1, \cdots, u_n) \propto \prod_{j \in \mathcal{M}} f_j(u_0 = \mathrm{T}, c_j) = 0 \\
\Longrightarrow & P_{U_0}(u_0 = \mathrm{T}) = 0
\end{array}
$$

3SAT の事例 U, C が mn の多項式の長さで記述でき、決定的 TM を用いて，その長さの多項式の時間で PRD の事例が計算

例 4.10

$U=\{u_1,u_2,u_3,u_4\}$, $C=\{\{u_1,u_2,u_3\},\{\overline{u}_1,\overline{u}_2,u_3\},\{u_2,\overline{u}_3,u_4\}\}$ であれば, $\mathcal{N}=\{1,2,3,4\}$, $\mathcal{M}=\{1,2,3\}$ $\mathcal{N}(1)=\{0,1,2,3\}$, $\mathcal{N}(2)=\{0,1,2,3\}$, $\mathcal{N}(3)=\{0,2,3,4\}$, $u_0\neq *$ のとき,

$$f_1(u_0,u_1,u_2,u_3)=\begin{cases}1, & u_0=\mathrm{F} \text{ または } c_1=\mathrm{T},\\ 0, & \text{それ以外},\end{cases}$$

$$f_2(u_0,u_1,u_2,u_3)=\begin{cases}1, & u_0=\mathrm{F} \text{ または } c_2=\mathrm{T},\\ 0, & \text{それ以外},\end{cases}$$

$$f_3(u_0,u_2,u_3,u_4)=\begin{cases}1, & u_0=\mathrm{F} \text{ または } c_3=\mathrm{T}\\ 0, & \text{それ以外}\end{cases}$$

$$P_{U_0U_1U_2U_3U_4}(u_0,u_1,u_2,u_3,u_4) \propto f_1(u_0,u_1,u_2,u_3)f_2(u_0,u_1,u_2,u_3)f_3(u_0,u_2,u_3,u_4)$$

例 4.10

$$C(u_1, u_2, u_3, u_4) = \mathrm{T} \iff f_1(u_0 = \mathrm{T}, u_1, u_2, u_3) f_2(u_0 = \mathrm{T}, u_1, u_2, u_3) \cdot f_3(u_0 = \mathrm{T}, u_2, u_3, u_4) = 1$$

u_0	u_1	u_2	u_3	u_4	c_1	c_2	c_3	C	f_1	f_2	f_3	$f_1 f_2 f_3$
F	F	F	F	F	F	T	T	F	1	1	1	1
F	F	F	F	T	F	T	T	F	1	1	1	1
F	F	F	T	F	T	T	F	F	1	1	1	1
F	F	F	T	T	T	T	T	T	1	1	1	1
F	F	T	F	F	T	T	T	T	1	1	1	1
F	F	T	F	T	T	T	T	T	1	1	1	1
F	F	T	T	F	T	T	T	T	1	1	1	1
F	F	T	T	T	T	T	T	T	1	1	1	1
F	T	F	F	F	T	T	T	T	1	1	1	1
F	T	F	F	T	T	T	T	T	1	1	1	1
F	T	F	T	F	T	T	F	F	1	1	1	1
F	T	F	T	T	T	T	T	T	1	1	1	1
F	T	T	F	F	T	F	T	F	1	1	1	1
F	T	T	F	T	T	F	T	F	1	1	1	1
F	T	T	T	F	T	T	T	T	1	1	1	1
F	F	T	T	T	T	T	T	T	1	1	1	1
T	F	F	F	F	F	T	T	F	0	1	1	0
T	F	F	F	T	F	T	T	F	0	1	1	0
T	F	F	T	F	T	T	F	F	1	1	0	0

PR が解ければ、SAT が解ける

アルゴリズム 4.1 を適用:

1. $n_{i\to j}(u_i) := 1,\ \ m_{j\to i}(u_i) := 1,\ \ i \in \mathcal{N}, j \in \mathcal{M}$
2. $j \in \mathcal{M}$ について,

$$m_{j\to i}(u_i) := \sum_{u_l,\ l\in\mathcal{N}(j)\backslash\{i\}} f_j(u_0, c_j) \prod_{k\in\mathcal{N}(j)\backslash\{i\}} n_{k\to j}(u_k),$$

$$n_{i\to j}(u_i) := \prod_{k\in\mathcal{M}(i)\backslash\{j\}} m_{k\to i}(u_i).$$

3. $b_0(u_0) \propto \prod_{j\in\mathcal{M}} m_{j\to 0}(u_0)$

2., 3. を繰り返す。

系 4.1

系 4.1, Shimony, 1994

MAPD は $\mathcal{NP}$ 完全，すなわち MAP は $\mathcal{NP}$ 困難である。

証明: 定理 4.3 の証明で，(21) の K の値は高々2^{n+1} なので，

$$P_{U_0\cdots U_n}(u_0=\mathrm{T},\ u_1,\cdots,u_n)\left\{\begin{array}{ll}\geq 2^{-n-1}, & u_1,\cdots,u_m\ satsifiable\\ =0, & u_1,\cdots,u_m\ non-satisfiable\end{array}\right.$$

$p := 2^{-n-1}$

$\mathcal{NP}$ 困難の注意点

- $\mathcal{NP}$ 困難性は、すべての場合に計算時間がかかることを意味しない
- 問題の中に計算時間がかかる事例が含まれていることのみを主張
- 条件を加えれば、$\mathcal{NP}$ 困難性が崩れることがある

例: PRD や MAPD でも，因子グラフが巡回経路を含まない場合，$\mathcal{P}$ 問題

接合木における $\mathcal{NP}$ 困難問題

接合木における $\mathcal{NP}$ 困難問題 (第 2 章)

$\max_{\mathbf{C}\in\mathbf{E}_T}|\mathbf{C}|$ を最小にする $T=(\mathbf{U}_T,\mathbf{E}_T)$ を見いだす問題が $\mathcal{NP}$ 困難

Partial K-Tree が $\mathcal{NP}$ 完全であることによる

事例　：$G=(\mathbf{U},\mathbf{E})$ および正整数 K

質問　：以下を満足する chordal な $G'=(\mathbf{U},\mathbf{E}')$ が存在するか。

(1) $\mathbf{E}\subseteq\mathbf{E}'$

(2) G' の各クリークの頂点数が高々 K 以下。

- 無向グラフに辺を加えて chordal にすることはいつでも可能
- 任意の無向グラフで、各頂点の最大値を最小にする接合木がわかれば、Partial K-Tree の任意の K について解けたことになる。