チョムスキーの標準形 (CNF)

12 月 22 日 (水) に予定していた補講は中止します。今年の授業は 12/8(水) の中間試験で最後で、来年は 1/7(金) から授業を行います。

チョムスキーの標準形 (CNF と略記)
とは、文脈自由文法 $G = (N, T, P, S)$ で生成規則は

$$
\begin{aligned}
A &\rightarrow BC, \quad A, B, C \in N, \\
A &\rightarrow a, \quad A \in N, a \in T
\end{aligned}
$$

という形のものだけであるもの

定理： ϵ を生成しないすべての文脈自由文法は生成する言語を変えずに CNF に変換できる

CNF がどのように役に立つのかはもうすこし先まで勉強するとわかります。

CNF への変換の仕方１

CFG(文脈文法自由文法) $G = (N, T, P, S)$ を $\epsilon \notin L(G)$ かつ ϵ 生成規則を含まない文法とする (ϵ 生成規則の除き方は前回やった)

ステップ１： $A \to B$ という形の生成規則 (単位生成規則) を除く

$$U(A) = \{B \in N \mid A \overset{+}{\underset{G}{\Longrightarrow}} B\}$$

$=$ A から１回以上の書き換えで導出できる非終端記号の集合

$U(A)$ の定義が教科書と少し違うので注意

例： $G = (\{S, A, B\}, \{a, b\}, P, S)$,

$$P = \{S \to A|ABA,\ A \to aA|a|B,\ B \to bB|b\}$$

$U(S) = \{A, B\}$, $U(A) = \{B\}$, $U(B) = \emptyset$

P から単位生成規則をすべて除いた後、$B \in U(A)$ となるすべての組 (A, B) について代わりに $A \to \alpha$ (但し $B \to \alpha$ かつ $\alpha \notin N$) を P に追加する

例： 上の例で $A \in U(S)$ なので、

$$S \to aA|a$$

を追加。他にも追加すべき規則を加えると

$$\begin{array}{rcl} S & \to & aA|a|bB|b|ABA, \\ A & \to & aA|a|bB|b, \\ B & \to & bB|b \end{array}$$

になる

CNFへの変換の仕方 2

ステップ 2：右辺の長さが 2 以上でかつ終端記号を含む生成規則の変形
$A \to \alpha$ をそのような規則とする
右辺が終端記号 c を含むとする
新しい非終端記号 X_c を N に (まだ追加していなければ) 追加し、$X_c \to c$ という規則を P に (まだ追加していなければ) 追加する
α の中の c を X_c に置き換える
$A \to \alpha$ の右辺に含まれる終端記号に同じ処理を行う
他の生成規則についても同じ処理を行う

例：さきほどの例に上記処理を行うと

$$
\begin{aligned}
S &\to X_aA|a|X_bB|b|ABA,\\
A &\to X_aA|a|X_bB|b,\\
B &\to X_bB|b,\\
X_a &\to a,\\
X_b &\to b
\end{aligned}
$$

CNF への変換の仕方 3

ステップ 3：右辺の長さが 3 以上の生成規則を除去
長さが 3 以上なら、右辺は非終端記号だけからなる $A \to B_1 B_2 \cdots B_n$ という生成規則があったら,
$A \to B_1 Y_2$, $Y_2 \to B_2 Y_3$, $\ldots$, $Y_{n-1} \to B_{n-1} B_n$ という規則を追加し、元の規則は P から削除する。但し Y_1, $\ldots$, Y_{n-1} は新しく追加した非終端記号である

例：唯一の右辺の長さが 3 以上の規則は $S \to ABA$ なので、$S \to AY_1$, $Y_1 \to BA$ という規則を追加して、元の規則を削除する

$$
\begin{array}{rcl}
S & \to & X_a A | a | X_b B | b | A Y_1, \\
Y_1 & \to & BA, \\
A & \to & X_a A | a | X_b B | b, \\
B & \to & X_b B | b, \\
X_a & \to & a, \\
X_b & \to & b
\end{array}
$$

上記を生成規則の集合 P として、求める文法は $(\{S, A, B, X_a, X_b, Y_1\}, \{a, b\}, P, S)$ となる

前回演習解説

問 43 $\{a^n b^{2n} \mid n = 0, 1, 2, \ldots\}$ を生成する文脈自由文法を (N, T, P, S) の形で書け

$(\{S\}, \{a, b\}, \{S \to \epsilon | aSbb\}, S)$

問 44 文法 $G = (\{S, A, B\}, \{a, b\}, P, S)$,

$$P = \{S \to AB, \quad A \to SA|BB|bB, \quad B \to b|aA|\epsilon\}$$

と等価で、$S \to \epsilon$ 以外の ϵ 生成規則を持たない文法 G' を (N, T, P, S) の形で書け

1. ϵ 生成規則を除いた生成規則の集合は $\{S \to AB$, $A \to SA|BB|bB$, $B \to b|aA\}$ である.
2. ϵ 生成記号は S, A, B である。3 回書き換えて ϵ になる S も ϵ 生成記号であることに注意
3. ϵ 生成記号を置き換えて、$S' \to \epsilon|S$ を追加して、最終的な生成規則の集合は $\{S' \to \epsilon|S$, $S \to AB|A|B$, $A \to SA|A|S|BB|B|bB|b$, $B \to b|aA|a\}$ である。この生成規則の集合を P' として、答えは $(\{S', S, A, B\}, \{a, b\}, P', S')$ である

前回演習解説 2

問 45 左から読んでも右から読んでも同じである記号列 $w \in \{a, b\}^*$ からなる言語を L_{45} とする。L_{45} が正規言語ではないことを証明せよ。

ポンプの補題を用いて、ある言語 L が正規言語ではないことを証明するには
1. 言語 L を受理する DFA の状態数 n を固定する。状態数をどのような記号 (例えば n) で表すか書かないと減点
2. 言語 L からポンプの補題が成り立たない記号列 z を選ぶ。z の長さは n 以上でないとポンプの補題を使えないから、z は n の値によって変わる記号列として定義する必要がある
3. $z = wxy$ (但し $|x| \geq 1$, $|wx| \leq n$) と分ける分け方をすべて場合分けする
4. z の各々の分け方について $wx^iy \notin L$ となる i を明示する

解答例：
1. L_{45} を受理する状態数 n の DFA が存在するとして矛盾を導く
2. $z = a^n b^{2n} a^n$ とする
3. $z = wxy$ (但し $|x| \geq 1$, $|wx| \leq n$) と分けると, ど

のような分け方をしても $x = a^m$ (但し $1 \leq m \leq n$) である
4. $wy = a^{n-m}b^{2n}a^n$ は右から読んだときと左から読んだときに違うので L_{45} に属さず矛盾が起きている

演習問題

問題 47 文脈自由文法 $G = (\{S, A, B\}, \{a, b\}, P, S\})$, $P = \{S \rightarrow \epsilon | ASB,\ A \rightarrow a | aAS,\ B \rightarrow bb | A | SbS\}$ に対して、$L(G)$ から ϵ を除いた言語を生成する文脈自由文法の CNF を (N, T, P, S) の形で書け

問題 48 アルファベット $\{a, b\}$ 上の言語 $L_{48} = \{ww \mid w \in \{a, b\}^*\}$ が正規言語ではないことを証明せよ。

問題 49 今日の授業でわかりにくい所や要望を書いて下さい