形式言語の理論

5. 文脈依存言語

Chomsky階層の4言語族

- 句構造
- 文脈依存言語
- 文脈自由言語
- 正規言語

句構造とは,

(1) $\boldsymbol{N}, \Sigma$ は $N \cap \Sigma = \emptyset$ なる空でない有限集合.

(2) $\boldsymbol{P}$ は $\alpha \rightarrow \beta$ の形をした記号列の有限集合
ただし, $\rightarrow \notin N \cup \Sigma$ で, $\alpha, \beta \in (N \cup \Sigma)^*$ かつ
α は少なくとも1個の N の要素を含む.

(3) $\boldsymbol{S} \in \boldsymbol{N}$.

5.1. 文脈依存文法とその標準形

- 文脈依存文法
- 定理5.1
  文脈依存文法　⇔　単調文法

- 補題5.2
  *n*次の単調文法は***n*-1**次の単調文法に変換できる。

- 補題5.3
  任意の単調文法は次数2の単調文法に変形できる

- 定理5.4
  任意の文脈依存文法は線形有界文法に変形できる

文脈依存文法

- 句構造文法 $\mathbf{G = (N, \Sigma, P, S)}$ のすべての生成規則が
  $\alpha A\beta \to \alpha\gamma\beta \quad (\alpha, \beta, \gamma \in (N \cup \Sigma)^*, \ \gamma \neq \varepsilon, \ A \in N)$
  の形をしているとき, $\mathbf{G}$ は文脈依存文法とよばれる.

- 文脈依存文法(context-sensitive grammar: **CSG**)は, ε を生成しないことに注意.

- 例5.1 $\mathbf{L = \{a^n b^n c^n \mid n \geq 1\}}$ を生成する **CSG**.

$$
\begin{array}{ll}
S \to a\,S\,B\,C & a\,B \to a\,b \\
S \to a\,B\,C & b\,B \to b\,b \\
C\,B \to A\,B & b\,C \to b\,c \\
A\,B \to A\,C & c\,C \to c\,c \\
A\,C \to B\,C &
\end{array}
$$

単調文法

- 句構造文法 $G = (N, \Sigma, P, S)$ が単調であるとは, G のすべての生成規則 $\alpha \to \beta \in P$ が $|\alpha| \le |\beta|$ を満たすことをいう.

- 単調文法と文脈依存文法は表現力が同等である.

- 例5.1 の は単調文法である.

定理5.1 （文脈依存文法⇔単調文法）

- 任意の言語 $L \subseteq \Sigma^*$ に対して，L が文脈依存言語であるための必要十分条件は，L が単調文法で生成されることである.

［ 証明 ］

- 文脈依存文法は単調文法なので，必要性は明らか.
- 任意の単調文法に対して，等価な文脈依存文法が存在することを示せばよい.

補題5.2 （単調文法の次数は下げられる）

- 任意の次数 n ($n \geq 3$) の単調文法 G に対して，G と等価な次数 n–1 の単調文法 G' が存在する．

次数：生成規則のうちの右辺の最大長

まず，任意の単調文法 G に対して，G と等価な単調文法 $G' = (N', \Sigma, P', S')$ で，すべての生成規則が

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

もしくは

$$\alpha \to \beta \quad (\alpha, \beta \in N'^{+})$$

の形をしているものが存在することに注意．

補題5.2の証明

- $A \to a$ の形をしていない生成規則には非終端記号しか現れないと考えて良い.
- $\alpha \to \beta$ を G の生成規則とする.
  $|\beta| \le 2$ の場合はそのまま. それ以外の場合には,
  $\alpha = A\alpha', \ \beta = BCD\beta' \quad (A, B, C, D \in N, \ \alpha', \beta' \in N^*)$
  とおける

$\alpha' = \varepsilon$ の場合

$$A \to A_1 A_2$$
$$A_1 \to B\, C$$
$$A_2 \to D\, \beta'$$

$\alpha' \ne \varepsilon \ (\alpha' = E\alpha'')$の場合

$$A\, E \to A'\, E'$$
$$A' \to B$$
$$E'\alpha'' \to C\, D\, \beta'$$

補題5.3
（任意の単調文法は次数2の単調文法に変形できる）

- 任意の単調文法 G に対して，G と等価な次数 2 の単調文法が存在する.

[証明]

補題5.2 を繰り返して得られる G が補題5.3を満たすのは明らか.

定理5.1の証明

- 補題5.3より，G は次数2の単調文法としてよい．

$$AB \rightarrow CD \quad (\mathrm{A} \neq \mathrm{C} \text{ かつ } B \neq D)$$

  の形をした生成規則を次のように置き換えて新たな単調文法 G' とする．

$$AB \rightarrow A'B$$
$$A'B \rightarrow A'D$$
$$A'D \rightarrow CD$$

- G' は文脈依存文法であり，かつ G と等価．

例5.2

- 単調文法 $G = (\{S, B, C\}, \{a, b, c\}, P, S)$

  $S \to a\,S\,B\,C$

  $S \to a\,B\,C$

  $C\,B \to B\,C$

  $a\,B \to a\,b$

  $b\,B \to b\,b$

  $b\,C \to b\,c$

  $c\,C \to c\,c$

- 例5.1の文脈依存文法はこの生成規則の $C\,B \to B\,C$ を書き換えたものである.

線形有界(黒田標準形)

- 文脈依存文法 $G = (N, \Sigma, P, S)$ が線形有界であるとは，P 中のすべての生成規則が次のいずれかの形をしていることをいう.
ただし，$A, B, C, D \in N \cup \Sigma - \{S\}$ である.

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

単調文法と同一視していることに注意

定理5.4
（任意の文脈依存文法は線形有界文法に変形できる）

- 任意の文脈依存文法 G に対し，G と等価な線形有界文法 G'が存在する．
- 線形有界文法の性質
  - 線形有界文法においては，$S \to S\ A$ の形以外の生成規則は導出される記号列の長さを変えない．
  - 開始記号 S を右辺にもつ生成規則は，$S \to S\ A$ に限られる．
  - 任意の線形有界文法 $G = (N, \Sigma, P, S)$ と任意の $\alpha, \beta \in (N \cup \Sigma)^*$ に対して，$S \overset{*}{\Rightarrow}_G \alpha S \beta$ ならば，$\alpha = \varepsilon$ かつ $\beta \in (N \cup \Sigma - \{S\})^*$ である．