KYUSHU UNIVERSITY 100th 2011
知の新世紀を拓く

形式言語理論

Formal Language Theory

10.　文脈自由文法の標準形

九州大学
KYUSHU UNIVERSITY

本日の内容

- 文脈自由文法の簡略化
- Chomsky標準形
- Greibach標準形

文脈自由文法の簡略化

無用な記号(1)

- 右のCFGにおいて，
  記号 C から始まる導出
  $$C \underset{G}{\Rightarrow}^{*} Ca \underset{G}{\Rightarrow}^{*} Caa \underset{G}{\Rightarrow}^{*} \cdots$$
  は終端記号列を産み出さない．
- よって，C を含む生成規則
  $$B \to cCD$$
  $$C \to Ca,$$
  を削除してよい．
- これにより，終端記号 c も
  削除される．

$$\begin{aligned} G &= (N, \Sigma, P, S) \\ &\quad N = \{S, A, B, C, D\} \\ &\quad \Sigma = \{a, b, c, d\} \\ &\quad P = \{S \to aAB, \\ &\qquad\quad A \to aBB, \\ &\qquad\quad B \to ab \mid cCD, \\ &\qquad\quad C \to Ca, \\ &\qquad\quad D \to d\} \end{aligned}$$

$$\Downarrow$$

$$\begin{aligned} G &= (N, \Sigma, P, S) \\ &\quad N = \{S, A, B, D\} \\ &\quad \Sigma = \{a, b, d\} \\ &\quad P = \{S \to aAB, \\ &\qquad\quad A \to aBB, \\ &\qquad\quad B \to ab, \\ &\qquad\quad D \to d\} \end{aligned}$$

無用な記号(2)

- 右のCFGにおいて，開始記号 S から始まるいかなる導出も記号 D を含まない．
- 終端記号 d も同様である．
- 記号 D, d を含む生成規則
$$D \rightarrow d$$
を削除してよい．

$$
\begin{aligned}
&G = (N, \Sigma, P, S) \\
&\quad N = \{S, A, B, D\} \\
&\quad \Sigma = \{a, b, d\} \\
&\quad P = \{S \rightarrow aAB, \\
&\qquad\quad A \rightarrow aBB, \\
&\qquad\quad B \rightarrow ab, \\
&\qquad\quad D \rightarrow d\}
\end{aligned}
$$

$$\Downarrow$$

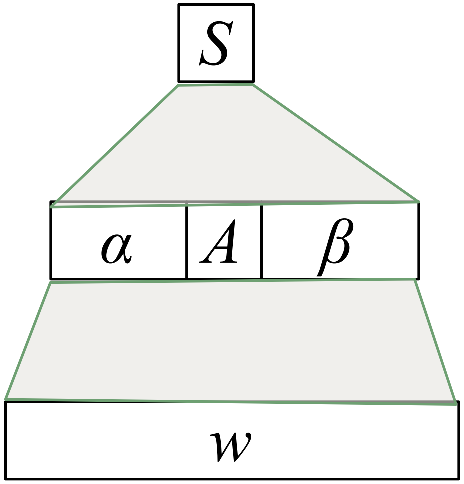
$$
\begin{aligned}
&G = (N, \Sigma, P, S) \\
&\quad N = \{S, A, B\} \\
&\quad \Sigma = \{a, b\} \\
&\quad P = \{S \rightarrow aAB, \\
&\qquad\quad A \rightarrow aBB, \\
&\qquad\quad B \rightarrow ab\}
\end{aligned}
$$

有用な記号・無用な記号

【定義】 $G = (N, \Sigma, P, S)$ を CFG とする．記号 $A \in N \cup \Sigma$ が
が**有用**(usefull)であるとは，

$$S \underset{G}{\Rightarrow}^* \alpha A \beta \underset{G}{\Rightarrow}^* w$$

となる $\alpha, \beta \in (N \cup \Sigma)^*$, $w \in \Sigma^*$ が存在するときをいう．有用でない記号は**無用**(useless)であるという．



CFGの既約性

【定義】 CFG $G = (N, \Sigma, P, S)$ が既約であるとは，$P = \varnothing$ であるか，または，すべての記号 $A \in N \cup \Sigma$ が有用であるときをいう.

【定理4.3】 任意の CFG $G = (N, \Sigma, P, S)$ に対し，
G と等価で以下を満たす既約な CFG $G' = (N', \Sigma', P', S)$ が一意に存在する.

$$N' \subseteq N,\ \Sigma' \subseteq \Sigma,\ P' \subseteq P$$

定理4.3の証明

- 無用な記号およびそれに関わる生成規則を除去すればよい.
- CFG $G = (N, \Sigma, P, S)$ に対して有用な記号の集合を U とする.
- $S \notin U$ ならば $G' = (\{S\}, \Sigma, \varnothing, S)$ とする.
- $S \in U$ ならば $G' = (N', \Sigma', P', S)$ を次で定める.
  - $N' = U \cap N$.
  - $\Sigma' = U \cap \Sigma$.
  - $P' = \{ A \to \gamma \in P \mid A \in N'$ かつ $\gamma \in (N' \cup \Sigma')^* \}$.

有用な非終端記号

有用な終端記号

有用な記号のみから成る生成規則

QED

既約なCFG構築アルゴリズム:ステップ1

- $Gen = \{A \in N \mid \exists w \in \Sigma^*, A \underset{G}{\Rightarrow}^* w\}$とおく.
- 次の命題に基づき Gen を求める.
- Gen に属さない記号を N から削除する.
- Gen に属さない記号を含む生成規則を削除する.

【命題】 CFG $G = (N, \Sigma, P, S)$ に対し，系列 $V_1, V_2, \ldots$ を漸化式

$$V_1 = \{ A \in N \mid \exists w \in \Sigma^*, A \to w \in P \}$$

$$V_{n+1} = V_n \cup \{ A \in N \mid \exists \gamma \in (V_n \cup \Sigma)^*, \ A \to \gamma \in P \}$$

によって定義する．このとき，$V_{|N|} = Gen$ が成り立つ．

既約なCFG構築アルゴリズム：ステップ2

- $Reach=\{B \in N \cup \Sigma \mid \exists \alpha, \exists \beta \in (N \cup \Sigma)^*,\ S \underset{G}{\Rightarrow}^* \alpha B \beta\}$とおく.
- 次の命題に基づき $Reach$ を求める.
- $Reach$ に属さない記号を N, Σ から削除する.
- $Reach$ に属さない記号を含む生成規則を削除する.

【命題】 CFG $G=(N, \Sigma, P, S)$ に対し，系列 $U_0, U_1, \ldots$ を漸化式

$$U_0 = \{S\}$$

$$U_{n+1} = U_n \cup \{B \in N \cup \Sigma \mid \exists A \in U_n, \exists \alpha, \exists \beta \in (N \cup \Sigma)^*, A \to \alpha B \beta \in P\}$$

によって定義する．このとき，$U_{|N|} = Reach$ が成り立つ.

注意

- ステップ2, ステップ1の順番で実行すると無用な記号が残ることがある.

$$
\begin{aligned}
G &= (N, \Sigma, P, S) \\
&\quad N = \{S, A, B, C, D\} \\
&\quad \Sigma = \{a, b, c, d\} \\
&\quad P = \{S \to aAB, \\
&\qquad\quad A \to aBB, \\
&\qquad\quad B \to ab \mid cCD, \\
&\qquad\quad C \to Ca, \\
&\qquad\quad D \to d\}
\end{aligned}
$$

ステップ2
実行後も
変わらない

A, B, C, Dは
すべて
Sから到達可能

$$
\begin{aligned}
G &= (N, \Sigma, P, S) \\
&\quad N = \{S, A, B, C, D\} \\
&\quad \Sigma = \{a, b, c, d\} \\
&\quad P = \{S \to aAB, \\
&\qquad\quad A \to aBB, \\
&\qquad\quad B \to ab \mid cCD, \\
&\qquad\quad C \to Ca, \\
&\qquad\quad D \to d\}
\end{aligned}
$$

Chomsky標準形

Chomsky標準形

【定義】 CFG $G = (N, \Sigma, P, S)$ が Chomsky標準形であるとは，生成規則が次のいずれかの形式であるときをいう．

(1) $A \rightarrow BC \quad (A, B, C \in N)$

(2) $A \rightarrow a \quad (A \in N,\ a \in \Sigma)$

与えられたCFG G に対し，
それと等価なChomsky標準形の
CFG G' を求めたい

ε-生成規則消去定理

【定理4.4】 CFG $G=(N, \Sigma, P, S)$ に対し，G と等価で生成規則が次の(1)(2)のいずれかの形式である CFG $G'=(N', \Sigma, P', S')$ を構成できる．

(1) $S' \rightarrow \varepsilon$

(2) $A \rightarrow \gamma \quad (A \in N',\ \gamma \in ((N' - \{S'\}) \cup \Sigma)^{+})$

生成規則の右辺は S' を含まない．

$G=(N, \Sigma, P, S)$

$N=\{S, A, B, C\}$

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

$P=\{S \rightarrow AB,$

$A \rightarrow aBC,$

$B \rightarrow C|b$

$C \rightarrow c|\varepsilon\ \}$

$G'=(N', \Sigma, P', S')$

$N'=N \cup \{S'\}$

$P'=\{S' \rightarrow S,$

$S \rightarrow AB|A,$

$A \rightarrow aBC|aB|aC|a,$

$B \rightarrow C|b$

$C \rightarrow c\ \}$

例

- 右のCFGにおいて ε-生成規則
  $C \to \varepsilon$
  を削除すると，以下の導出ができなくなる.
  $C \underset{G}{\Rightarrow}^{*} \varepsilon \quad B \underset{G}{\Rightarrow}^{*} \varepsilon$
- そこで，右辺に B, C を含む生成規則に対し B, C を削除した規則を加えておく.

$S \to AB$ ➡ $S \to A$ を追加

$A \to aBC$ ➡ $A \to aB|aC|a$ を追加

$$
\begin{aligned}
G &= (N, \Sigma, P, S) \\
&\quad N = \{S, A, B, C\} \\
&\quad \Sigma = \{a, b, c\} \\
&\quad P = \{S \to AB, \\
&\qquad\quad A \to aBC, \\
&\qquad\quad B \to C|b \\
&\qquad\quad C \to c|\varepsilon
\end{aligned}
$$

⬇

$$
\begin{aligned}
G' &= (N', \Sigma, P', S') \\
&\quad N' = N \cup \{S'\} \\
&\quad P' = \{S' \to S, \\
&\qquad\quad S \to AB|A, \\
&\qquad\quad A \to aBC|aB|aC|a, \\
&\qquad\quad B \to C|b \\
&\qquad\quad C \to c\ \}
\end{aligned}
$$

定理4.4の証明

- CFG $G = (N, \Sigma, P, S)$ に対して次を求める.
  - $Null = \{ A \in N \mid A \overset{*}{\underset{G}{\Rightarrow}} \varepsilon \}$
- CFG $G' = (N', \Sigma', P', S')$ を次のように定める.
  - $N' = N \cup \{S'\}$.
  - $P' = \{S' \to S\} \cup \{S' \to \varepsilon \mid \varepsilon \in L(G)\}$

$$\cup \left\{ A \to B_1 \ldots B_k \;\middle|\; \begin{array}{l} k \geq 1, \\ A \to \alpha_0 B_1 \alpha_1 \ldots \alpha_{k-1} B_k \alpha_k \in P, \\ B_1, \ldots, B_k \in N \cup \Sigma, \\ \alpha_0, \alpha_1, \ldots, \alpha_k \in Null^* \end{array} \right\}$$

QED

記号の集合 $Null$ を求める

- 次の命題に基づき $Null$ を求めることができる.

【命題】 CFG $G = (N, \Sigma, P, S)$ に対し，系列 $E_1, E_2, \ldots$ を漸化式

$$E_1 = \{A \in N \mid A \to \varepsilon \in P\}$$

$$E_{n+1} = E_n \cup \{A \in N \mid A \to \gamma \in P,\ \gamma \in E_n^{*}\}$$

によって定義する．このとき，$E_{|N|} = Null$ が成り立つ．

鎖生成規則消去定理

$A \to B \;\; (A, B \in N)$の形式の規則

【定理4.5】 CFG $G = (N, \Sigma, P, S)$ に対し，G と等価で生成規則が次の(1)-(3)のいずれかの形式である CFG $G' = (N, \Sigma, P', S)$ を構成できる．

(1) $S \to \varepsilon$

(2) $A \to \gamma \quad (A \in N, \gamma \in ((N - \{S\}) \cup \Sigma)^*, \; |\gamma| \geq 2)$

(3) $A \to a \quad (A \in N, a \in \Sigma)$

$G = (N, \Sigma, P, S)$

$N = \{S, A, B, C\}$

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

$P = \{S \to A,$
$A \to AB \mid a,$
$B \to C \mid b$
$C \to A \mid c\}$

⇒

$G' = (N, \Sigma, P', S)$

$P' = \{ S \to AB \mid a,$
$A \to AB \mid a,$
$B \to b \mid a \mid c \mid AB$
$C \to c \mid a \mid AB \}$

定理4.5の証明：アイデア

$G=(N, \Sigma, P, S)$
$N=\{S, A, B, C\}$
$\Sigma=\{a, b, c\}$
$P=\{S \to A,$
$A \to AB \mid a,$
$B \to C \mid b$
$C \to A \mid c\}$

- 右のCFGにおいて鎖状生成規則

  $S \to A,\ B \to C,\ C \to A$

  を削除すると，以下の導出ができなくなる．

  $S \underset{G}{\Rightarrow}^* A,\ \ B \underset{G}{\Rightarrow}^* C,\ \ C \underset{G}{\Rightarrow}^* A,\ \ B \underset{G}{\Rightarrow}^* A$

- そこで，鎖状生成規則を介した導出をダイレクトに行う規則を追加．

$S \underset{G}{\Rightarrow}^* A \quad A \to AB \mid a$ ➡ $S \to AB \mid a$ を追加

$B \underset{G}{\Rightarrow}^* C \quad C \to A \mid c$ ➡ $B \to c$ を追加

$C \underset{G}{\Rightarrow}^* A \quad A \to AB \mid a$ ➡ $C \to AB \mid a$ を追加

$B \underset{G}{\Rightarrow}^* A \quad A \to AB \mid a$ ➡ $B \to AB \mid a$ を追加

⬇

$G'=(N, \Sigma, P', S)$
$P'=\{S \to AB \mid a,$
$A \to AB \mid a,$
$B \to b \mid a \mid c \mid AB$
$C \to c \mid a \mid AB\}$

定理4.5の証明

- CFG $G = (N, \Sigma, P, S)$ の生成規則は定理4.4の(1)(2)を満たすと仮定する.
- 各$A \in N$ に対して次のようにおく.
  - $Chain(A) = \{ B \in N \mid A \underset{G}{\Rightarrow}^* B \}$
- CFG $G' = (N, \Sigma, P', S)$ を次のように定める.
  - $P' = \{A \rightarrow \gamma \mid \exists B \in Chain(A), B \rightarrow \gamma \in P \wedge \gamma \notin N \}$

QED

記号の集合 $Chain(A)$ を求める

- 次の命題に基づき $Chain(A)$ を求めることができる.

【命題】 CFG $G = (N, \Sigma, P, S)$ と各 $A \in N$ に対し,
系列 $F_1(A), F_2(A), \ldots$ を漸化式

$$F_1(A) = \{B \in N \mid A \to B \in P\}$$

$$F_{n+1}(A) = F_n(A) \cup \{ C \in N \mid B \in F_n(A), B \to C \in P \}$$

によって定義する. このとき, $F_{|N|}(A) = Chain(A)$ が成り立つ.

Chomsky標準形への変換

【定理4.6】任意のCFG G に対し，$L(G)-\{\varepsilon\}=L(G')$となるChomsky標準形の CFG G' を構成できる.

$$
\begin{aligned}
G &= (N, \Sigma, P, S)\\
&\quad N=\{S, A\}\\
&\quad \Sigma=\{a\}\\
&\quad P=\{S\to AaAA,\\
&\qquad\qquad A\to a\}
\end{aligned}
$$

$\Rightarrow$

$$
\begin{aligned}
G' &= (N, \Sigma, P', S)\\
&\quad N' = \{S, A, B_a, Y_1, Y_2\}\\
&\quad P'=\{ S\to AY_1,\\
&\qquad\qquad Y_1\to B_aY_2,\\
&\qquad\qquad Y_2\to AA,\\
&\qquad\qquad B_a\to a,\\
&\qquad\qquad A\to a\}
\end{aligned}
$$

定理4.6の証明(1/2)

- 定理4.5より，CFG $G=(N,\Sigma,P,S)$ の生成規則は，次のいずれかの形式と仮定できる．
  - ① $S \to \varepsilon$
  - ② $A \to \gamma \quad (A \in N, \gamma \in ((N-\{S\}) \cup \Sigma)^*,\ |\gamma| \geq 2)$
  - ③ $A \to a \quad (A \in N,\ a \in \Sigma)$
- ②のみを考慮すればよい．
  - ①は言語 $L(G)-\{\varepsilon\}$ の生成に関与しない．
  - ③は Chomsky標準形の制約を満たす．
- ②の右辺 γ は $\gamma \in (N-\{S\})^*$ としてよい．
  - 右辺 γ に現れる終端記号 a を新たな非終端記号 B_a で置換し，生成規則 $B_a \to a$ を新たに作成すればよい．
- 生成規則 $A \to \gamma \ (\gamma \in (N-\{S\}))^*,\ |\gamma| \geq 2)$ のみ考える．

定理4.6の証明(2/2)

- 生成規則 $A \rightarrow X_1 \ldots X_m (X_i \in N - \{S\}, m > 2)$ を次の生成規則で置換する.

$$
\begin{aligned}
&A \rightarrow X_1 Y_1, \\
&Y_1 \rightarrow X_2 Y_2, \\
&\quad \cdots \\
&Y_{m-3} \rightarrow X_{m-2} Y_{m-2} \\
&Y_{m-2} \rightarrow X_{m-1} X_m
\end{aligned}
$$

ここで, $Y_1, \ldots, Y_{m-2}$ は新たに導入した非終端記号である.

QED

Greibach標準形

Greibach標準形

【定義】 CFG $G = (N, \Sigma, P, S)$ が Greibach標準形であるとは，生成規則が次の形式であるときをいう.

$$A \rightarrow a\gamma \quad (A \in N,\ a \in \Sigma,\ \gamma \in N^*)$$

与えられたCFG G に対し，
それと等価なGreibach標準形の
CFG G' を求めたい

Greibach標準形への変換

【定理4.7】 CFG G に対し, $L(G)-\{\varepsilon\}=L(G')$ となるGreibach標準形の CFG G' を構成できる.

$$
\begin{aligned}
G &= (N, \Sigma, P, S)\\
&\quad N=\{S, A, B\}\\
&\quad \Sigma=\{a, b\}\\
&\quad P=\{S\to AB,\\
&\qquad\quad A\to SB \mid a,\\
&\qquad\quad B\to b\}
\end{aligned}
$$

$$
\begin{aligned}
G' &= (N, \Sigma, P', S)\\
&\quad N'=\{S, A, B, C\}\\
&\quad P'=\{S\to aCB \mid aB,\\
&\qquad\quad A\to aC \mid a,\\
&\qquad\quad B\to b,\\
&\qquad\quad C\to bA \mid bAC\}
\end{aligned}
$$

証明は省略する