文脈自由文法

離散数学・オートマトン
2022 年後期
佐賀大学理工学部　只木進一

1. **形式文法**: Formal grammars
2. **正規文法**: Regular Grammar
3. **文脈自由文法**: Context Free Grammar
4. **文脈自由言語 L を受理する** NPDA: NPDA Accepting CFL L
5. **空スタックで受理する** NPDA **に対応する文脈自由文法**: CFL corresponding to NPDA

言語と文法: Languages and grammars

- 言語の構成要素
  - 文法
  - 語
  - 文
- 文法
  - 語の配置規則
  - 文の生成規則
  - 生成文法: Generative Grammars
    N. Chomsky, *Syntactic Structure*, 1957.
- 有限オートマトン、プッシュダウンオートマトンの受理言語に対応する文法とは

形式文法: Formal Grammar

文法の一般的定式化

$$G = \langle N, \Sigma, P, S_0 \rangle \tag{1.1}$$

- N: 非終端アルファベット: 文法の要素（品詞など）に相当
- Σ: アルファベット: 語に相当
- P: 生成規則
- $S_0 \in N$: 開始記号

生成文法 (Generative Grammars) とも言う

- 開始記号から終端記号の列を生成

正規文法: Regular Grammar

- 正規表現に対応した正規言語を生成
- 生成規則

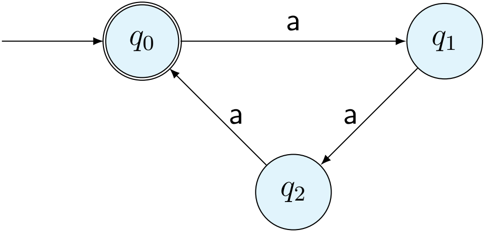
$$P : N \to \Sigma N | \Sigma \tag{2.1}$$

  ただし、$S_0 \to \epsilon$ も許す
- 記号 | は、or を表す

例 2.1:

$$N = \{S_0, S_1, S_2\}, \Sigma = \{\mathsf{a}\},$$
$$P = \{S_0 \to \epsilon | \mathsf{a}S_1, S_1 \to \mathsf{a}S_2, S_2 \to \mathsf{a}S_0 | \mathsf{a}\}$$

$$S_0 \Rightarrow \mathsf{a}S_1 \Rightarrow \mathsf{aa}S_2 \Rightarrow \mathsf{aaa}S_0$$
$$\Rightarrow \mathsf{aaaa}S_1 \Rightarrow \mathsf{aaaaa}S_2 \Rightarrow \mathsf{aaaaaa}$$



$$L = \{a^{3n} | n \in N \cup \{0\}\} = (\mathbf{aaa})^*$$

例 2.2:

$$N = \{S, A\}, \Sigma = \{\mathsf{a}, \mathsf{b}\}$$
$$P = \{S \to \epsilon \mid \mathsf{a}A, A \to \mathsf{b}S\}$$

正規文法が正規表現と同等であること

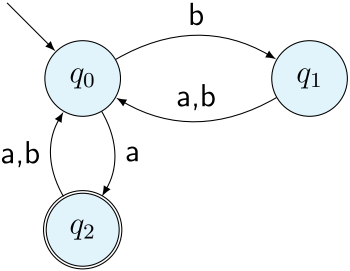
- 決定性有限オートマトンを正規文法に翻訳できること
  - 遷移関数を生成規則に翻訳
- 正規文法を非決定性有限オートマトンに翻訳できること
  - 生成規則を遷移関数に翻訳

決定性有限オートマトンから正規文法へ

Algorithm 1 $M = \langle Q, \Sigma, \delta, q_0, F \rangle \Rightarrow G = \langle N, \Sigma, P, q_0 \rangle$

```
N = Q
for all q′ = δ(q, a) do                         ▷ 遷移関数の各場合を変換
    P に q → aq′ を追加
    if q′ = δ(q, a) ∈ F then    ▷ 終状態への変換の場合は、終端記号へ
        P に q → a を追加
    end if
end for
```

例 2.3:



- 正規表現

$$((\mathsf{a}+\mathsf{b})(\mathsf{a}+\mathsf{b}))^*\,\mathsf{a}$$

- 対応する正規文法の非終端記号と終端記号

$$N = \{q_0, q_2, q_1\}$$
$$\Sigma = \{\mathsf{a}, \mathsf{b}\}$$

遷移関数	生成規則
$\delta(q_0, \mathsf{a}) = q_2$	$q_0 \to \mathsf{a}q_2 \mid \mathsf{a}$
$\delta(q_0, \mathsf{b}) = q_1$	$q_0 \to \mathsf{b}q_1$
$\delta(q_1, \mathsf{a}) = q_0$	$q_1 \to \mathsf{a}q_0$
$\delta(q_1, \mathsf{b}) = q_0$	$q_1 \to \mathsf{b}q_0$
$\delta(q_2, \mathsf{a}) = q_0$	$q_2 \to \mathsf{a}q_0$
$\delta(q_2, \mathsf{b}) = q_0$	$q_2 \to \mathsf{b}q_0$

以上まとめて

$$P = \{q_0 \to \mathsf{a}q_2|\mathsf{b}q_1|\mathsf{a},\ q_1 \to \mathsf{a}q_0|\mathsf{b}q_0,\ q_2 \to \mathsf{a}q_0|\mathsf{b}q_0\}$$

導出例

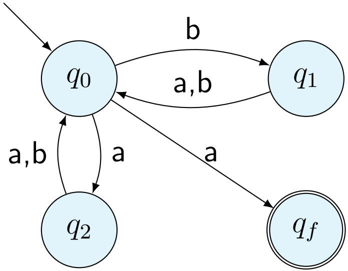
$$\begin{aligned} q_0 &\Rightarrow \mathsf{a}q_2 \Rightarrow \mathsf{ab}q_0 \Rightarrow \mathsf{abb}q_1 \\ &= \mathsf{abba}q_0 \Rightarrow \mathsf{abbaa} \end{aligned}$$

正規文法から非決定性有限オートマトンへ

Algorithm 2 $G = \langle N, \Sigma, P, S_0 \rangle \Rightarrow M = \langle Q, \Sigma, \delta, S_0, \rangle$

$Q = N$ ▷ 非終端記号を内部状態に読み替える
$F = \{q_f\}$
for all $A \to aB$ **do** ▷ 非終端記号を含む場合
  $B \in \delta(A, a)$
end for
for all $A \to a$ **do** ▷ 終端記号のみの場合は終状態へ
  $q_f \in \delta(A, a)$
end for
if $S_0 \to \epsilon \in P$ **then**
  $q_f \in \delta(S_0.\epsilon)$
end if

$$N = \{q_0, q_2, q_1\}$$
$$\Sigma = \{\mathsf{a}, \mathsf{b}\}$$
$$P = \{q_0 \to \mathsf{a}q_2|\mathsf{a}|\mathsf{b}q_1, q_1 \to \mathsf{a}q_0|\mathsf{b}q_0, q_2 \to \mathsf{a}q_0|\mathsf{b}q_0\}$$



文脈自由文法: Context Free Grammar

- 生成規則

$$P : N \to (\Sigma \cup N)^* \tag{3.1}$$

- 例

$$N = \{S_0\}, \ \Sigma = \{\mathsf{a}, \mathsf{b}\}$$
$$P = \{S_0 \to \epsilon | \mathsf{a} S_0 \mathsf{b}\}$$
$$S_0 \Rightarrow \mathsf{a} S_0 \mathsf{b} \Rightarrow \mathsf{aa} S_0 \mathsf{bb} \Rightarrow \mathsf{aaa} S_0 \mathsf{bbb} \Rightarrow \mathsf{aaabbb}$$

なぜ「文脈自由」なのか

- 生成規則の左辺は、非終端記号一つ
- 非終端記号や終端記号との繋がり（文脈）を無視

二つの標準形

- チョムスキー標準形 (Chomsky normal form, CNF)
  - $A \to BC (B, C \in N)$
  - $A \to a (a \in \Sigma)$
  - $S \to \epsilon$ も可
- グライバッハ標準形 (Greibach normal form, GNF)
  - $A \to a\alpha \ (a \in \Sigma, \alpha \in N^*)$
  - $S \to \epsilon$ も可

PDA の 3 種類の受理

- 入力終了時にスタックが空

$$L_A(M) = \{w \in \Sigma * \mid (q_0, w, Z) \vdash^* (q, \epsilon, \epsilon)\} \tag{4.1}$$

- 入力終了時に終状態

$$L_A(M) = \{w \in \Sigma * \mid (q_0, w, Z) \vdash^* (q, \epsilon, \gamma), q \in F\} \tag{4.2}$$

- 入力終了時にスタックが空、かつ終状態

$$L_A(M) = \{w \in \Sigma * \mid (q_0, w, Z) \vdash^* (q, \epsilon, \epsilon), q \in F\} \tag{4.3}$$

文脈自由言語 L を受理する NPDA

$$G = \langle N, \Sigma, P, S \rangle \tag{4.4}$$

- $L \cap \{\epsilon\} = \emptyset$ とする: 明示的に ϵ は受理する場合を除く
- 生成規則は GNF (Greibach 標準形)
- 最左導出 (一番左の非終端記号から生成規則を適用)
- 等価な NPDA: 入力終了時にスタックが空になる

$$M = \langle \{q\}, \Sigma, N, \delta, q, S, \emptyset \rangle \tag{4.5}$$

- 最左導出

$$S \Rightarrow a_1 A_1 \gamma_1 \Rightarrow a_1 a_2 A_2 \gamma_2 \Rightarrow^* a_1 a_2 \cdots a_{n-1} A_{n-1} \Rightarrow a_1 a_2 \cdots a_n$$

- 対応する動作

$$\begin{aligned}(q, a_1 a_2 \cdots a_n, S) &\vdash (q, a_2 \cdots a_n, A_1 \gamma_1)\\ &\vdash (q, a_3 \cdots a_n, A_2 \gamma_2)\\ &\cdots\\ &\vdash (q, a_n, A_{n-1})\\ &\vdash (q, \epsilon, \epsilon)\end{aligned}$$

- 遷移関数
  - 生成規則 $A \to a\gamma$ があり、かつその限り

$$(q, \gamma) \in \delta(q, a, A)$$

例 4.1:

$$G = \langle \{S, A, B\}, \{\mathsf{a}, \mathsf{b}\}, P, S \rangle$$
$$P = \{S \to \mathsf{a}|\mathsf{b}|\mathsf{a}SA|\mathsf{b}SB, A \to \mathsf{a}, B \to \mathsf{b}\}$$

$$S \Rightarrow \mathsf{a}SA \Rightarrow \mathsf{ab}SBA \Rightarrow \mathsf{aba}SABA \Rightarrow \mathsf{abaa}ABA$$
$$\Rightarrow \mathsf{abaaa}BA \Rightarrow \mathsf{abaaab}A \Rightarrow \mathsf{abaaaba}$$

$$M = \langle \{q\}, \{\mathsf{a}, \mathsf{b}\}, \{S, A, B\}, \delta, S, \emptyset \rangle$$

$\delta(q, \mathsf{a}, S) = \{(q, \epsilon), (q, SA)\}$ $S \to \mathsf{a} \mid \mathsf{a}SA$ より

$\delta(q, \mathsf{b}, S) = \{(q, \epsilon), (q, SB)\}$ $S \to \mathsf{b} \mid \mathsf{b}SB$ より

$\delta(q, \mathsf{a}, A) = \{(q, \epsilon)\}$ $A \to \mathsf{a}$ より

$\delta(q, \mathsf{b}, B) = \{(q, \epsilon)\}$ $S \to \mathsf{b}$ より

$$
\begin{aligned}
(q, \text{abaaaba}, S) &\vdash (q, \text{baaaba}, SA) \\
&\vdash (q, \text{aaaba}, SBA) \\
&\vdash (q, \text{aaba}, SABA) \\
&\vdash (q, \text{aba}, ABA) \\
&\vdash (q, \text{ba}, BA) \\
&\vdash (q, \text{a}, A) \\
&\vdash (q, \epsilon, \epsilon)
\end{aligned}
$$

空スタックで受理する NPDA に対応する文脈自由文法

$$M = \langle Q, \Sigma, \Gamma, \delta, q_0, Z, \emptyset \rangle$$

$$G = \langle N, \Sigma, P, S \rangle$$

$q, q' \in Q, A \in \Gamma$に対して $[qAq'] \in N$

- $\forall q \in Q$ に対して $S \to [q_0 Z q]$ を作る
- $(q_1, B_1 \cdots B_k) \in \delta(q, a, A)$ に対して
  - $\forall q_2, \cdots, q_{k+1}$ に対して

$$[qAq_{k+1}] \to [q_1 B_1 q_2][q_2 B_2 q_3] \cdots [q_k B_k q_{k+1}]$$

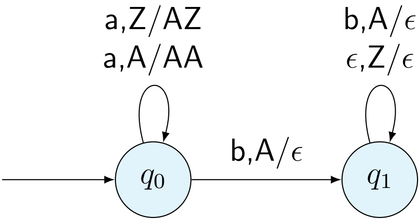
  - ただし $(q_1, \epsilon) \in \delta(q, a, A)$ に対しては

$$[qAq_1] \to a$$

例 5.1:

$$M = \langle \{q_0, q_1\}, \{\mathsf{a}, \mathsf{b}\}, \{A, Z\}, \delta, q_0, Z, \emptyset \rangle$$

$$\begin{aligned}
&\delta(q_0, \mathsf{a}, Z) = \{(q_0, AZ)\}, && \delta(q_0, \mathsf{a}, A) = \{(q_0, AA)\}, \\
&\delta(q_0, \mathsf{b}, A) = \{(q_1, \epsilon)\}, && \\
&\delta(q_1, \mathsf{b}, A) = \{(q_1, \epsilon)\}, && \delta(q_1, \epsilon, Z) = \{(q_1, \epsilon)\}.
\end{aligned}$$



$$
\begin{aligned}
(q_0, \mathsf{aaabbb}, Z) &\vdash (q_0, \mathsf{aabbb}, \mathsf{AZ}) \\
&\vdash (q_0, \mathsf{abbb}, \mathsf{AAZ}) \\
&\vdash (q_0, \mathsf{bbb}, \mathsf{AAAZ}) \\
&\vdash (q_1, \mathsf{bb}, \mathsf{AAZ}) \\
&\vdash (q_1, \mathsf{b}, \mathsf{AZ}) \\
&\vdash (q_1, \epsilon, Z) \\
&\vdash (q_1, \epsilon, \epsilon)
\end{aligned}
$$

対応する CFG を構成

$$G = \langle N, \{\mathsf{a}, \mathsf{b}\}, P, S \rangle$$

- 開始記号

$$S \to [q_0 Z q_0] \mid [q_0 Z q_1]$$

- $\delta(q_0, \mathsf{a}, Z) = \{(q_0, \mathsf{AZ})\}$ より

$$[q_0 Z q_0] \to \mathsf{a}\,[q_0 A q_0]\,[q_0 Z q_0] \mid \mathsf{a}\,[q_0 A q_1]\,[q_1 Z q_0]$$
$$[q_0 Z q_1] \to \mathsf{a}\,[q_0 A q_0]\,[q_0 Z q_1] \mid \mathsf{a}\,[q_0 A q_1]\,[q_1 Z q_1]$$

- $\delta\left(q_0, \mathsf{a}, A\right) = \{(q_0, \mathsf{AA})\}$ より

$$
\begin{aligned}
\left[q_0 A q_0\right] &\to \mathsf{a}\left[q_0 A q_0\right]\left[q_0 A q_0\right] | \mathsf{a}\left[q_0 A q_1\right]\left[q_1 A q_0\right] \\
\left[q_0 A q_1\right] &\to \mathsf{a}\left[q_0 A q_0\right]\left[q_0 A q_1\right] | \mathsf{a}\left[q_0 A q_1\right]\left[q_1 A q_1\right]
\end{aligned}
$$

- $\delta\left(q_0, B, A\right) = \{(q_1, \epsilon)\}$ より

$$[q_0 A q_1] \to \mathsf{b}$$

- $\delta\left(q_1, B, A\right) = \{(q_1, \epsilon)\}$ より

$$[q_1 A q_1] \to \mathsf{b}$$

- $\delta\left(q_1, \epsilon, Z\right) = \{(q_1, \epsilon)\}$ より

$$[q_1 Z q_1] \to \epsilon$$

生成規則: 暫定

$$
\begin{aligned}
S &\to [q_0 Z q_0] \mid [q_0 Z q_1] \\
[q_0 Z q_0] &\to \mathsf{a}\,[q_0 A q_0]\,[q_0 Z q_0] \mid \mathsf{a}\,[q_0 A q_1]\,[q_1 Z q_0] \\
[q_0 Z q_1] &\to \mathsf{a}\,[q_0 A q_0]\,[q_0 Z q_1] \mid \mathsf{a}\,[q_0 A q_1]\,[q_1 Z q_1] \\
[q_0 A q_0] &\to \mathsf{a}\,[q_0 A q_0]\,[q_0 A q_0] \mid \mathsf{a}\,[q_0 A q_1]\,[q_1 A q_0] \\
[q_0 A q_1] &\to \mathsf{a}\,[q_0 A q_0]\,[q_0 A q_1] \mid \mathsf{a}\,[q_0 A q_1]\,[q_1 A q_1] \mid \mathsf{b} \\
[q_1 A q_1] &\to \mathsf{b} \\
[q_1 Z q_1] &\to \epsilon
\end{aligned}
$$

- 終端記号を導かない N の要素

$$[q_0 Z q_0]\,,\,[q_1 Z q_0]\,,\,[q_0 A q_0]\,,\,[q_1 A q_0]$$

生成規則

$$
\begin{aligned}
S &\rightarrow [q_0 Z q_1] \\
[q_0 Z q_1] &\rightarrow \mathsf{a}\, [q_0 A q_1]\, [q_1 Z q_1] \\
[q_0 A q_1] &\rightarrow \mathsf{a}\, [q_0 A q_1]\, [q_1 A q_1] \,|\mathsf{b} \\
[q_1 A q_1] &\rightarrow \mathsf{b} \\
[q_1 Z q_1] &\rightarrow \epsilon
\end{aligned}
$$

導出例

$$
\begin{aligned}
S &\Rightarrow [q_0 Z q_1] \\
&\Rightarrow \mathsf{a}\,[q_0 A q_1]\,[q_1 Z q_1] \\
&\Rightarrow \mathsf{aa}\,[q_0 A q_1]\,[q_1 A q_1]\,[q_1 Z q_1] \\
&\Rightarrow \mathsf{aaa}\,[q_0 A q_1]\,[q_1 A q_1]\,[q_1 A q_1]\,[q_1 Z q_1] \\
&\Rightarrow \mathsf{aaab}\,[q_1 A q_1]\,[q_1 A q_1]\,[q_1 Z q_1] \\
&\Rightarrow \mathsf{aaabb}\,[q_1 A q_1]\,[q_1 Z q_1] \\
&\Rightarrow \mathsf{aaabbb}\,[q_1 Z q_1] \\
&\Rightarrow \mathsf{aaabbb}
\end{aligned}
$$