9　形式言語論 （2） –文脈自由文法と導出木–

- **9-1：有向木**

グラフとは, 頂点と, それらを結ぶ辺からなる図形のことである．ここでは, 各辺に向き付けがなされているようなグラフを扱う．

定義 9.1. 2 つの集合 V, E と 2 つの写像 $s,t : E \to V$ からなる 4 つ組 $Q = (V, E, s, t)$ を **有向グラフ**という．このとき, V の元を **頂点** といい, E の元を **矢印** と呼ぶ．矢印 $\alpha \in E$ に対して, $s(\alpha)$ を α の **始点**といい, $t(\alpha)$ を α の **終点** と呼ぶ．矢印 α は次のように図示される:

$$s(\alpha) \xrightarrow{\alpha} t(\alpha)$$

頂点 $a, b \in V$ について，w が始点 a から終点 b への長さ $k \geq 1$ の **道** とは，矢印 α_i $(i = 1, \ldots, k)$ の列 $w = (a|\alpha_1, \alpha_2, \ldots, \alpha_k|b)$ であって，

$$s(\alpha_1) = a, \quad t(\alpha_i) = s(\alpha_{i+1}) \ \ (i = 1, \ldots, k-1), \quad t(\alpha_k) = b$$

を満たすものである．道はしばしば単純に $\alpha_1\alpha_2 \cdots \alpha_k$ と表す．

有向グラフの中で特別なものが有向木である．

定義 9.2. 有向グラフ $Q = (V, E, s, t)$ が **有向木** であるとは, 以下の条件を満たすときをいう．

(DT1) 任意の矢印 $\alpha \in E$ に対して, $t(\alpha) \neq v_0$ となるような頂点 v_0 が唯一つ存在する．この v_0 を **根** と呼ぶ．
(DT2) 任意の頂点 $v \in V$ に対して, 始点 v_0 から終点 v への道が存在する．
(DT3) 根ではないような頂点 $v \in V$ に対して, ある矢印 $\alpha \in E$ で $t(\alpha) = v$ となるようなものが唯一つ存在する．

頂点 $v \in V$ に対して, v を始点とする矢印の終点となる頂点を, v の **子** という．子を持たない頂点を **葉** という．頂点 v の **子孫** とは, v もしくは v から道があるような頂点のことをいう．

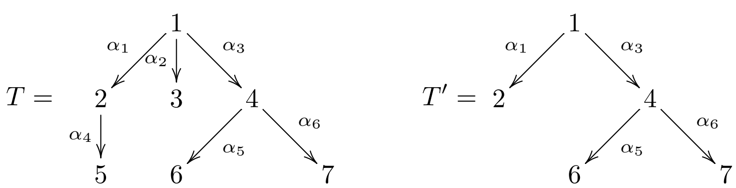
有向木 $T = (V, E, s, t)$ に対して, $T' = (V', E', s', t')$ が T の **部分木** であるとは, $V' \subset V$, $E' \subset E$ であって $s', t' : E' \to V'$ が

$$s'(\alpha) = s(\alpha), \quad t'(\alpha) = t(\alpha) \qquad (\alpha \in E')$$

が成り立つような有向木のことである．

以後, 頂点は自然数でラベル付けされているとし, 頂点 v の子が $n_1, n_2, \ldots, n_k$（ただし, $i < j$ のとき $n_i < n_j$）であるとき, これを図示するときには左から順に $n_1, n_2, \ldots, n_k$ に配置する．

例 9-1　以下の T は有向木の例を与えている．また, T' は T の部分木である．



T において, 頂点 1 が根であり, 頂点 5, 6, 7 が葉である．頂点 1 の子は頂点 2, 3, 4 となる．また T' において, 頂点 2, 6, 7 が葉である．

レポート 9-1　$V = \{1, 2, \ldots, 11\}$, $E = \{\alpha_1, \ldots, \alpha_{10}\}$ とし, 有向木 $Q = (V, E, s, t)$ を以下で定める.

$$s : \alpha_1 \mapsto 1,\ \alpha_2 \mapsto 1,\ \alpha_3 \mapsto 1,\ \alpha_4 \mapsto 1,\ \alpha_5 \mapsto 3,\ \alpha_6 \mapsto 3,\ \alpha_7 \mapsto 4,\ \alpha_8 \mapsto 5,\ \alpha_9 \mapsto 7,\ \alpha_{10} \mapsto 9$$
$$t : \alpha_1 \mapsto 2,\ \alpha_2 \mapsto 3,\ \alpha_3 \mapsto 4,\ \alpha_4 \mapsto 5,\ \alpha_5 \mapsto 6,\ \alpha_6 \mapsto 7,\ \alpha_7 \mapsto 8,\ \alpha_8 \mapsto 9,\ \alpha_9 \mapsto 10,\ \alpha_{10} \mapsto 11$$

有向木 Q を図示せよ. また, これの根と葉はなにか. また, 頂点 3 の子孫をすべて答えよ.

• 9-2 : 導出木

定義 9.3. $G = (V_N, V_T, S, P)$ を文脈自由文法とする. G の **導出木** とは, 次の条件を満たす有向木 $T = (V, E, s, t)$ のことである.

(DET1) 写像 $V \to V_N \cup V_T$ が存在する. すなわち, 頂点それぞれに対応する.

(DET2) 根には開始記号 S が対応する.

(DET3) 子をもつ頂点には, V_N の記号が対応する.

(DET4) 頂点 v の子が左から $n_1, n_2, \ldots, n_k$ であって, v には記号 $\mathtt{A}$ が対応しており, n_i には記号 $\mathtt{A}_i$ が対応しているとき生成規則

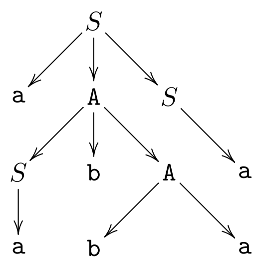
$$\mathtt{A} \longrightarrow \mathtt{A}_1\mathtt{A}_2 \cdots \mathtt{A}_k$$

が P に属する.

例 9-2　文脈自由文法 $G = (\{S, \mathtt{A}\}, \{\mathtt{a}, \mathtt{b}\}, P, S)$ を考えよう. ただし, P は以下の生成規則からなる集合である.

(i) $S \to \mathtt{aA}S$　(ii) $S \to \mathtt{a}$　(iii) $\mathtt{A} \to S\mathtt{bA}$　(iv) $\mathtt{A} \to \mathtt{ba}$　(v) $\mathtt{A} \to SS$

このとき, 以下の木は G の導出木の定義を満たす.



• 9-3 : 結果と導出

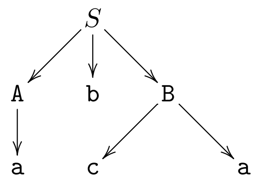
定義 9.4. 有向木 $T = (V, E, s, t)$ の頂点 $v \in V$ について, 根から v への道の長さを v の **深さ** という. v が根のときは, 深さは 0 と定義する.

例 9-3　**例 9-1** にある有向木 T について, 頂点 3 は深さ 1 であり, 頂点 6 は深さ 2 である.

定義 9.5. 文脈自由文法 G の導出木 T があったとき, 次のルールで得られる文字列 w を, T の **結果** と呼ぶ.

- 頂点 v_1, v_2 が頂点 v の子であって, v_1 が v_2 の左にあるとする. このとき, v_1 の子孫は v_2 とその子孫より **左出現** であるという.
- 木 T の葉に現れる文字を, 左出現する順で並べたものが w である.

例 9-4　以下のような導出木があったとする．



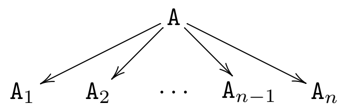
これの結果は abca である．

定理 9.6. $G = (V_N, V_T, P, S)$ を文脈自由文法とする．開始記号を変数 $\mathtt{A} \in V_N$ とする文脈自由文法 $G_{\mathtt{A}} = (V_N, V_T, P, \mathtt{A})$ と $\alpha \neq \varepsilon$ に対して，$\mathtt{A} \overset{*}{\underset{G_{\mathtt{A}}}{\Longrightarrow}} \alpha$ であるための必要十分条件は，α を結果としてもつ $G_{\mathtt{A}}$ の導出木が存在することである．

証明. **（必要条件）** $\mathtt{A} \overset{*}{\underset{G_{\mathtt{A}}}{\Longrightarrow}} \alpha$ が k ステップで得られるとする．導出のステップ数 k による帰納法で証明を行う．以下，$\alpha = \mathtt{A}_1\mathtt{A}_2\cdots\mathtt{A}_n$ とおく．

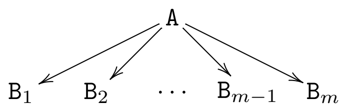
まず，$k = 1$ であるとき，$(\mathtt{A} \to \alpha) \in P$ であり，次のような導出木を考えればよい:

$T_{\mathtt{A}} :=$

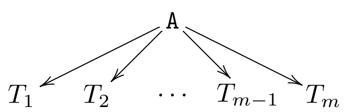


次に，k 未満のステップ数の導出 $\mathtt{B} \overset{*}{\underset{G_{\mathtt{B}}}{\Longrightarrow}} \beta$ については，$G_{\mathtt{B}}$ の導出木であって結果が β となるものが存在すると仮定する．k ステップで得られる導出 $\mathtt{A} \overset{*}{\underset{G_{\mathtt{A}}}{\Longrightarrow}} \alpha$ を考えよう．この導出のはじめが $\mathtt{A} \to \mathtt{B}_1\mathtt{B}_2\cdots\mathtt{B}_m$ であったとすれば，$G_{\mathtt{A}} = (V_N, V_T, P, \mathtt{A})$ は文脈自由文法なので，その生成規則の右辺には変数が高々 1 つしかないから，$\mathtt{A}_i$ は $\mathtt{B}_1, \ldots, \mathtt{B}_m$ のいずれかであるか，またはこれらのいずれから生成された記号である．はじめのステップを表す導出木は

$T' :=$



のようになる．このとき，α のなかで $\mathtt{B}_i$ から生成される部分を β_i とおくことにしよう．つまり，$\alpha = \beta_1\cdots\beta_m$ とする．導出 $\mathtt{B}_i \overset{*}{\underset{G_{\mathtt{B}_i}}{\Longrightarrow}} \beta_i$ を考えれば，これは k 未満のステップで得られる．帰納法の仮定から，結果が β_i であるような $G_{\mathtt{B}_i}$ の導出木 T_i が存在する．これは，根が $\mathtt{B}_i$ に対応するような有向木である．そこで，有向木 T' の葉 $\mathtt{B}_1, \mathtt{B}_2, \ldots, \mathtt{B}_m$ の部分を T_i に置き換えた



は $\beta_1\beta_2\cdots\beta_m = \alpha$ を結果とする $G_{\mathtt{A}}$ の導出木である．

（十分条件） α を結果とする $G_{\mathtt{A}}$ の導出木 T が存在すると仮定する．主張を，葉ではない頂点の個数 k に関する帰納法で証明する．

$k = 0$ のときは，根しかない木であり，$\mathtt{A} \overset{*}{\underset{G_{\mathtt{A}}}{\Longrightarrow}} \mathtt{A}$ は自明に成り立つ．$k = 1$ であるとき，その導出木は必要条件の証明で与えた $T_{\mathtt{A}}$ で与えられて，$(\mathtt{A} \to \alpha) \in P$ である．

$k - 1$ まで主張は正しいと仮定し，α が葉以外の頂点が k（> 1）個の導出木の結果であるとする．$k > 1$ なので，根の子のうち，少なくともひとつは葉ではない．根の子が $1, 2, \ldots, m$ して，それらは左から $\mathtt{B}_1, \ldots, \mathtt{B}_m$ と対応しているとする．このとき，導出木の定義から $(\mathtt{A} \to \mathtt{B}_1\cdots\mathtt{B}_m) \in P$ である．頂点 i を根とする T の部分木

を T_i とおく．T_i は文脈自由文法 $G_{\mathtt{B}_i}$ の導出木であり，その結果が β_i であるとする．すると，$\alpha = \beta_1 \cdots \beta_m$ である．各 T_i の頂点のうち，葉でない頂点の数は T のそれより少ないから，帰納法の仮定より $\mathtt{B}_i \underset{G_{\mathtt{B}_i}}{\overset{*}{\Longrightarrow}} \beta$ である．従って，

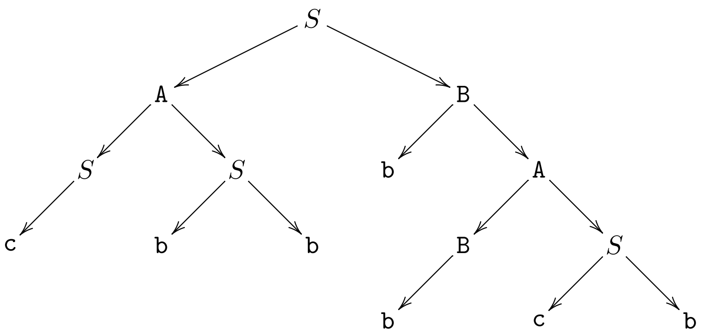
$$\mathtt{A} \underset{G_{\mathtt{A}}}{\Longrightarrow} \mathtt{B}_1\mathtt{B}_2 \cdots \mathtt{B}_m \underset{G_{\mathtt{B}_1}}{\overset{*}{\Longrightarrow}} \beta_1\mathtt{B}_2 \cdots \mathtt{B}_m \underset{G_{\mathtt{B}_2}}{\overset{*}{\Longrightarrow}} \beta_1\beta_2\mathtt{B}_3 \cdots \mathtt{B}_m \underset{G_{\mathtt{B}_3}}{\overset{*}{\Longrightarrow}} \cdots \underset{G_{\mathtt{B}_m}}{\overset{*}{\Longrightarrow}} \beta_1\beta_2 \cdots \beta_m$$

である．ところで，各 $i = 1, 2, \ldots, m$ について $G_{\mathtt{B}_i}$ の生成規則の集合は全て P であるから，

$$\mathtt{A} \underset{G_{\mathtt{A}}}{\overset{*}{\Longrightarrow}} \beta_1\beta_2 \cdots \beta_m = \alpha$$

である．よって十分性が示された． □

レポート 9-2　以下の図は，ある文脈自由文法 $G = (V_N, V_T, P, S)$ に関する導出木である．



この導出木の結果を答えよ．また，この導出木からわかる生成規則を全て述べよ．また，$\mathtt{bcbbcbc} \in \mathcal{L}(G)$ であることを示せ．