

文脈自由文法

離散数学・オートマトン

2024 年後期

佐賀大学工学部 只木進一

- 1 形式文法: Formal grammars
- 2 正規文法: Regular Grammars
- 3 文脈自由文法: Context Free Grammars
- 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: Elements of languages
 - 文法: Grammars
 - 語: Words
 - 文: Sentences
- 文法: Grammars
 - 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: General formalization of grammars

$$G = \langle N, \Sigma, P, S_0 \rangle \quad (1.1)$$

- 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 Grammars

- 正規表現に対応した正規言語を生成
- 生成規則: Productions

$$P : N \rightarrow \Sigma N | \Sigma \quad (2.1)$$

- ただし、必要ならば $S_0 \rightarrow \epsilon$ も許容する
- 非終端記号が、一つの終端記号、または一つの終端記号と一つの非終端記号の列に置き換わる
- 記号 $|$ は、or を表す
- 右辺に ΣN があることで、無限回の繰り返しが可能となる。

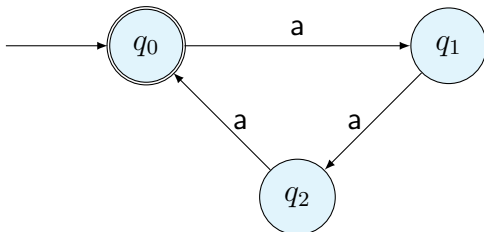
$$A \rightarrow aA | a \quad (2.2)$$

例 2.1:

$$N = \{S_0, S_1, S_2\}, \quad \Sigma = \{a\},$$

$$P = \{S_0 \rightarrow \epsilon | aS_1, S_1 \rightarrow aS_2, S_2 \rightarrow aS_0 | a\}$$

$$\begin{aligned} S_0 &\Rightarrow aS_1 \Rightarrow aaS_2 \Rightarrow aaaS_0 \\ &\Rightarrow aaaaS_1 \Rightarrow aaaaaS_2 \Rightarrow aaaaaa \end{aligned}$$



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

例 2.2:

$$N = \{S, A\}, \Sigma = \{a, b\}$$

$$P = \{S \rightarrow \epsilon \mid aA, A \rightarrow bS\}$$

正規文法が正規表現と同等であること: Regular grammars are equivalent to regular expressions

- 決定性有限オートマトンを正規文法に翻訳できること: DFAs can be translated to regular grammars
 - 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: Regular grammars can be translated to NFAs
 - 生成規則を遷移関数に翻訳
- 一般的な翻訳規則があることが重要: General translation rules are important

決定性有限オートマトンから正規文法へ: From DFA to regular grammar

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

$N = Q$

▷ 内部状態を非終端記号に読み替える

for all $q' = \delta(q, a)$ **do**

▷ 遷移関数の各場合を変換

P に $q \rightarrow aq'$ を追加

if $q' \in F$ **then**

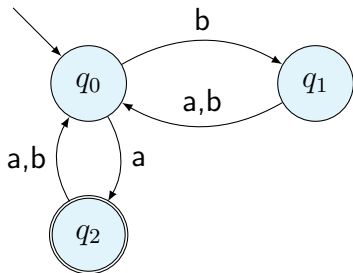
▷ 終状態への遷移の場合は、終端記号のみへ

P に $q \rightarrow a$ を追加

end if

end for

例 2.3:



- 正規表現

$$((a + b)(a + b))^* a$$

- 対応する正規文法の非終端記号と終端記号

$$N = \{q_0, q_2, q_1\}$$

$$\Sigma = \{a, b\}$$

遷移関数	生成規則
$\delta(q_0, a) = q_2$	$q_0 \rightarrow aq_2 a$
$\delta(q_0, b) = q_1$	$q_0 \rightarrow bq_1$
$\delta(q_1, a) = q_0$	$q_1 \rightarrow aq_0$
$\delta(q_1, b) = q_0$	$q_1 \rightarrow bq_0$
$\delta(q_2, a) = q_0$	$q_2 \rightarrow aq_0$
$\delta(q_2, b) = q_0$	$q_2 \rightarrow bq_0$

以上まとめて

$$P = \{q_0 \rightarrow aq_2 | bq_1 | a, q_1 \rightarrow aq_0 | bq_0, q_2 \rightarrow aq_0 | bq_0\}$$

導出例

$$\begin{aligned}
 q_0 &\Rightarrow aq_2 \Rightarrow abq_0 \Rightarrow abbq_1 \\
 &= abbaq_0 \Rightarrow abbaa
 \end{aligned}$$

正規文法から非決定性有限オートマトンへ:

From regular grammar to NFA

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 $A \rightarrow aB$ **do**

▷ 非終端記号を含む場合

$B \in \delta(A, a)$

end for

for all $A \rightarrow a$ **do**

▷ 終端記号のみの場合は終状態へ

$q_f \in \delta(A, a)$

end for

if $S_0 \rightarrow \epsilon \in P$ **then**

$q_f \in \delta(S_0, \epsilon)$

end if

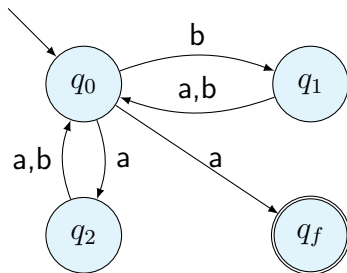
例 2.4:

$$N = \{q_0, q_2, q_1\}$$

$$\Sigma = \{a, b\}$$

$$P = \{q_0 \rightarrow aq_2 | a|bq_1, q_1 \rightarrow aq_0 | bq_0, q_2 \rightarrow aq_0 | bq_0\}$$

CFG	NFA
$q_0 \rightarrow aq_2 a bq_1$	$\delta(q_0, a) = \{q_1, q_f\}$ $\delta(q_0, b) = \{q_2\}$
$q_1 \rightarrow aq_0 bq_0$	$\delta(q_1, a) = \{q_0\}$ $\delta(q_1, b) = \{q_0\}$
$q_2 \rightarrow aq_0 bq_0$	$\delta(q_2, a) = \{q_0\}$ $\delta(q_2, b) = \{q_0\}$



文脈自由文法: Context Free Grammars

- 生成規則: 非終端記号が非終端記号または終端記号の長さ 0 以上の列へと変換

$$P : N \rightarrow (\Sigma \cup N)^* \quad (3.1)$$

- 例

$$N = \{S_0\}, \Sigma = \{a, b\}$$

$$P = \{S_0 \rightarrow \epsilon | aS_0b\}$$

$$S_0 \Rightarrow aS_0b \Rightarrow aaS_0bb \Rightarrow aaaS_0bbb \Rightarrow aaabbbb$$

なぜ「文脈自由」なのか: Why they are called context free

- 生成規則の左辺は、非終端記号一つ
- 前後の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: Standard forms

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

PDA の 3 種類の受理

- 入力終了時にスタックが空

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

- 入力終了時に終状態

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

- 入力終了時にスタックが空、かつ終状態

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

文脈自由言語 L を受理する NPDA: NPDA Accepting CFL L

$$G = \langle N, \Sigma, P, S \rangle \quad (4.4)$$

- $L \cap \{\epsilon\} = \emptyset$ とする: 明示的に ϵ を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): leftmost derivation
- 等価な NPDA: 入力終了時にスタックが空になる

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

- 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

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

- 遷移関数

- 生成規則 $A \rightarrow a\gamma$ があり、かつその限り

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

例 4.1:

$$G = \langle \{S, A, B\}, \{a, b\}, P, S \rangle$$

$$P = \{S \rightarrow a|b|aSA|bSB, A \rightarrow a, B \rightarrow b\}$$

$$\begin{aligned} S &\Rightarrow aSA \Rightarrow abSBA \Rightarrow abaSABA \Rightarrow abaaABA \\ &\Rightarrow abaaaBA \Rightarrow abaaabA \Rightarrow abaaaba \end{aligned}$$

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

$$\delta(q, a, S) = \{(q, \epsilon), (q, SA)\}$$

$$S \rightarrow a \mid aSA \text{より}$$

$$\delta(q, b, S) = \{(q, \epsilon), (q, SB)\}$$

$$S \rightarrow b \mid bSB \text{より}$$

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

$$A \rightarrow a \text{より}$$

$$\delta(q, b, B) = \{(q, \epsilon)\}$$

$$B \rightarrow b \text{より}$$

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

空スタックで受理する NPDA に対応する 文脈自由文法

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

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

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

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

$$[q A q_{k+1}] \rightarrow a [q_1 B_1 q_2] [q_2 B_2 q_3] \cdots [q_k B_k q_{k+1}]$$

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

$$[q A q_1] \rightarrow a$$

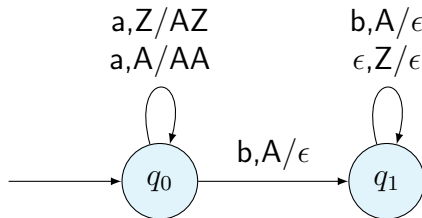
例 5.1:

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

$$\delta(q_0, a, Z) = \{(q_0, AZ)\}, \quad \delta(q_0, a, A) = \{(q_0, AA)\},$$

$$\delta(q_0, b, A) = \{(q_1, \epsilon)\},$$

$$\delta(q_1, b, A) = \{(q_1, \epsilon)\}, \quad \delta(q_1, \epsilon, Z) = \{(q_1, \epsilon)\}.$$



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

対応する CFG を構成

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

- 開始記号

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

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

$$[q_0 Z q_0] \rightarrow a [q_0 A q_0] [q_0 Z q_0] \mid a [q_0 A q_1] [q_1 Z q_0]$$

$$[q_0 Z q_1] \rightarrow a [q_0 A q_0] [q_0 Z q_1] \mid a [q_0 A q_1] [q_1 Z q_1]$$

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

$$[q_0 A q_0] \rightarrow a [q_0 A q_0] [q_0 A q_0] \mid a [q_0 A q_1] [q_1 A q_0]$$

$$[q_0 A q_1] \rightarrow a [q_0 A q_0] [q_0 A q_1] \mid a [q_0 A q_1] [q_1 A q_1]$$

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

$$[q_0 A q_1] \rightarrow b$$

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

$$[q_1 A q_1] \rightarrow b$$

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

$$[q_1 Z q_1] \rightarrow \epsilon$$

生成規則: 暫定

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

$$[q_0 Z q_0] \rightarrow \mathbf{a} [q_0 A q_0] [q_0 Z q_0] \mid \mathbf{a} [q_0 A q_1] [q_1 Z q_0]$$

$$[q_0 Z q_1] \rightarrow \mathbf{a} [q_0 A q_0] [q_0 Z q_1] \mid \mathbf{a} [q_0 A q_1] [q_1 Z q_1]$$

$$[q_0 A q_0] \rightarrow \mathbf{a} [q_0 A q_0] [q_0 A q_0] \mid \mathbf{a} [q_0 A q_1] [q_1 A q_0]$$

$$[q_0 A q_1] \rightarrow \mathbf{a} [q_0 A q_0] [q_0 A q_1] \mid \mathbf{a} [q_0 A q_1] [q_1 A q_1] \mid \mathbf{b}$$

$$[q_1 A q_1] \rightarrow \mathbf{b}$$

$$[q_1 Z q_1] \rightarrow \epsilon$$

生成規則から終端記号に至る非終端記号を探す

- 直接終端記号に至る左辺記号探す

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

- 上記がわかった上で終端記号に至る左辺記号探す

$$[q_0 Z q_1]$$

- 上記がわかった上で終端記号に至る左辺記号探す

$$S$$

- 上記がわかった上で終端記号に至る左辺記号探すが、新たな記号が見つからない

終端記号を導く N の要素

- 終端記号を導く N の要素

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

- 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

$$S \rightarrow [q_0 Z q_1]$$

$$[q_0 Z q_1] \rightarrow \mathbf{a} [q_0 A q_1] [q_1 Z q_1]$$

$$[q_0 A q_1] \rightarrow \mathbf{a} [q_0 A q_1] [q_1 A q_1] \mid \mathbf{b}$$

$$[q_1 A q_1] \rightarrow \mathbf{b}$$

$$[q_1 Z q_1] \rightarrow \epsilon$$

導出例

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