

万能 CF 文法 と 文法の準同型写像

京大数理解析研 笠井琢美

定義 1 $G = (N, \Sigma, P, S)$ を CF 文法 とする。ただし

N : nonterminal symbols の集合, Σ : terminal symbols の集合

P : productions の集合, S : initial symbol

$V = N \cup \Sigma$ とおく。

$$x_0 \xrightarrow[p_1]{p_1} x_1 \xrightarrow[p_2]{p_2} \cdots \xrightarrow[p_k]{p_k} x_k$$

が G における leftmost derivation (x_{i-1} から x_i への変換に production p_i が適用される) で $\alpha = p_1 p_2 \cdots p_k \in P^*$ のとき

$$x_0 \xrightarrow[\alpha]{\alpha} x_k$$

と書く。さらに languages $L(G) \subset \Sigma^*$ と $A_0(G) \subset P^*$ を

$$L(G) = \{ w \in \Sigma^* \mid S \xrightarrow[\alpha]{\alpha} w, \alpha \in P^* \}$$

$$A_0(G) = \{ \alpha \in P^* \mid S \xrightarrow[\alpha]{\alpha} w, w \in \Sigma^* \}$$

で定義する。

注 $S \xrightarrow{\alpha} w$, $w \in \Sigma^*$ のとき α を w の left parse といい

$A_e(G)$ を G の left parse language といい (Aho & Ullman [6]).

w の left parse α は w の derivation tree とみなすことができる.
したがって $A_e(G)$ を G の derivation trees 全体からなる集合とみなすことができる.

定義 2 $G = (N, \Sigma, P, S)$ を CF 文法, $C \subseteq P^*$ とすると

$$L_e(G; C) = \{ w \in \Sigma^* \mid S \xrightarrow{\alpha} w, \alpha \in C \}$$

を C を control set として G によって生成される言語という.

定義 3 $G = (N, \Sigma, P, S)$ と $\bar{G} = (\bar{N}, \Sigma, \bar{P}, \bar{S})$ を

CF 文法とするとき G から $\bar{G}$ への left-homomorphism ($h: G \rightarrow \bar{G}$ と書く) とは monoid homomorphism $h: P^* \rightarrow \bar{P}^*$ 2

$$h(A_e(G)) \subseteq A_e(\bar{G})$$

を満たすものをいう.

注 以後 left-homomorphism を単に homomorphism と呼び、

$A_e, \Rightarrow, L_e$ などの添字は省略する.

homomorphism $h: G \rightarrow \bar{G}$ が proper であるとは 任意の $w \in \Sigma^*$ に対し

$$S \xrightarrow{\alpha} w \text{ in } G \text{ iff } \bar{S} \xrightarrow{h(\alpha)} w \text{ in } \bar{G}$$

が成立するときをいう.

$h: G \rightarrow \bar{G}$ が proper な homomorphism なる:

$$\begin{aligned} L(\bar{G}; h(A(G))) &= \{ w \in \Sigma^* \mid \bar{S} \xrightarrow{h(w)} w \text{ in } \bar{G}, \alpha \in P^* \} \\ &= L(\bar{G}; h(P)^*) \end{aligned}$$

となることに注意する.

目的 以下で次のような CF 文法 G_U の存在を示す: 任意の CF 文法 G に対し proper な homomorphism $h: G \rightarrow G_U$ が存在する. この結果の一般の文法への拡張は Hart [3] によって試みられている.

Notation Σ を与えられた alphabet とする. (以後 Σ は固定する.) $c, \bar{c}, d, \bar{d}$ を Σ の元でない symbol とし

$$\Delta = \Sigma \cup \{c, \bar{c}, d, \bar{d}\}$$

とおく. $\sim$ を次で定義される Δ^* 上の関係とする.

$$x c \bar{d} \bar{c} y \sim x d y \quad x, y \in \Delta^*$$

$$x d \bar{d} \bar{d} y \sim x \bar{d} y \quad x, y \in \Delta^*$$

$$x a y \sim x y \quad a \in \Sigma, y \in (\Delta - \bar{c}) \Delta^*$$

$\sim^*$ を $\sim$ の reflexive transitive closure とし

$$D = \{ w \in \Delta^* \mid w \sim^* \bar{d} \}$$

とおく.

定義 4 $G = (\{X_0, X_1, \dots, X_n\}, \Sigma, P, X_0)$ を任意の CF 文法 とする.

$$\xi: V^* \rightarrow \Delta^* \quad h: P^* \rightarrow \Delta^*$$

を次のように定義する.

$$\begin{cases} \xi(a) = a & a \in \Sigma \\ \xi(X_i) = d c^i & 0 \leq i \leq n \end{cases}$$

$$h(X_i \rightarrow x) = \bar{c}^i \xi(x^R) \bar{d}, \quad X_i \rightarrow x \in P.$$

Lemma 1. $X_i \xrightarrow{\alpha} w$, for some $w \in \Sigma^*$

$$\text{iff } d c^i \bar{d} h(\alpha) \sim^* \bar{d}$$

系 $\alpha \in A(G) \quad \text{iff} \quad h(\alpha) \in D$

定義 5 $G_v = (\{S, X\}, \Sigma, P_v, S)$

$$P_v = \{ a: S \rightarrow Sa \mid a \in \Sigma \}$$

$$\cup \left\{ \begin{array}{ll} c: S \rightarrow SX, & \bar{c}: X \rightarrow \lambda \\ d: S \rightarrow SS, & \bar{d}: S \rightarrow \lambda \end{array} \right\}$$

以後 P_v と Δ を同視する. すなわち production $p: X \rightarrow x \in P_v$ を文字 $p \in \Delta$ とみなす.

Lemma 2 $A(G_U) = D$

系 $h: P^* \rightarrow \Delta^* (\Delta = P_U)$ は G から G_U への homomorphism.

Theorem $h: G \rightarrow G_U$ は proper homomorphism.

系1 任意の context-free language $L \subset \Sigma^*$ に対し,

$$L = L(G_U; C)$$

となる regular control set C が存在する. (特に C を $C = \{\alpha_1, \alpha_2, \dots, \alpha_k\}^*$, $\alpha_i \in P_U^*$, に選ぶことができる. matrix grammar)

系2 $\{L(G_U; C) \mid C \text{ は regular control set}\}$

$= \Sigma$ 上の context-free languages の族.

系3 任意の CF文法 G と G の regular control set

C_1, C_2 に対し $L(G; C_1) = L(G; C_2)?$, $L(G; C_1) \cap L(G; C_2) = \emptyset?$, $L(G; C_1) - L(G; C_2)$ は finite? などは recursively unsolvable.

注 G が unambiguous ならば 系3の各問題は solvable となる.

Acknowledgment:

有益な御討論をしていただきました、高須達教授
と林健志氏に感謝いたします。

文献

- [1] T. KASAI A hierarchy between context-free and
 context-sensitive languages JCSS 4 (1970)
- [2] T. KASAI A universal context-free grammar
 Information & Control 28 (1975)
- [3] J.M. Hart Two extension of KASAI's
 universal context-free grammar.
 Univ. of Kentucky. technical report No. 22-75
- [4] J.M. Hart Derivation languages and syntactical
 categories. Information & Control 28 (1975)
- [5] R.L. Cannon State grammar parsing
 Doctoral Dissertation. Univ. of North Carolina.
- [6] A.V. Aho and J.D. Ullman The Theory of Parsing
 Translation and Compiling.