2021 年 11 月 9 日

「 オートマトンと形式言語」補足資料 (6)

1　文脈自由文法 (CFG) の例

例 1 CFG として，

$$\mathcal{G}: \qquad S \;\to\; \mathtt{1}S\mathtt{1} \mid \sharp$$

を考える．導出の例をいくつか示す:

$$\begin{aligned} &S \Rightarrow \sharp \\ &S \Rightarrow \mathtt{1}S\mathtt{1} \Rightarrow \mathtt{1}\sharp\mathtt{1} \\ &S \Rightarrow \mathtt{1}S\mathtt{1} \Rightarrow \mathtt{11}S\mathtt{11} \Rightarrow \mathtt{11}\sharp\mathtt{11} \end{aligned}$$

このとき，$L(\mathcal{G}) = \{\mathtt{1}^n\sharp\mathtt{1}^n \mid n \geq 0\}$ となる．

例 2 CFG として，

$$\mathcal{G}: \qquad S \;\to\; \mathtt{0}S\mathtt{11} \mid \mathtt{00}$$

を考える．導出の例をいくつか示す:

$$\begin{aligned} &S \Rightarrow \mathtt{00} \\ &S \Rightarrow \mathtt{0}S\mathtt{11} \Rightarrow \mathtt{00011} \\ &S \Rightarrow \mathtt{0}S\mathtt{11} \Rightarrow \mathtt{00}S\mathtt{1111}DFA \Rightarrow \mathtt{00001111} \end{aligned}$$

このとき，$L(\mathcal{G}) = \{\mathtt{0}^{n+2}\mathtt{1}^{2n} \mid n \geq 0\}$ となる．

例 3 CFG として，

$$\begin{aligned} \mathcal{G}: \qquad S &\;\to\; \mathtt{0}S\mathtt{1} \mid T \\ T &\;\to\; \mathtt{0}T \mid \varepsilon \end{aligned}$$

を考える．導出の例をいくつか示す:

$$\begin{aligned} &S \Rightarrow T \Rightarrow \varepsilon \\ &S \Rightarrow T \Rightarrow \mathtt{0}T \Rightarrow \mathtt{0}\varepsilon = \mathtt{0} \\ &S \Rightarrow T \Rightarrow \mathtt{0}T \Rightarrow \mathtt{00}T \Rightarrow \mathtt{00} \\ &S \Rightarrow \mathtt{0}S\mathtt{1} \Rightarrow \mathtt{0}T\mathtt{1} \Rightarrow \mathtt{01} \\ &S \Rightarrow \mathtt{0}S\mathtt{1} \Rightarrow \mathtt{0}T\mathtt{1} \Rightarrow \mathtt{00}T\mathtt{1} \Rightarrow \mathtt{001} \\ &S \Rightarrow \mathtt{0}S\mathtt{1} \Rightarrow \mathtt{00}S\mathtt{11} \Rightarrow \mathtt{00}T\mathtt{11} \Rightarrow \mathtt{0011} \\ &S \Rightarrow \mathtt{0}S\mathtt{1} \Rightarrow \mathtt{00}S\mathtt{11} \Rightarrow \mathtt{00}T\mathtt{11} \Rightarrow \mathtt{000}T\mathtt{11} \Rightarrow \mathtt{00011} \end{aligned}$$

このとき，$L(\mathcal{G}) = \{\mathtt{0}^i\mathtt{1}^j \mid i \geq j\}$ となる．

例 4 CFGとして，

$$\mathcal{G}: \quad \begin{array}{lcl} S & \to & S\sharp S \mid T \\ T & \to & \mathsf{0}T\mathsf{1} \mid \mathsf{01} \end{array}$$

を考える．導出の例をいくつか示す:

$$S \Rightarrow T \overset{*}{\Rightarrow} \mathsf{0}^3\mathsf{1}^3$$
$$S \Rightarrow S\sharp S \Rightarrow T\sharp S \overset{*}{\Rightarrow} \mathsf{0}^3\mathsf{1}^3\sharp S \Rightarrow \mathsf{0}^3\mathsf{1}^3\sharp T \overset{*}{\Rightarrow} \mathsf{0}^3\mathsf{1}^3\sharp\mathsf{0}^2\mathsf{1}^2$$
$$S \Rightarrow S\sharp S \Rightarrow S\sharp S\sharp S \overset{*}{\Rightarrow} T\sharp T\sharp T \overset{*}{\Rightarrow} \mathsf{01}\sharp\mathsf{0}^2\mathsf{1}^2\sharp\mathsf{0}^3\mathsf{1}^3$$

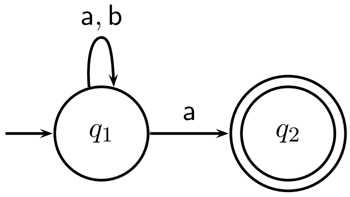
このとき，$L(\mathcal{G}) = \{\mathsf{0}^{i_1}\mathsf{1}^{i_1}\sharp\mathsf{0}^{i_2}\mathsf{1}^{i_2}\sharp\cdots\sharp\mathsf{0}^{i_k}\mathsf{1}^{i_k} \mid k \geq 1, i_1, \ldots, i_k \geq 1\}$となる．

2 NFAから文脈自由文法への変換

NFAから等価な文脈自由文法は，教科書p.122にある，DFAから作るやり方と同じ．つまり，状態遷移図から，

1. 状態を変数にとり，開始状態を開始変数にする．
2. それぞれの状態遷移 $q_i \xrightarrow{a} q_j$ について，$q_i \to aq_j$ を生成規則に追加する．
3. それぞれの受理状態 $q_i \in F$ について，$q_i \to \varepsilon$ を生成規則に追加する．

例 5 アルファベットを $\Sigma = \{\mathsf{a}, \mathsf{b}\}$ とする．以下の状態遷移図で表わされる NFA N_1 を考えてみる．



このとき，q_1, q_2 を変数にとり，以下のような CFG を構成すればよい．

$$\begin{array}{lcl} q_1 & \to & \mathsf{a}q_1 \mid \mathsf{b}q_1 \mid \mathsf{a}q_2 \\ q_2 & \to & \varepsilon \end{array}$$

DFAやNFAが与えられれば，等価なCFGが構成できるので，正規言語は，文脈自由言語でもある．一方，例1の文脈自由言語のように，文脈自由言語のなかには正規言語でないものもある．したがって，以下のような関係が成り立つ．

定理 6 正規言語のクラスは，文脈自由言語のクラスの真部分クラスである．

3 文脈自由言語のクラスの閉包性

文脈自由言語のクラスは，正規言語のクラスと同様に，和集合，連結演算，スター演算について閉じている．一方，正規言語のクラスとは対照的に，積集合と補集合については，閉じていない．

和集合に関する閉包性　2 つの CFG $\mathcal{G}_1$ と $\mathcal{G}_2$ が与えられたとする．

$$\begin{array}{llll} \mathcal{G}_1: & S_1 & \to & \dots \\ & \multicolumn{3}{l}{\dots\dots\dots\dots} \\ \\ \mathcal{G}_2: & S_2 & \to & \dots \\ & \multicolumn{3}{l}{\dots\dots\dots\dots} \end{array}$$

このとき，変数は適当に変更して，$\mathcal{G}_1$ と $\mathcal{G}_2$ では，同じ変数を使っていないようにする．その上で，$\mathcal{G}_1$ と $\mathcal{G}_2$ の両者の生成規則と合併するとともに，新しい開始変数 S_0 を用意して，生成規則 $S_0 \to S_1 \mid S_2$ を追加する．

$$\begin{array}{llll} \mathcal{G}: & S_0 & \to & S_1 \mid S_2 \\ & S_1 & \to & \dots \\ & \multicolumn{3}{l}{\dots\dots\dots\dots\dots} \\ & S_2 & \to & \dots \\ & \multicolumn{3}{l}{\dots\dots\dots\dots\dots} \end{array}$$

この CFG を $\mathcal{G}$ とおくと，

$$L(\mathcal{G}) = L(\mathcal{G}_1) \cup L(\mathcal{G}_2)$$

が成立している．

連結演算に関する閉包性　2 つの CFG $\mathcal{G}_1$ と $\mathcal{G}_2$ が与えられたとする．

$$\begin{array}{llll} \mathcal{G}_1: & S_1 & \to & \dots \\ & \multicolumn{3}{l}{\dots\dots\dots\dots} \\ \\ \mathcal{G}_2: & S_2 & \to & \dots \\ & \multicolumn{3}{l}{\dots\dots\dots\dots} \end{array}$$

このとき，変数は適当に変更して，$\mathcal{G}_1$ と $\mathcal{G}_2$ では，同じ変数を使っていないようにする．その上で，$\mathcal{G}_1$ と $\mathcal{G}_2$ の両者の生成規則と合併するとともに，新しい開始変数 S_0 を用意して，生成規則 $S_0 \to S_1S_2$ を追加する．

$$\begin{array}{llll} \mathcal{G}: & S_0 & \to & S_1S_2 \\ & S_1 & \to & \dots \\ & \multicolumn{3}{l}{\dots\dots\dots\dots\dots} \\ & S_2 & \to & \dots \\ & \multicolumn{3}{l}{\dots\dots\dots\dots\dots} \end{array}$$

この CFG を $\mathcal{G}$ とおくと，

$$L(\mathcal{G}) = L(\mathcal{G}_1) \circ L(\mathcal{G}_2)$$

が成立している．

スター演算に関する閉包性　CFG $\mathcal{G}_1$ が与えられたとする．

$$\begin{array}{llll} \mathcal{G}_1: & S_1 & \to & \dots \\ & \multicolumn{3}{l}{\dots\dots\dots\dots} \end{array}$$

このとき，新しい開始変数 S_0 を用意して，生成規則 $S_0 \to S_1 S_0$ および $S_0 \to \varepsilon$ を追加する．

$$\begin{array}{llll} \mathcal{G}: & S_0 & \to & S_1 S_0 \mid \varepsilon \\ & S_1 & \to & \ldots \\ & \multicolumn{3}{l}{\ldots\ldots\ldots\ldots\ldots\ldots} \end{array}$$

この CFG を $\mathcal{G}$ とおくと，

$$L(\mathcal{G}) = L(\mathcal{G}_1)^*$$

が成立している．

積集合に関する閉包性がないこと　以下の 2 つの言語 L_1 と L_2 を考える．

$$\begin{array}{lll} L_1 & = & \{\mathtt{a}^n\mathtt{b}^n\mathtt{c}^m \mid n, m \geq 0\} \\ L_2 & = & \{\mathtt{a}^n\mathtt{b}^m\mathtt{c}^m \mid n, m \geq 0\} \end{array}$$

このとき，

$$\begin{array}{llll} \mathcal{G}_1: & S_1 & \to & T_1 U_1 \\ & T_1 & \to & \mathtt{a} T_1 \mathtt{b} \mid \varepsilon \\ & U_1 & \to & \mathtt{c} U_1 \mid \varepsilon \\ \mathcal{G}_2: & S_2 & \to & U_2 T_2 \\ & T_2 & \to & \mathtt{b} T_2 \mathtt{c} \mid \varepsilon \\ & U_2 & \to & \mathtt{a} U_2 \mid \varepsilon \end{array}$$

をとると，$L(\mathcal{G}_1) = L_1$，$L(\mathcal{G}_2) = L_2$ となるので，L_1 と L_2 は文脈自由言語となる．このとき，

$$L_1 \cap L_2 = \{\mathtt{a}^n\mathtt{b}^n\mathtt{c}^n \mid n \geq 0\}$$

となるが，後々示されるように (例 2.36)，これは文脈自由言語でない．つまり，L_1 と L_2 が文脈自由言語だからといって，$L_1 \cap L_2$ が文脈自由言語とは限らない．

補集合に関する閉包性がないこと　もし補集合に関する閉包性があるとする．このとき，和集合に関する閉包性を利用して，任意の文脈自由言語 L_1 と L_2 について，

$$L_1 \cap L_2 = \overline{\overline{L_1} \cap \overline{L_2}}$$

が，いつも文脈自由言語になることになってしまう．しかし，これは，積集合に関する閉包性がないことに矛盾する．

　つまり，L_1 が文脈自由言語だからといって，その補集合 $\overline{L_1}$ が文脈自由言語とは限らない．