11　形式言語論 （4） –有限オートマトンと正規文法–

• 11-1：有限オートマトンと正規文法

正規文法によって生成される言語と有限オートマトンによって受理される文字列の集合との間の関係を取り上げよう．まずは具体的に次のような正規文法 $G = (V_N, V_T, P, S)$ を考えてみる．

$$V_N := \{S, \mathtt{A}, \mathtt{B}\}, \quad V_T := \{\mathtt{0}, \mathtt{1}\}$$

として, 生成規則の集合 P は

$$S \to \mathtt{0A}, \quad \mathtt{A} \to \mathtt{0A}, \quad \mathtt{A} \to \mathtt{1B}, \quad \mathtt{A} \to \mathtt{1}, \quad \mathtt{B} \to \mathtt{1B}, \quad \mathtt{B} \to \mathtt{1}$$

であるとする．このとき, $\mathcal{L}(G)$ は 1 個以上の 0 が続いたあとに 1 個以上の 1 が続くような文字列全体の集合である．例えば, G は文字列 000111 を生成する．実際,

$$S \underset{G}{\Longrightarrow} \mathtt{0A} \underset{G}{\Longrightarrow} \mathtt{00A} \underset{G}{\Longrightarrow} \mathtt{000A} \underset{G}{\Longrightarrow} \mathtt{0001B} \underset{G}{\Longrightarrow} \mathtt{00011B} \underset{G}{\Longrightarrow} \mathtt{000111}$$

となることが簡単にわかるだろう．これを見れば, 生成過程のそれぞれにおいて非終端記号は高々 1 つしか現れておらず, それは右端にしか現れないことがわかる．これは, 非終端記号を状態, そして終端記号を受け取る文字列と思ってみれば, 「ある状態から文字を受け取って次の状態へ遷移する」という状況にとても似ている．つまり, 正規文法における非終端記号はオートマトンの状態の役割を果たし, 逆もまた然り, という推測がたつ．そしてこれは実際に正しい．

定理 11.1. 以下の主張が成り立つ．

(1) 有限オートマトン M に対して, ある正規文法 G で $\mathcal{T}(M) = \mathcal{L}(G)$ となるものが存在する．
(2) 正規文法 G に対して, ある有限オートマトン M で $\mathcal{L}(G) = \mathcal{T}(M)$ となるものが存在する．

• 11-2：命題 11-1 の証明

では, **命題 11-1** の証明を与えよう．

(1) **定理 10-4** によって, 有限オートマトン $M = (K, \Sigma, \delta, q_0, F)$ は決定性オートマトンであると仮定しても良い．まず, $q_0 \notin F$ のケースを考える．文法 $G = (V_N, V_T, P, S)$ を以下のように定義しよう．まず, $V_N = K$, $V_T = \Sigma$, $S = q_0$ と定める．生成規則の集合 P は

$$P = \{q \to \mathtt{bC} \mid \delta(q, \mathtt{b}) = \mathtt{C}\} \cup \{q \to \mathtt{b} \mid \delta(q, \mathtt{b}) \in F\}$$

と定義する．明らかに, 文法 G は正規文法である．以下, $\mathcal{T}(M) = \mathcal{L}(G)$ を示そう．

任意に $w \in \mathcal{T}(M)$ をとれば, $q_0 \notin F$ なので w は空文ではない．文字列 w を $w = w_1 w_2 \cdots w_n$ $(w_i \in \Sigma)$ であるとすれば, M はこれを受理するので, 以下を満たす状態の列 $r_0, r_1, \ldots, r_n \in K$ が存在する．

(i) $r_0 = q_0$
(ii) $i = 0, 1, \ldots, n-1$ に対して, $r_{i+1} = \delta(r_i, w_{i+1})$
(iii) $r_n \in F$

(ii) より $i = 0, 1, \ldots, n-1$ に対して, 生成規則 P には「$r_i \to w_{i+1} r_{i+1}$」が属している．よって

$$r_0 \underset{G}{\Longrightarrow} w_1 r_1 \underset{G}{\Longrightarrow} w_1 w_2 r_2 \underset{G}{\Longrightarrow} \cdots \underset{G}{\Longrightarrow} w_1 w_2 \cdots w_{n-1} r_{n-1}$$

となる．(i) から $q_0 \overset{*}{\underset{G}{\Longrightarrow}} w_1 w_2 \cdots w_{n-1} r_{n-1}$ である．(iii) によって $\delta(r_{n-1}, w_n) = r_n \in F$ なので, 「$r_{n-1} \to w_n$」が P に属するから結局, $q_0 \overset{*}{\underset{G}{\Longrightarrow}} w_1 w_2 \cdots w_{n-1} w_n$ となり $w \in \mathcal{L}(G)$ がわかった．

逆に $w = w_1w_2\cdots w_n \in \mathcal{L}(G)$ $(w_i \in V_T)$ をとれば, $q_0 \overset{*}{\underset{G}{\Longrightarrow}} w_1w_2\cdots w_n$ である. 生成規則の定義より, ある状態 $r_{n-1} \in K$ が存在して「$r_{n-1} \to w_n$」が P に属するものがとれる. これは, $\delta(r_{n-1}, w_n) \in F$ ということであり,

$$q_0 \overset{*}{\underset{G}{\Longrightarrow}} w_1w_2\cdots w_{n-1}r_{n-1} \underset{G}{\Longrightarrow} w_1w_2\cdots w_{n-1}w_n$$

である. 次に最後から 2 番目に適用した生成規則を考えば, 生成規則の定義より, ある状態 $r_{n-2} \in K$ が存在して「$r_{n-2} \to w_{n-1}r_{n-1}$」が P に属するものがとれる. これは, $\delta(r_{n-2}, w_{n-1}) = r_{n-1}$ を意味しており,

$$q_0 \overset{*}{\underset{G}{\Longrightarrow}} w_1w_2\cdots w_{n-2}r_{n-2} \underset{G}{\Longrightarrow} w_1w_2\cdots w_{n-1}r_{n-1} \underset{G}{\Longrightarrow} w$$

である. そのあと, 最後から 3 番目に適用した生成規則, 最後から 4 番目に適用した生成規則,... と順に同様に考えることによって, $i = 1, \ldots, n-2$ に対して, $r_{i+1} = \delta(r_i, w_{i+1})$ を得て,

$$q_0 \underset{G}{\Longrightarrow} w_1r_1 \underset{G}{\Longrightarrow} w_1w_2r_2 \underset{G}{\Longrightarrow} \cdots \underset{G}{\Longrightarrow} w_1w_2\cdots w_{n-1}r_{n-1} \underset{G}{\Longrightarrow} w_1w_2\cdots w_{n-1}w_n$$

である. 一番初めに適用する生成規則は「$q_0 \to w_1r_1$」であり, これは $\delta(q_0, w_1) = r_1$ を意味する. こうして得られた状態の列 $q_0, r_1, r_2, \ldots, r_{n-1}, \delta(r_{n-1}, w_n)$ をみれば, M は w を受理していることがわかる.

以上で $\mathcal{T}(M) = \mathcal{L}(G)$ が証明された.

最後に, $q_0 \in F$ のケースを考える. このとき, 空文 $\varepsilon \in \mathcal{T}(M)$ であるが, この場合は先に定義した正規文法 G について $\mathcal{L}(G) = \mathcal{T}(M) \setminus \{\varepsilon\}$ である. しかし, この場合は**命題 8-6** によって新たな正規文法 G' で $\mathcal{L}(G') = \mathcal{L}(G) \cup \{\varepsilon\} = \mathcal{T}(M)$ となるものを構成することができる. 以上で証明が完了した.

(2) 正規文法 $G = (V_N, V_T, P, S)$ に対して, 非決定性オートマトン $M = (K, \Sigma, \delta, q_0, F)$ を以下のように定義しよう. 記号 $\mathtt{A}$ を, V_N に属していない新たな記号であるとする. まず,

$$K := V_N \cup \{\mathtt{A}\}, \quad \Sigma := V_T, \quad q_0 := S, \quad F := \begin{cases} \{S, \mathtt{A}\} & S \to \varepsilon \in P \text{ のとき}, \\ \{\mathtt{A}\} & S \to \varepsilon \notin P \text{ のとき} \end{cases}$$

とおく. 遷移関数 $\delta : K \times \Sigma \cup \{\varepsilon\} \to 2^K$ を任意の非終端記号 $X \in K (= V_N \cup \{\mathtt{A}\})$ と $a \in \Sigma$ に対して,

$$\delta(X, a) := \begin{cases} \{Y \in K \mid X \to aY \in P\} \cup \{\mathtt{A}\} & X \in V_N,\ X \to a \in P \text{ のとき}, \\ \{Y \in K \mid X \to aY \in P\} & X \in V_N,\ X \to a \notin P \text{ のとき}, \\ \varnothing & X = \mathtt{A} \text{ のとき} \end{cases}$$

と定義する. このように構成した有限オートマトン M について, $\mathcal{L}(G) = \mathcal{T}(M)$ であることを確認しよう.

任意に $w = w_1w_2\cdots w_n \in \mathcal{L}(G)$ $(w_i \in V_T)$ をとれば, G は正規文法であるからある変数 $\mathtt{A}_1, \ldots, \mathtt{A}_{n-1} \in V_N$ が存在して

$$S \underset{G}{\Longrightarrow} w_1\mathtt{A}_1 \underset{G}{\Longrightarrow} w_1w_2\mathtt{A}_2 \underset{G}{\Longrightarrow} \cdots \underset{G}{\Longrightarrow} w_1w_2\cdots w_{n-1}\mathtt{A}_{n-1} \underset{G}{\Longrightarrow} w_1w_2\cdots w_n$$

である. 遷移関数 δ の定義によって, $\mathtt{A}_1 \in \delta(S, w_1)$, $\mathtt{A}_2 \in \delta(\mathtt{A}_1, w_2), \ldots, \mathtt{A}_{n-1} \in \delta(\mathtt{A}_{n-2}, w_{n-1})$ が従う. 最後に適用した生成規則が $\mathtt{A}_{n-1} \to w_n \in P$ であるので, $\mathtt{A} \in \delta(\mathtt{A}_{n-1}, w_n)$ である. こうして, 状態の列 $S, \mathtt{A}_1, \ldots, \mathtt{A}_{n-1}, \mathtt{A}$ によって, $w = w_1w_2\cdots w_n$ を M は受理することになる. よって, $w \in \mathcal{T}(M)$ がわかった.

逆に, $w = w_1w_2\cdots w_n \in \mathcal{T}(M)$ $(w_i \in \Sigma)$ を任意にとる. すると, M は w を受理する非決定性オートマトンなので以下を満たす状態の列 $r_0, r_1, \ldots, r_n \in K$ が存在する.

(i) $r_0 = S$

(ii) $i = 0, 1, \ldots, n-1$ に対して, $r_{i+1} \in \delta(r_i, w_{i+1})$

(iii) $r_n \in F$

ゆえに生成規則の集合 P は, 「$S \to w_1 r_1$」, 「$r_1 \to w_2 r_2$」,..., 「$r_{n-2} \to w_{n-1} r_{n-1}$」および「$r_{n-1} \to w_n$」という規則を含む. 従って,

$$S \underset{G}{\Longrightarrow} w_1 r_1 \underset{G}{\Longrightarrow} w_1 w_2 r_2 \underset{G}{\Longrightarrow} \cdots \underset{G}{\Longrightarrow} w_1 w_2 \cdots w_{n-1} r_{n-1} \underset{G}{\Longrightarrow} w_1 w_2 \cdots w_n$$

は G の導出であるから, $w \in \mathcal{L}(G)$ がわかった.

もし, $S \to \varepsilon \in P$ のときは, S がどの生成規則の右辺にも現れないのだから, 状態が S 以外のところから S に戻ることはありえない. 以上で $\mathcal{L}(G) = \mathcal{T}(M)$ が示されて証明が完了した.

- **11-3: 正規文法からオートマトンの構成例**

冒頭で紹介した正規文法 $G = (V_N, V_T, P, S)$ を考えてみる. つまり,

$$V_N := \{S, \mathtt{A}, \mathtt{B}\}, \quad V_T := \{\mathtt{0}, \mathtt{1}\}$$

として, 生成規則の集合 P は

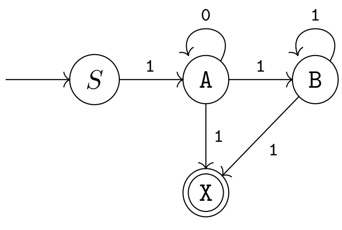
$$S \to \mathtt{0A}, \quad \mathtt{A} \to \mathtt{0A}, \quad \mathtt{A} \to \mathtt{1B}, \quad \mathtt{A} \to \mathtt{1}, \quad \mathtt{B} \to \mathtt{1B}, \quad \mathtt{B} \to \mathtt{1}$$

であるような正規文法である. これと同じ言語を受理する有限オートマトンを構成しよう. まず, V_N に属していない新たな記号 X を用意する. 非決定性オートマトン $M = (K, \Sigma, \delta, q_0, F)$ を証明に倣って構成しよう.

- $K := V_N \cup \{\mathtt{X}\} = \{S, \mathtt{A}, \mathtt{B}, \mathtt{X}\}$
- $\Sigma := V_T = \{\mathtt{0}, \mathtt{1}\}$
- $q_0 := S$
- $F := \{\mathtt{X}\}$
- 遷移関数 $\delta : K \times \Sigma \cup \varepsilon \to 2^K$ は次の通り:

δ	ε	0	1
S	$\varnothing$	$\{\mathtt{A}\}$	$\varnothing$
A	$\varnothing$	$\{\mathtt{A}\}$	$\{\mathtt{A}, \mathtt{X}\}$
B	$\varnothing$	$\varnothing$	$\{\mathtt{B}, \mathtt{X}\}$
X	$\varnothing$	$\varnothing$	$\varnothing$

これを図示すると, 次のような形になる.



レポート 11-1　以下で定義される正規文法 $G = (V_N, V_T, P, S)$ と同じ言語を受理する非決定性オートマトン M を証明の構成に倣ってひとつ与えよ. また, この非決定性オートマトン M と同じ言語を受理する決定性オートマトン M' を構成せよ.

$$V_N := \{S, \mathtt{B}\}, \quad V_T := \{\mathtt{0}, \mathtt{1}\}, \quad P = \{S \to \mathtt{0B},\ \mathtt{B} \to \mathtt{0B},\ \mathtt{B} \to \mathtt{1}S,\ \mathtt{B} \to \mathtt{0}\}$$