第18講 正規文法と正規言語

林　恒俊

正規言語と正規文法

- 正規文法で定義される言語と今まで正規表現や FSA で定義された言語が同じ種類であることを確認する。それには

  1. 与えられた正規文法から同じ言語を受理する FSA が構成できる
  2. 与えられた FSA から同じ言語を受理する正規文法が構成できる

  ことを言わなくてはならない。

正規文法から正規言語

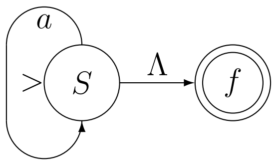
- 次のように文法から NFA を構成して前半を証明する。
  - 終端記号集合は同一
  - 文法の非終端記号集合を状態集合とする
  - 開始記号を初期状態とする
  - $A \to \sigma B$ ただし $A, B \in N$, $\sigma \in \Sigma^*$ という生成規則について (A, σ, B) を遷移表に追加する
  - $A \to \sigma$ ただし $A \in N, \sigma \in \Sigma^*$ という生成規則について (A, σ, f) を遷移表に追加する。なお f は新規の終了状態で必要なら状態集合に追加する。
- 構成された NFA は文法が定義する言語を受理する。なぜなら
  - 生成規則 $A \to \sigma B$ で生成された記号列について NFA は規則の左辺記号 A の状態から右辺の記号 B の状態へ σ により遷移する

- ◦ 生成規則 $A \to \sigma$ で生成された記号列について NFA は規則の左辺記号 A の状態から σ を生成して終了する

からである。

正規文法から正規言語の例

- 正規文法
$$G = (\{S\}, \{a\}, \{S \to aS, S \to \Lambda\}, S)$$
について上記に従って NFA を構築する。なおこの文法が定義する言語 $L(G) = \{a^n \mid n \geq 0\}$ である。
  - 状態集合は $\{S, f\}$, アルファベットは $\{a\}$, 初期状態は S, 終了状態集合は $\{f\}$
  - 状態遷移関係は $\{(S, a, S), (S, \Lambda, f)\}$
  - 状態遷移図は次の通り



たしかにこの NFA は $L(G)$ を受理する

正規言語から正規文法

- 次のように与えられた言語に対応する DFA から文法を構成して後半の証明をおこなう。
  - DFA のアルファベットを終端記号集合とする
  - DFA の状態集合を非終端記号集合とする
  - DFA の初期状態を開始記号とする
  - (q_i, x, q_j) ただし $q_i, q_j \in Q,\ x \in \Sigma$ という状態遷移について $q_i \to xq_j$ という生成規則を追加する

  - すべての終了状態 $f \in F$ について $f \to \Lambda$ という生成規則を追加する

- このように構成された文法は DFA が受理する記号列を生成することが可能である。なぜなら
  - DFA は初期状態では入力記号列は何も読まない。また導出過程で最初の文型式が開始記号のみの場合は生成された文はない。
  - 状態遷移 (q_i, x, q_j) では入力記号列から x を読み次状態に遷移する。生成規則 $q_i \to xq_j$ による導出過程では最後尾の 1 個の非終端記号 q_i を書換える。すなわち文型式の後尾に x を追加して非終端記号が q_j になる。なお正規文法では文型式中に非終端記号が常に最後尾に 1 個だけ存在する。
  - DFA はすべて入力を読んだ後終了状態になれば受理する。文法で終了状態に対応する非終端記号が最後尾にあれば空列で置換される。結果非終端記号が消去され文を生成する。
  - DFA による文の受理と文法による文の生成は同等に行われる。

  からである。

正規言語から正規文法の例

- 次の DFA M_e

$$M_e = (\{q_0, q_1\}, \{0, 1\}, \delta, q_0, \{q_0\}) \quad \text{ただし}$$
$$\delta = \{(q_0, 0, q_0),\ (q_0, 1, q_1),\ (q_1, 0, q_1),\ (q_1, 1, q_0)\}$$

  は 1 が偶数個含まれる記号列を受理する。

- これと同じ言語を生成する正規文法 $G_e = (N, \Sigma, P, S)$ は
  - 非終端記号集合 $N = \{q_0, q_1\}$
  - 終端記号集合 $\Sigma = \{0, 1\}$
  - 生成規則

$$P = \{q_0 \to 0q_0, q_0 \to 1q_1, q_1 \to 0q_1, q_1 \to 1q_0, q_0 \to \Lambda\}$$

  - ◦ 開始記号 $S = q_0$

  である。

- この文法が生成する記号列にはたしかに 1 が偶数個含まれる。なぜなら非終端記号が文型式から取去られるためには q_0 でなければならないし、q_0 が文型式に現れるためには 1 が偶数個現れなければならないからである。
- 次の G_e の文 0110 の生成列を考察するとよい。

$$q_0 \Rightarrow 0q_0 \Rightarrow 01q_1 \Rightarrow 011q_0 \Rightarrow 0110q_0 \Rightarrow 0110$$

正規言語は正規文法で定義可能

- 以上の考察から

$$\mathcal{L}_{\mathbf{RL}} = \mathcal{L}_{\mathbf{FSA}} = \mathcal{L}_{\mathbf{RG}}$$

  という結論が得られる。