アルゴリズム論 1
第 4 回
文脈自由文法 (1)

関川 浩

2017/05/10

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 4 回の目標:

- 文脈自由文法の導入
- 文脈自由文法で生成される言語が正規言語を真に含むこと

1. 文脈自由文法
   - 文脈自由文法の重要性
   - 文脈自由文法の定義
   - 例題
   - 文脈自由言語の性質
2. 文脈自由文法と正規言語
   - 正規言語と文脈自由言語の関係

文脈自由文法の重要性 (1/2)

- 形式言語において, 文法の典型的クラス
- 表現できる言語の重要性:
  多くのプログラミング言語がこのクラス
- プッシュダウンオートマトンと言語の表現能力が等価

文脈自由文法の重要性 (2/2)

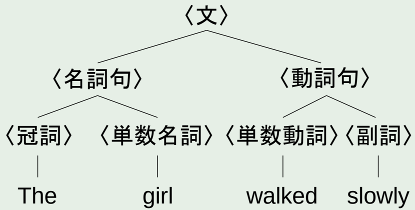
「文法」のイメージに合う

例

英文法における文の構造の解析

“The girl walked slowly.”

⟨ 文 ⟩	→	⟨ 名詞句 ⟩⟨ 動詞句 ⟩
⟨ 名詞句 ⟩	→	⟨ 冠詞 ⟩⟨ 単数名詞 ⟩
⟨ 動詞句 ⟩	→	⟨ 単数動詞 ⟩⟨ 副詞 ⟩
⟨ 冠詞 ⟩	→	The
⟨ 単数名詞 ⟩	→	girl
⟨ 単数動詞 ⟩	→	walked
⟨ 副詞 ⟩	→	slowly



文脈自由文法 (cfg) の定義

定義 1 (文脈自由文法 (cfg))

cfg (context-free grammar): $G = (V, \Sigma, P, S)$

- V, Σ, P: 空集合ではない有限集合, $V \cap \Sigma = \emptyset$
- 変数: V の要素. 大文字で表わすことが多い
  終端記号: Σ の要素. 小文字で表わすことが多い
- 生成規則: P の要素. "$A \to \alpha$" の形 ($A \in V$, $\alpha \in (V \cup \Sigma)^*$)
- 開始記号: 特別な変数 S

注意

"$A \to \alpha$" という生成規則は, A がどのような文脈に現れるかに無関係に A を α に置き換えることを許す

$\Longrightarrow$ 文脈自由という呼び名

cfg の例

例

$G_1 = (\{S\}, \{0,1\}, \{S \to \varepsilon, S \to 0S1\}, S)$ は cfg

「導出」という操作によって $x \in \{0,1\}^*$ を導く

$$S \Rightarrow 0S1 \Rightarrow 00S11 \Rightarrow 000S111 \Rightarrow 000111$$

G_1 が表現する言語:

導出により得られる $x \in \{0,1\}^*$ の全体 $= \{0^n 1^n \mid n \geq 0\}$

注意

$\{0^n 1^n \mid n \geq 0\}$ は正規言語ではない (第 3 回の定理 3)

一般に, 変数は複数, 各変数に適用可能な生成規則も複数

$\Longrightarrow$ 自由度が高く, 多くの列を導出することが可能

諸定義

- $G = (V, \Sigma, P, S)$: cfg
  - 列 $u, v \in (V \cup \Sigma)^*$ が以下の条件を満たすとき, v は u から**1 ステップで導出される**といい, $u \Rightarrow v$ と書く

    条件: ある $\alpha, \gamma \in (V \cup \Sigma)^*$ と生成規則 $B \to \beta$ が存在して,
    $u = \alpha B \gamma$ かつ $v = \alpha \beta \gamma$
  - $u_0 \Rightarrow u_1 \Rightarrow \cdots \Rightarrow u_n$ $(n \geq 0)$ のとき, u_n は u_0 から**導出される**といい, $u_0 \stackrel{*}{\Rightarrow} u_n$ と書く
  - $x \in \Sigma^*$ に対し,

    x が G によって**生成される** $\stackrel{\text{def}}{\Longleftrightarrow}$ x が S から導出される
  - G が**生成する言語** $L(G)$: G が生成する列の全体
- **文脈自由言語**: cfg によって生成される言語

例題 1 (1/2)

例題 1

$\{0^m 1^n \mid m,\ n \geq 0,\ m \neq n\}$ を生成する cfg を求めよ

解答

$G_2 = (\{S, S', A, B\}, \{0, 1\}, P, S)$, ただし,

$$P = \{S \to AS',\ S \to S'B,\ S' \to \varepsilon,\ S' \to 0S'1,\\ A \to 0A,\ A \to 0,\ B \to B1,\ B \to 1\}$$

- $0^m 1^n$ $(m,\ n \geq 0,\ m \neq n)$ が G_2 で生成されること
  - 0 の方が多い列 $0^{i+j}1^i$ $(j \geq 1)$ は以下の通り

$$S \Rightarrow AS' \overset{*}{\Rightarrow} A0^iS'1^i \Rightarrow A0^i1^i \overset{*}{\Rightarrow} 0^{j-1}A0^i1^i \Rightarrow 0^j0^i1^i$$

  - 0^i1^{i+j} $(j \geq 1)$ も同様

例題 1 (2/2)

解答 (再掲)

$G_2 = (\{S, S', A, B\}, \{0, 1\}, P, S)$, ただし,

$$P = \{S \to AS',\ S \to S'B,\ S' \to \varepsilon,\ S' \to 0S'1,\\ A \to 0A,\ A \to 0,\ B \to B1,\ B \to 1\}$$

- $0^m 1^n$ $(m,\ n \geq 0,\ m \neq n)$ のみが G_2 で生成されること
  最初に使える生成規則は $S \to AS'$, $S \to S'B$ のみ
  - $S \to AS'$
    $A \overset{*}{\Rightarrow} x \in \{0, 1\}^*$ なら, $x = 0^j$ $(j \geq 1)$
    $S' \overset{*}{\Rightarrow} y \in \{0, 1\}^*$ なら, $y = 0^i 1^i$ $(i \geq 0)$
    よって, $AS' \overset{*}{\Rightarrow} z \in \{0, 1\}^*$ なら, $z = 0^j 0^i 1^i$ $(j \geq 1,\ i \geq 0)$
  - $S \to S'B$ も同様

例題 2

例題 2

$G_3 = (\{S\}, \{0,1\}, P, S)$ が生成する言語は何か.

ただし, $P = \{S \to \varepsilon, S \to 0S0, S \to 1S1\}$

解答

$L(G_3) = \{xx^{\mathrm{R}} \mid x \in \{0,1\}^*\}$

注意

$\{xx \mid x \in \{0,1\}^*\}$ は文脈自由言語ではない (次回示す)

例題 3 (1/2)

例題 3

$L = \{0,1\}^* \setminus \{xx^{\mathrm{R}} \mid x \in \{0,1\}^*\}$ を生成する cfg を求めよ

解答 (1/2)

$x \in L \iff$ (a) または (b) が成立

(a) $|x|$ は奇数

(b) $x = \alpha 0 \beta 1 \alpha^{\mathrm{R}}$ または $x = \alpha 1 \beta 0 \alpha^{\mathrm{R}}$ ($\exists \alpha, \beta \in \{0,1\}^*$)

例題 3 (2/2)

解答 (2/2)

生成規則のみ示す (開始記号は S)

$$
\begin{aligned}
&S \to S_1,\ S \to S_2,\\
&S_1 \to 0,\ S_1 \to 1,\ S_1 \to 00S_1,\ S_1 \to 01S_1,\ S_1 \to 10S_1,\ S_1 \to 11S_1,\\
&S_2 \to 0S_20,\ S_2 \to 1S_21,\ S_2 \to 0A1,\ S_2 \to 1A0,\\
&A \to \varepsilon,\ A \to 0A,\ A \to 1A
\end{aligned}
$$

- "$S_1 \overset{*}{\Rightarrow} x \in \Sigma^*$" $\iff$ x は条件 (a) を満たす
- "$S_2 \overset{*}{\Rightarrow} x \in \Sigma^*$" $\iff$ x は条件 (b) を満たす

$$
S_2 \overset{*}{\Rightarrow} xS_2x^{\mathrm{R}} \Rightarrow \begin{cases} x0A1x^{\mathrm{R}} \\ x1A0x^{\mathrm{R}} \end{cases}
$$

A からは任意の列が生成されることに注意

文脈自由言語の性質 (1/2)

定理 1

L_1 と L_2 が文脈自由言語なら, 以下の言語も文脈自由言語

(1) $L_1 \cup L_2$

(2) $L_1 L_2$

(3) L_1^*

証明 (1/2)

$G_1 = (V_1, \Sigma, P_1, S_1)$: L_1 を生成する cfg

$G_2 = (V_2, \Sigma, P_2, S_2)$: L_2 を生成する cfg

$V_1 \cap V_2 = \emptyset$ と仮定してよい (必要なら変数を変更)

文脈自由言語の性質 (2/2)

証明 (2/2)

(1) $G = (V, \Sigma, P, S)$ は $L_1 \cup L_2$ を生成する cfg. ただし,
- $V = V_1 \cup V_2 \cup \{S\}$ $(S \notin V_1 \cup V_2)$
- $P = P_1 \cup P_2 \cup \{S \to S_1,\ S \to S_2\}$

(2) $G = (V, \Sigma, P, S)$ は $L_1 L_2$ を生成する cfg. ただし,
- $V = V_1 \cup V_2 \cup \{S\}$ $(S \notin V_1 \cup V_2)$
- $P = P_1 \cup P_2 \cup \{S \to S_1 S_2\}$

(3) $G = (V, \Sigma, P, S_1)$ は L_1^* を生成する cfg. ただし,
- $P = P_1 \cup \{S_1 \to S_1 S_1\}$ ■

定理 1 に関する注意

注意

- L_1, L_2 が正規言語のとき
  - $L_1 \cup L_2$, $L_1 \cap L_2$, $\Sigma^* \setminus L_1$, $L_1 L_2$, L_1^* も正規言語 (第 3 回の定理 2)
- L_1, L_2 が文脈自由言語のとき
  - $L_1 \cup L_2$, $L_1 L_2$, L_1^* も文脈自由言語 (定理 1)
  - $L_1 \cap L_2$, $\Sigma^* \setminus L_1$ は文脈自由言語になるとは限らない

    $\Longrightarrow$ 具体例を第 5 回 ($L_1 \cap L_2$), 第 7 回 ($\Sigma^* \setminus L_1$) に

最左導出

定義 2 (最左導出)

一番左の変数を常に置き換えるような導出を最左導出と呼ぶ

定理 2

cfg G によって，終端記号のみからなる列 x が導出されるとき，x は最左導出によっても導出される

あとで導出木を用いて証明

例題 4 (1/3)

例題 4

以下の cfg G_4 が生成する言語は何か (開始記号は S)

$$S \to SS, \quad S \to \varepsilon, \quad S \to aB, \quad S \to bA,$$
$$A \to a, \quad A \to bAA, \quad B \to b, \quad B \to aBB$$

解答

$L(G_4) = \{x \in \{a, b\}^* \mid x$ 中の a, b の個数が等しい $\}$

注意

カッコ文は以下で生成される (a が左カッコ, b が右カッコ)

$$S \to SS, \quad S \to \varepsilon, \quad S \to aB, \quad B \to b, \quad B \to aBB$$

例題 4 (2/3)

解答の証明 (1/2)

(1) $x \in L(G_4)$ なら, x 中の a, b の個数が等しいこと

- 開始記号 S に対し, 以下の式の値は 0

$$\{(a \text{ の個数}) + (A \text{ の個数})\} - \{(b \text{ の個数}) + (B \text{ の個数})\} \quad (\#)$$

- すべての生成規則の適用前後で (#) の値は不変

$\Longrightarrow$ x 中の a, b の個数は等しい (x 中に A, B はない)

例題 4 (3/3)

解答の証明 (2/2)

(2) x 中の a, b の個数が等しいなら, $x \in L(G_4)$ となること

1. $x = \varepsilon$ なら, x は $S \to \varepsilon$ で生成
2. $x \neq \varepsilon$ なら, x を左から見ていき, a, b の個数が等しくなったら区切りを入れて部分列に分ける

   例: $x = abbaabaababb$ なら $ab|ba|ab|aababb$

3. $S \to SS$ を複数回適用して, 部分列の数だけ S を並べる
4. a で始まる部分列: $S \to aB$, $B \to b$, $B \to aBB$ で最左導出
   ((B の個数) = (a の個数) − (b の個数))

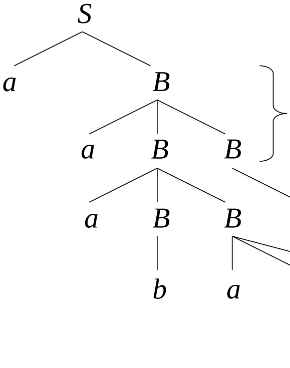
   b で始まる部分列: $S \to bA$, $A \to a$, $A \to bAA$ で最左導出
   ((A の個数) = (b の個数) − (a の個数))

導出木

定義 3 (導出木)

cfg と, ある導出が与えられると決まる木.

ある変数 (図では S) を根として, 葉を左から右に読むと, 導出された列 (図では $aaababbb$) となるもの



生成規則
$B \rightarrow aBB$ に対応

注意

複数の生成規則が適用できるとき,
導出木には適用順序の情報なし
$\Longrightarrow$ 定理 2 の証明になっている

文脈自由文法と正規言語

文脈自由文法: 正規言語ではない言語を生成可能

例: $\{0^n 1^n \mid n \geq 0\}$

すべての正規言語は文脈自由言語か?

Yes なら「文脈自由言語が正規言語を真に含むこと」がいえる

正規言語と文脈自由言語の関係 (1/2)

定理 3

L が正規言語なら, L は文脈自由言語

証明

$(\{s_0, \ldots, s_n\}, \Sigma, \delta, s_0, F)$: L を認識する dfa

- $V = \{S_0, \ldots, S_n\}$
- $P = \{S_i \to aS_j \mid s_i \text{ から } s_j \text{ へ } a \text{ による遷移がある }\}$
  $\cup \{S_i \to \varepsilon \mid s_i \in F\}$

$\Longrightarrow$ cfg $G = (V, \Sigma, P, S_0)$ は L を生成 ∎

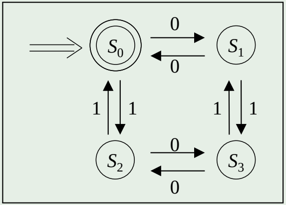
正規言語と文脈自由言語の関係 (2/2)

例

(cfg $G = (\{S_0, S_1, S_2, S_3\}, \{0, 1\}, P, S_0)$ が生成する言語)
$=$ (図の fa M が認識する言語)

ただし,

$$
\begin{aligned}
P = \{&S_0 \to 0S_1,\ S_0 \to 1S_2,\\
&S_1 \to 0S_0,\ S_1 \to 1S_3,\\
&S_2 \to 0S_3,\ S_2 \to 1S_0,\\
&S_3 \to 0S_2,\ S_3 \to 1S_1,\\
&S_0 \to \varepsilon\}
\end{aligned}
$$



G による導出 $S_0 \Rightarrow 0S_1 \Rightarrow 01S_3 \Rightarrow 010S_2 \Rightarrow 0101S_0 \Rightarrow 0101$

$\overset{\text{対応}}{\Longleftrightarrow}$ M が 0101 を読んで s_0 から s_1, s_3, s_2, s_0 と遷移して受理