アルゴリズム論 1 第 5 回 文脈自由文法 (2)

関川 浩

2017/05/17

第4回から第7回の目標

第4回から第7回の目標

正規表現と fa: よくできたシステムだが能力が低い

より能力が高いシステムを導入する

- 文脈自由文法 (第 4, 5 回)
- プッシュダウンオートマトン (第 6, 7 回)

第 5 回の目標:

- 文脈自由文法の標準形の導入
- 文脈自由文法では生成できない言語の例

- ① 文脈自由文法の標準形
 - Chomsky 標準形, Greibach 標準形
 - 単一規則の排除
 - Chomsky 標準形の存在
 - A 生成規則
 - Greibach 標準形の存在
- 2 文脈自由文法では生成できない言語
 - cfg の能力の限界
 - uvwxy 定理
 - 例題

- 1 文脈自由文法の標準形
- ② 文脈自由文法では生成できない言語

文脈自由文法の標準形

この節の仮定

扱う文脈自由言語は ε を含まない (一般性を大きく失う制限ではない) (\iff 「 $A \to \varepsilon$ 」の形の生成規則がない cfg で生成可能)

定義 1 (Chomsky 標準形)

G の生成規則が以下の形のみのとき, Chomsky 標準形という $A \to BC$ $(B, C \in V)$, $A \to a$ $(a \in \Sigma)$

定義 2 (Greibach 標準形)

G の生成規則が以下の形のみのとき, Greibach 標準形という

$$A \to a\alpha \quad (a \in \Sigma, \ \alpha \in V^*)$$

注意

Greibach 標準形が長さ n の列を導出 \Longrightarrow 規則の適用回数は n 回

Greibach 標準形の応用

cfg P_1 (Chomsky 標準形) と P_2 (Greibach 標準形) (開始記号 S)

$$P_1 = \{S \to AX, \ S \to CC, \ X \to SB, \ A \to 0, \ B \to 1, \ C \to 2\}$$

$$P_2 = \{S \to 0SB, \ S \to 2A, \ A \to 2, \ B \to 1\}$$

 \Longrightarrow 同じ言語 L を生成

問題

 $x = 00221 \in L \ \text{th}$?

解答

P2 を使えば簡単

 $x \in L$ と仮定すると, x の最左導出は以下しかあり得ない

$$S\Rightarrow 0SB\Rightarrow 00SBB\Rightarrow 002ABB\Rightarrow 0022BB\Rightarrow 00221B$$

B が残るのでx は導出できない

単一規則の排除 (1/3)

定義 3 (単一規則)

<mark>単一規則</mark>: $A \rightarrow B$ (A, B) は変数) の形をした生成規則

補題 1

 $\operatorname{cfg} G$ に対し, L(G') = L(G) かつ単一規則がない $\operatorname{cfg} G'$ が存在

証明 (1/3)

$$G=(V,\Sigma,P,S)$$
 とする. 以下の P' は単一規則を含まない
$$P'=\{p\mid p\in P\$$
は単一ではない生成規則 $\}$
$$\cup \{A\to\alpha\mid G\$$
により $A\stackrel{*}{\Rightarrow}B\$ ($A,B\in V$) かつ
$$B\to\alpha\$$
は P の単一ではない生成規則 $\}$ $G'=(V,\Sigma,P',S)$ に対し, $L(G')=L(G)$ を示せばよい

単一規則の排除 (2/3)

証明 (2/3)

• $L(G) \subseteq L(G')$ $x \in L(G)$ に対し, G による最左導出を考える

$$S = \alpha_0 \Rightarrow \alpha_1 \Rightarrow \cdots \Rightarrow \alpha_n = x$$

G の単一ではない生成規則で $\alpha_i \Rightarrow \alpha_{i+1}$ なら, G' でも $\alpha_i \Rightarrow \alpha_{i+1}$

単一規則の適用後,必ず単一ではない生成規則が適用される

- (a) $\alpha_i \Rightarrow \alpha_{i+1} \Rightarrow \cdots \Rightarrow \alpha_j$ はすべて G の単一規則
- (b) $\alpha_j \Rightarrow \alpha_{j+1}$ は G の単一ではない生成規則
- (a) において置き換えられる変数はすべて同じ位置
- $\Longrightarrow P'$ のある一つの生成規則によって $\alpha_i \Rightarrow \alpha_{j+1}$
- $\Longrightarrow G'$ でも $S \stackrel{*}{\Rightarrow} x$ となり $L(G) \subseteq L(G')$

単一規則の排除 (3/3)

証明 (3/3)

• $L(G') \subseteq L(G)$ $A \to \alpha$ が P' に属すなら, G により $A \stackrel{*}{\Rightarrow} \alpha$ $\Longrightarrow G'$ により $S \stackrel{*}{\Rightarrow} x \in \Sigma^*$ なら, G によっても $S \stackrel{*}{\Rightarrow} x$ $\Longrightarrow L(G') \subseteq L(G)$

Chomsky 標準形の存在 (1/2)

定理1

L が文脈自由言語なら, L を生成する, Chomsky 標準形である cfg が存在

証明 (1/2)

$$G = (V, \Sigma, P, S)$$
: $L = L(G)$ となる cfg. P は単一規則を含まない

$$G_1 = (V \cup \{X_t \mid t \in \Sigma\}, \ \Sigma, \ P_1, \ S)$$
, ただし, P_1 は以下の通り

- $P \ \ \, \mathbb{K} \ \, X_t \to t \ \, (t \in \Sigma) \ \,$ を追加
- P の $A \rightarrow \alpha$ ($|\alpha| \ge 2$) は, α 中の $t \in \Sigma$ を変数 X_t で置き換え

例: $A \rightarrow aABb$ は $A \rightarrow X_aABX_b$ に置き換え

注意

 $(A \rightarrow \alpha) \in P$ に対し, $|\alpha| = 1$ なら $\alpha \in \Sigma$ (P には単一規則がないから)

Chomsky 標準形の存在 (2/2)

証明 (2/2)

P₁ に属する生成規則は以下のいずれか

- (a) 右辺は終端記号 1 個
- (b) 右辺は変数のみからなり, 長さが 2
- (c) 右辺は変数のみからなり, 長さが3以上

 $A \to B_1 \dots B_n \ (n \ge 3)$ に対し、新しい変数 Y_1, \dots, Y_{n-2} を導入以下の置き換えをすれば Chomsky 標準形になる

$$A \to B_1 Y_1,$$

 $Y_1 \to B_2 Y_2, \quad Y_2 \to B_3 Y_3, \quad \dots, \quad Y_{n-3} \to B_{n-2} Y_{n-2},$
 $Y_{n-2} \to B_{n-1} B_n$

例題 1 (1/4)

例題 1

以下の生成規則で与えられる cfg G を Chomsky 標準形に直せ (開始記号は A)

$$A \to B, \quad A \to C, \quad B \to D, \quad D \to E, \quad E \to B, \quad C \to F,$$

 $B \to b, \quad E \to ADa, \quad C \to ABB$

注意

左辺が F の生成規則はないから, 実は F と $C \rightarrow F$ は不要

解答 (1/4)

与えられた生成規則のうち単一規則ではないものは,

$$B \to b$$
, $E \to ADa$, $C \to ABB$

例題 1 (2/4)

解答 (2/4)

まず、 Ψ 一規則を排除し、新しい規則を追加 (補題 1) するため、 各変数 X に対し $X \stackrel{\Rightarrow}{\to} Y$ となる変数 Y をすべて求める

- (1) X に番号 0 をつけ, i = 0 として (2) へ
- (2) 以下の条件を満たす変数を探す

番号 *i* がついている変数から

- 1 ステップで導出でき
- ・まだ番号のついていない変数

あれば、そういうすべての変数に番号 i+1 をつけ (3) へなければ終了. 番号のついている変数が答

(3) iを1増やして(2)へ

例題 1 (3/4)

解答 (3/4)

● A からはすべての変数に到達できることが分かるので

$$A \to b$$
, $A \to ABB$, $A \to ADa$

が追加される規則

B からは B, D, E に到達可能なので

$$B \to b$$
, $B \to ADa$

が追加される規則

• . . .

その後,単一規則を除く

例題 1 (4/4)

解答 (4/4)

最終的には生成規則は以下の通り

$$\begin{array}{llll} A \rightarrow b, & A \rightarrow AY_1, & Y_1 \rightarrow DX_a, & A \rightarrow AY_2, & Y_2 \rightarrow BB, \\ B \rightarrow b, & B \rightarrow AY_3, & Y_3 \rightarrow DX_a, \\ D \rightarrow b, & D \rightarrow AY_4, & Y_4 \rightarrow DX_a, \\ E \rightarrow b, & E \rightarrow AY_5, & Y_5 \rightarrow DX_a, & C \rightarrow AY_6, & Y_6 \rightarrow BB, \\ X_a \rightarrow a, & X_b \rightarrow b \end{array}$$

補題 2 (1/2)

定義 4 (A 生成規則)

A 生成規則: 左辺が変数 A である生成規則

補題 2

$$G = (V, \Sigma, P, S)$$
: cfg

- $A \to \alpha_1 B \alpha_2$: P に属する生成規則
 - $\{B \rightarrow \beta_i \mid i=1, 2, ..., r\}$: すべての B 生成規則の集合
 - $P' = (P \setminus \{A \to \alpha_1 B \alpha_2\}) \cup \{A \to \alpha_1 \beta_i \alpha_2 \mid i = 1, 2, \dots, r\}$

$$G' = (V, \Sigma, P', S)$$
 とすると, $L(G) = L(G')$

証明 (1/2)

 $\bullet \ L(G') \subseteq L(G)$

G' による導出中に $A \Rightarrow \alpha_1 \beta_i \alpha_2$ というステップがあれば, G では $A \Rightarrow \alpha_1 B \alpha_2 \Rightarrow \alpha_1 \beta_i \alpha_2$ とすればよい

補題 2 (2/2)

証明 (2/2)

• $L(G) \subseteq L(G')$ $A \to \alpha_1 B \alpha_2$: G にあって G' にはない唯一の生成規則

 $A \to \alpha_1 B \alpha_2$ が G による導出に出現 \Longrightarrow 変数 B は, $B \to \beta_i$ によっていつかは書き換え

導出結果は生成規則の適用順によらないから

$$A \Rightarrow \alpha_1 B \alpha_2 \Rightarrow \alpha_1 \beta_i \alpha_2$$

としてよい

これは, G' で $A \Rightarrow \alpha_1 \beta_i \alpha_2$ に置き換えが可能

$A \rightarrow A\alpha$ の除去 (1/3)

補題 3

$$G = (V, \Sigma, P, S)$$
: cfg

- (1) $(A \rightarrow A\alpha_i) \in P$ $(i=1,\ldots,r)$: A が右辺の左端にある A 生成規則
- (2) $(A \to \beta_j) \in P$ (j = 1, ..., s): 残りの A 生成規則

P': P から (1) を削除し, 以下を追加したもの (Z: 新しい変数)

$$Z \to \alpha_i, \ Z \to \alpha_i Z \quad (i = 1, \ldots, r)$$

 $A \to \beta_j Z \quad (j = 1, \ldots, s)$

$$G' = (V \cup \{Z\}, \Sigma, P', S)$$
 とすると, $L(G') = L(G)$

$A \rightarrow A\alpha$ の除去 (2/3)

証明 (1/2)

• $L(G) \subseteq L(G')$ $x \in L(G)$ とする

G における x の最左導出中, $P \setminus P'$ に属する生成規則が現れる導出

$$\gamma_1 A \gamma_2 \Rightarrow \gamma_1 A \alpha_{j_1} \gamma_2 \Rightarrow \gamma_1 A \alpha_{j_2} \alpha_{j_1} \gamma_2 \Rightarrow \cdots$$
$$\Rightarrow \gamma_1 A \alpha_{j_p} \dots \alpha_{j_1} \gamma_2 \Rightarrow \gamma_1 \beta_i \alpha_{j_p} \dots \alpha_{j_1} \gamma_2$$

は, G' においては以下で導出できる

$$\gamma_1 A \gamma_2 \Rightarrow \gamma_1 \beta_i Z \gamma_2 \Rightarrow \gamma_1 \beta_i \alpha_{j_p} Z \gamma_2 \Rightarrow \cdots$$
$$\Rightarrow \gamma_1 \beta_i \alpha_{j_p} \dots \alpha_{j_2} Z \gamma_2 \Rightarrow \gamma_1 \beta_i \alpha_{j_p} \dots \alpha_{j_1} \gamma_2$$

よって, $x \in L(G')$, すなわち, $L(G) \subseteq L(G')$

$A \rightarrow A\alpha$ の除去 (3/3)

証明 (2/2)

- $L(G')\subseteq L(G)$ $x\in L(G')$ とし、G' における x の最左導出を考える
 - $P' \setminus P$ に属する生成規則 (Z を含む生成規則) の適用があれば、以後、生成規則の適用順を変更、Z を左辺に持つものを優先
 - \bullet G' において, Z が現れてから消えるまでの部分

$$\gamma_1 A \gamma_2 \Rightarrow \gamma_1 \beta_i \mathbf{Z} \gamma_2 \Rightarrow \gamma_1 \beta_i \alpha_{j_p} \mathbf{Z} \gamma_2 \Rightarrow \cdots$$
$$\Rightarrow \gamma_1 \beta_i \alpha_{j_p} \dots \alpha_{j_2} \mathbf{Z} \gamma_2 \Rightarrow \gamma_1 \beta_i \alpha_{j_p} \dots \alpha_{j_1} \gamma_2$$

は,Gにおいては以下で導出可能

$$\gamma_1 A \gamma_2 \Rightarrow \gamma_1 A \alpha_{j_1} \gamma_2 \Rightarrow \gamma_1 A \alpha_{j_2} \alpha_{j_1} \gamma_2 \Rightarrow \cdots$$
$$\Rightarrow \gamma_1 A \alpha_{j_p} \dots \alpha_{j_1} \gamma_2 \Rightarrow \gamma_1 \beta_i \alpha_{j_p} \dots \alpha_{j_1} \gamma_2$$

よって, $x \in L(G)$, すなわち, $L(G') \subseteq L(G)$

Greibach 標準形の存在 (1/5)

定理 2

L が文脈自由言語なら, L を生成する, Greibach 標準形である cfg が存在

証明 (1/5)

 $G = (\{A_1, \ldots, A_m\}, \Sigma, P, A_1)$: L を生成する Chomsky 標準形

- (1) 生成規則を修正して,以下の二条件を満たすようにする
 - (#) $A_i \rightarrow A_i \gamma$ なら i < j
 - (b) 生成規則の右辺は高々 1 個の終端記号のあとに 0 個以上の 変数の列

Chomsky 標準形は (b) を満たすことに注意

Greibach 標準形の存在 (2/5)

証明 (2/5)

- (1) の続き
- i が小さい方から順に生成規則を修正していく
 - ullet i=1 のとき $A_1
 ightarrow A_1 \gamma$ という生成規則が存在すれば、補題 3 を適用して

 $(新しく導入する変数を <math>Z_1$ とする), (\sharp), (\flat) を満たすようにできる

Greibach 標準形の存在 (3/5)

証明 (3/5)

- (1) の続き
 - $i \le k$ で (\sharp), (\flat) を満たすとして, i = k+1 のとき
 - ullet $A_{k+1}
 ightarrow A_j \gamma \ (k+1>j)$ には補題 2 を適用 $(A_{k+1}$ 生成規則は変更していないので, γ は一つの変数)

$${A_j \rightarrow \beta_{jp}}: A_j$$
 生成規則の全体

$$A_{k+1} \rightarrow A_j \gamma$$
 を削除し $A_{k+1} \rightarrow \beta_{jp} \gamma$ を追加 (β_{ip}) は (b) を満たすので, $\beta_{ip} \gamma$ もそう)

$$\beta_{jp}$$
 の左端が A_q なら、帰納法の仮定より、 $q>j$ ⇒ 有限ステップで $A_{k+1}\to A_l\gamma$ $(k+1\leq l)$ の形に

ullet $A_{k+1}
ightarrow A_{k+1} \gamma$ には補題 3 を適用 (新しい変数を Z_{k+1})

Greibach 標準形の存在 (4/5)

証明 (4/5)

- (2) 条件 (‡), (b) を満たすようになると, 生成規則は以下の形のみ
 - (a) $A_k \to A_l \gamma$, $k < l, \gamma \in (V \cup \{Z_1, \dots, Z_m\})^*$
 - (b) $A_k \to \alpha \gamma$, $\alpha \in \Sigma$, $\gamma \in (V \cup \{Z_1, \dots, Z_m\})^*$
 - (c) $Z_k \to \gamma$, $\gamma \in (V \cup \{Z_1, \ldots, Z_m\})^*$

(a), (b) の場合:

- k=m のときは、右辺の左端は終端記号
- k=m-1 のときは、右辺の左端は終端記号か A_m
 - $\implies A_m$ のときは補題 2 を適用して, 右辺が終端記号で始まるものに置き換え可能. (\flat) も満たす
- $k=m-2, \ldots, k=1$ も同様にすれば, A_i 生成規則はすべて (b) の形にできる

Greibach 標準形の存在 (5/5)

証明 (5/5)

- (2) の続き
- (c) の Z_i 生成規則の場合:
 - (*) Z_i 生成規則の右辺の左端が A_j の場合 補題 2 を適用して A_j を置き換え, 右辺の左端を終端記号に
 - (†) Z_i 生成規則の右辺の左端が Z_j の場合補題 2 を適用して Z_j を置き換え

左端の置き換えを繰り返す

 \Longrightarrow いつか必ず A_l が左端に現れるので, (*) に帰着 $(Z_k$ は変数の列にのみ置き換わるから)

- ① 文脈自由文法の標準形
- 2 文脈自由文法では生成できない言語

cfg の能力の限界

cfg は正規表現や fa より真に能力が高いしかし, cfg でも生成できない言語が存在

- fa では認識できない言語の存在証明 記憶が有限という性質を利用
- cfg では生成できない言語の存在証明 複数の繰り返し構造を関連づける能力には制限があることを利用

注意

fa では、複数の繰り返し構造に関連をつけることができない

例: $\{0^n1^n \mid n \ge 0\}$ は正規言語ではない (fa では認識できない)

この節の仮定

この節の仮定

扱う文脈自由言語の生成規則は以下の 2 条件を満たす

- 右辺が ε の生成規則は、あるとすれば「 $S \to \varepsilon$ 」のみ (S は開始記号)
- 「 $S \rightarrow \varepsilon$ 」があるとき, S が右辺に現れる生成規則なし

命題1

任意の文脈自由言語は上記 2 条件を満たす cfg で生成可能

証明

文脈自由言語についての 5ページの主張

 ε を含まない \iff 「 $A \to \varepsilon$ 」の形の生成規則がない cfg で生成可能

から明らか

uvwxy 定理 (1/3)

定理 3 (uvwxy 定理)

L: 文脈自由言語, 無限集合

L によって決まる定数 K が存在して, $|z| \ge K$ である任意の列 $z \in L$ に対し, 以下の 4 条件を満たす列 u, v, w, x, y が存在

- \bullet z = uvwxy
- 任意の $i \ge 0$ に対して $uv^i wx^i y \in L$
- $vx \neq \varepsilon$
- $|vwx| \leq K$

注意

- u, v, w, x, y は L の元でなくてもよい
- 定理 3 は, 挿入定理, 反復定理, ポンプ定理 (補題) などともいう

uvwxy 定理 (2/3)

証明 (1/2)

L=L(G) となる cfg $G=(V,\Sigma,P,S)$ を取る $r=\max\{|\alpha|\mid (A o\alpha)\in P\}$ (= G の導出木の枝分れの最大数)

 \implies 根から葉までの最長のパス (枝をたどる経路) の長さが m である 導出木は、高々 r^m 個の葉しかもたない

 $K = r^{|V|+2}$ とおき, $|z| \ge K$ を満たす $z \in L$ を取る

- ullet $S \stackrel{*}{\Rightarrow} z$: 対応する導出木 T の頂点数が最少となる導出
- \bullet γ : T の最長のパスの一つ

$$(T$$
 の葉の数) $\geq |z| \geq K > r^{|V|+1}$

- $\Longrightarrow \gamma$ の長さは |V|+2 以上
- $\Longrightarrow \gamma$ の経路上, 葉から |V|+2 個目の頂点までに二回以上現れる $A \in V$ が存在

uvwxy 定理 (3/3)

証明 (2/2)

- A を根とする小さい方の部分木 T_1 : $A \stackrel{*}{\Rightarrow} w$
- A を根とする大きい方の部分木 T_2 : $A \stackrel{*}{\Rightarrow} vwx$

右図の通り z = uvwxy と分解

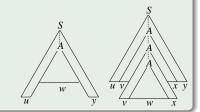


- ullet T_2 を T_1 で置き換え \Longleftrightarrow $S \stackrel{*}{\Rightarrow} uwy$
- ullet T_1 を T_2 で i-1 回 $(i\geq 1)$ 置き換え \Longleftrightarrow $S \stackrel{*}{\Rightarrow} uv^iwx^iy$

 $v = x = \varepsilon$ なら, $S \stackrel{*}{\Rightarrow} uwy = z$ の導出木の頂点数は T の頂点数より少 \Longrightarrow 矛盾

 T_2 の最長パスの長さは高々 |V|+2

$$\implies |vwx| \le r^{|V|+2} = K$$



例題 2 (1/2)

例題 2

 $L = \{a^n b^n c^n \mid n \ge 1\}$ は文脈自由言語ではないことを示せ

証明 (1/2)

L が文脈自由言語であると仮定

K:L に対する uvwxy 定理の定数

 $a^K b^K c^K \in L$ を uvwxy 定理の 4 条件を満たすよう uvwxy と分解

 $\implies vx \neq \varepsilon \ \text{\sharp b } v \neq \varepsilon \ \text{\sharp b } \delta v \ \text{\sharp } \epsilon \neq \varepsilon$

 $v \neq \varepsilon$ とする $(x \neq \varepsilon)$ の場合も同様)

(1) $v = a^i b^j$ $(i, j \ge 1)$ の場合 $(v = b^i c^j (i, j \ge 1))$ の場合も同様) $uv^2 wx^2 y = a^K b^j a^i \cdots \notin L$ となり uvwxy 定理に矛盾

例題 2 (2/2)

証明 (2/2)

- (2) $v = a^i$ の場合 ($v = b^i$, $v = c^i$ の場合も同様)
 - $x = \varepsilon$ なら, $uv^2wx^2y = a^{K+i}b^Kc^K \notin L$ 一方, uvwxy 定理より $uv^2wx^2y \in L$ だから矛盾
 - $x \neq \varepsilon$ なら、(1) より $x = a^j$ 、 b^j 、 c^j のいずれか $\Longrightarrow b$ 、c のどちらかは v にも x にも現れない よって、b、c のどちらかは uvwxy、 uv^2wx^2y に同数出現 a の個数は後者の方が多い $\Longrightarrow uv^2wx^2y \not\in L$ 一方、uvwxy 定理より $uv^2wx^2y \in L$ だから矛盾

文脈自由言語の共通部分が文脈自由言語とならない例

- $L_1 = \{a^nb^nc^m \mid n, m \ge 1\}$, $L_2 = \{a^mb^nc^n \mid n, m \ge 1\}$ はいずれも 文脈自由言語
 - L₁ を生成する生成規則 (開始記号は S)

$$S \to S_1C$$
, $S_1 \to aS_1b$, $S_1 \to ab$, $C \to cC$, $C \to c$

 \bullet L_2 を生成する生成規則 (開始記号は S)

$$S \to AS_2$$
, $S_2 \to bS_2c$, $S_2 \to bc$, $A \to aA$, $A \to a$

• $L_1 \cap L_2 = \{a^n b^n c^n \mid n \ge 1\}$ は文脈自由言語ではない (例題 2)

例題 3 (1/2)

例題 3

 $L = \{z^2 \mid z \in \{a,b\}^*\}$ は文脈自由言語ではないことを示せ

注意

 $\{zz^{\mathbf{R}} \mid z \in \{a,b\}^*\}$ は文脈自由言語 (第 4 回の例題 2)

証明 (1/2)

L を文脈自由言語と仮定して<mark>矛盾</mark>を導く

 $z' = a^K b^K a^K b^K \in L$ とする (K: L に対する uvwxy 定理の定数)

 \implies uvwxy 定理 の 4 条件を満たす分解 z' = uvwxy が存在

 $|vwx| \leq K$ より、vwx は $a^K b^K$ あるいは $b^K a^K$ に入る

a^K	b^K	a^K	b^K

例題 3 (2/2)

証明 (2/2)

 $a^K b^K = pvwxq$ (あるいは $b^K a^K = pvwxq$) とする

注意: pv^2wx^2q の最初の K 個は a (b), 最後の K 個は b (a)

ullet vwx が最初の a^Kb^K に含まれる場合 (二番目の場合も同様)

• vwx が b^Ka^K に含まれる場合

いずれも矛盾