

「オートマトンと形式言語」補足資料(4)

1 正規表現の定義に関する補足 (教科書 p.74–75)

正規表現に使われる記号である $\varepsilon, \emptyset, \cup, \circ, *$ (教科書 p.74) と言語に対する正規演算を表わす $\varepsilon, \emptyset, \cup, \circ, *$ (教科書 p.51) とは, (それぞれが対応するので同じ記号を使うのは自然なのだが), 本来は別のものであるということに注意.

以下では, 違いを区別するため, 正規表現に使われる記号の方を, フォントや記号を若干変更して, 定義を書き下してみる.

アルファベット Σ 上の正規表現の集合 $\mathcal{RE}$ を, 以下のように, BNF にもとづき定義する:

$$R_1, R_2 \in \mathcal{RE} ::= a \mid \varepsilon \mid \emptyset \mid R_1 \cup R_2 \mid R_1 \bullet R_2 \mid R_1^*$$

(ただし, ここで, $a \in \Sigma$.)

このとき, $R \in \mathcal{RE}$ が表わす言語 $L(R)$ は, 以下のように再帰的に定義される.

$$\begin{aligned} L(a) &= \{a\} \\ L(\varepsilon) &= \{\varepsilon\} \\ L(\emptyset) &= \emptyset \\ L(R_1 \cup R_2) &= L(R_1) \cup L(R_2) \\ L(R_1 \bullet R_2) &= L(R_1) \circ L(R_2) \\ L(R_1^*) &= L(R_1)^* \end{aligned}$$

定義の左辺の L の中に出てくる記号は, 正規表現に使われる記号. 一方, 定義の右辺に使われている記号は, 言語に対する正規演算である.

教科書では, 更に, $L(\)$ を基本的には省略する (p.75) ので,

$$\begin{aligned} a &= \{a\} \\ \varepsilon &= \{\varepsilon\} \\ \emptyset &= \emptyset \\ R_1 \cup R_2 &= R_1 \cup R_2 \\ R_1 \bullet R_2 &= R_1 \circ R_2 \\ R_1^* &= R_1^* \end{aligned}$$

ということになる (!). ということ, 対応するので同じ記号を使うことは構わないのだが, 読むときには, どちらの意味で使われているのか, 読者側で認識する必要がある.

2 冗長な記号の導入について (教科書 p.74, p.75)

教科書でも紹介されているが, 正規表現の記法を簡潔にするため, しばしば, いくつかの冗長な記号を用いることがある.

1.

$$R^k \stackrel{\text{def}}{=} \overbrace{R \circ \cdots \circ R}^{k \text{ 回}}$$

正確には、以下のように再帰的に定義する：

$$\begin{aligned} R^0 &= \varepsilon \\ R^{k+1} &= R^k \circ R \end{aligned}$$

$R^0 = \varepsilon$ と定義するのは、 ε が $\circ$ の単位元となるからである（本補足資料の4節を参照のこと）。

2. アルファベットを $\Sigma = \{a_1, \dots, a_n\}$ とするとき、

$$\Sigma \stackrel{\text{def}}{=} a_1 \cup \cdots \cup a_n$$

なお、前節でも述べたように、文脈に応じて、 Σ が集合の意味で用いられているのか、正規表現として用いられているのかを区別するのは、読者の側の責任である。

3.

$$R^+ \stackrel{\text{def}}{=} R \circ R^*$$

なお、 $R \circ R^* = R^* \circ R$ という恒等式が成立するので、 $R^+ \stackrel{\text{def}}{=} R^* \circ R$ と考えてもよい。

3 正規表現の括弧づけについて（教科書 p.75）

正規表現の括弧の省略法について、整理しておく。

1. $R \circ S$ は、 $\circ$ を省略して、 RS と書く
2. 正規表現では、演算子の結合力として、

$$(\)^*, (\)^+, (\)^k \quad > \quad \circ \quad > \quad \cup$$

を考え、括弧を省略する。

3. $\circ$ と $\cup$ は結合性を満たす（本補足資料の4節を参照のこと）ので、 $\circ$ や $\cup$ が左結合か右結合かは特に区別しない。

例えば、 $(R_1 \cup R_2) \cup R_3$ や $R_1 \cup (R_2 \cup R_3)$ は、 $R_1 \cup R_2 \cup R_3$ と記す。 $(R \circ S) \circ T$ は、 RST と書かれる。 $R \circ S^*$ は、 $R \circ (S^*)$ の略であり、 $(R \circ S)^*$ の略ではない。

なお、以下のような記号もよく用いられる：

- $\cup$ の代わりに、 $+$ や $|$
- ε の代わりに、 1
- $\emptyset$ の代わりに、 0

4 正規表現の等価性に関する補足 (教科書 p.75–76)

教科書 p.76 にいくつか例があるように、正規表現の間では、さまざまな恒等式が成り立つ。ここでの、2つの正規表現 R と S の等価性とは、表わす言語が等しい (つまり、 $L(R) = L(S)$ となること) ことを意味する。

教科書でも後で示すように、2つの正規表現が等しい言語を表わすかどうかは、それらの正規表現を NFA、そして、DFA へと変換することで、DFA の等価性判定を用いて、チェックすることが出来る。ただ、実際にこれを手で行うのは煩雑だし、プログラムでやるのも場合によっては実際的にはできない場合もある。このため、正規表現の間でどのような等価変形ができるか知っておくと便利。

まず、正規表現の演算については、合同性が成立することを述べておく。以下の、推論規則は、上式が成立すれば下式も成立する、という意味になっている。

1. 合同性

1.

$$\frac{R_1 = R_2}{R_1 \cup S = R_2 \cup S} \quad \frac{R_1 = R_2}{S \cup R_1 = S \cup R_2}$$

2.

$$\frac{R_1 = R_2}{R_1 \circ S = R_2 \circ S} \quad \frac{R_1 = R_2}{S \circ R_1 = S \circ R_2}$$

3.

$$\frac{R = S}{R^* = S^*}$$

合同性により、部分表現を等価変形したものは、等価である。

2. $\cup$ と $\circ$ に関する恒等式

以下に記載する恒等式は、 ε や $\emptyset, \cup, \circ$ の意味を考えると明らかだろうと思われる。

1. $(R_1 \cup R_2) \cup R_3 = R_1 \cup (R_2 \cup R_3)$ ($\cup$ の結合律)

2. $R_1 \cup R_2 = R_2 \cup R_1$ ($\cup$ の交換律)

3. $R \cup R = R$ ($\cup$ のべき等律)

4. $\emptyset \cup R = R \cup \emptyset = R$ ($\cup$ の単位元)

5. $(R_1 \circ R_2) \circ R_3 = R_1 \circ (R_2 \circ R_3)$ ($\circ$ の結合律)

6. $\varepsilon \circ R = R \circ \varepsilon = R$ ($\circ$ の単位元)

7. $\emptyset \circ R = R \circ \emptyset = \emptyset$ ($\circ$ の零元)

8. $(R_1 \cup R_2) \circ S = (R_1 \circ S) \cup (R_2 \circ S)$
 $S \circ (R_1 \cup R_2) = (S \circ R_1) \cup (S \circ R_2)$ (分配律)

3. スター演算に関する恒等式 (1)

スター演算に関する恒等式では、そのままではあまり自明でないので、スター演算に関する性質である

$$\begin{aligned} L(R^*) &= L(R)^* \\ &= \{w_1 \cdots w_k \mid w_1, \dots, w_k \in L(R)\} \\ &= \{\varepsilon\} \cup L(R) \cup L(R^2) \cup \cdots = \bigcup_{i \geq 0} L(R^i) \end{aligned}$$

から考えるとよい。なお、下記の説明では、

$$R^* = \varepsilon \cup R \cup R^2 \cup \cdots$$

を頻繁に用いるが、正規表現には無限和を取る演算はないので、正確には、 $L(\)$ を使った表記に戻して、等価性を示す必要があることに注意しておく。

1. $R^* \circ R^* = R^*$

説明:

$$R^* \circ R^* = \overbrace{R \cdots R}^{0 \text{ 回以上}} \overbrace{R \cdots R}^{0 \text{ 回以上}} = \overbrace{R \cdots R}^{0 \text{ 回以上}} = R^*$$

2. $(R^*)^* = R^*$

説明:

$$\begin{aligned} (R^*)^* &= \varepsilon \cup R^* \cup (R^* \circ R^*) \cup (R^* \circ R^* \circ R^*) \cup \cdots \\ &= \varepsilon \cup R^* \cup R^* \cup R^* \cup \cdots \\ &= R^* \end{aligned}$$

3. $\varepsilon \cup R^* = R^* \cup \varepsilon = R^*$

説明:

$$\begin{aligned} \varepsilon \cup R^* &= \varepsilon \cup (\varepsilon \cup R \cup R^2 \cup R^3 \cup \cdots) \\ &= \varepsilon \cup R \cup R^2 \cup R^3 \cup \cdots \\ &= R^* \end{aligned}$$

4. $R \cup R^* = R^* \cup R = R^*$

説明:

$$\begin{aligned} R \cup R^* &= R \cup (\varepsilon \cup R \cup R^2 \cup R^3 \cup \cdots) \\ &= R \cup \varepsilon \cup R \cup R^2 \cup R^3 \cup \cdots \\ &= \varepsilon \cup R \cup R^2 \cup R^3 \cup \cdots \\ &= R^* \end{aligned}$$

5. $(\varepsilon \cup R)^* = R^*$

説明:

$$\begin{aligned} (\varepsilon \cup R)^* &= \varepsilon \cup (\varepsilon \cup R) \cup (\varepsilon \cup R)^2 \cup (\varepsilon \cup R)^3 \cup \cdots \\ &= \varepsilon \cup (\varepsilon \cup R) \cup (\varepsilon \cup R \cup R^2) \cup (\varepsilon \cup R \cup R^2 \cup R^3) \cup \cdots \\ &= \varepsilon \cup R \cup R^2 \cup R^3 \cup \cdots \\ &= R^* \end{aligned}$$

6. $\varepsilon^* = \varepsilon$

説明:

$$\begin{aligned} \varepsilon^* &= \varepsilon \cup \varepsilon^1 \cup \varepsilon^2 \cup \varepsilon^3 \cup \cdots \\ &= \varepsilon \cup \varepsilon \cup \varepsilon \cup \varepsilon \cup \cdots \\ &= \varepsilon \end{aligned}$$

7. $\emptyset^* = \varepsilon$

説明:

$$\begin{aligned}\varepsilon^* &= \varepsilon \cup \emptyset^1 \cup \emptyset^2 \cup \emptyset^3 \cup \dots \\ &= \varepsilon \cup \emptyset \cup \emptyset \cup \emptyset \cup \dots \\ &= \varepsilon\end{aligned}$$

8. $R \circ R^* = R^* \circ R$

説明:

$$\begin{aligned}R \circ R^* &= R \circ (\varepsilon \cup R \cup R^2 \cup R^3 \cup \dots) \\ &= R \cup R^2 \cup R^3 \cup R^4 \cup \dots \\ &= (\varepsilon \cup R \cup R^2 \cup R^3 \cup \dots) \circ R \\ &= R^* \circ R\end{aligned}$$

9. $R^* = \varepsilon \cup R \circ R^*$

説明:

$$\begin{aligned}R^* &= \varepsilon \cup R \cup R^2 \cup R^3 \cup \dots \\ &= \varepsilon \cup R \circ (\varepsilon \cup R \cup R^2 \cup \dots) \\ &= \varepsilon \cup R \circ R^*\end{aligned}$$

10. $(R_1 \circ R_2)^* \circ R_1 = R_1 \circ (R_2 \circ R_1)^*$

説明:

$$\begin{aligned}(R_1 \circ R_2)^* \circ R_1 &= (\varepsilon \cup (R_1 R_2) \cup (R_1 R_2)^2 \cup (R_1 R_2)^3 \cup \dots) \circ R_1 \\ &= \varepsilon \cup R_1 R_2 \cup R_1 R_2 R_1 R_2 \cup R_1 R_2 R_1 R_2 R_1 R_2 \cup \dots \circ R_1 \\ &= R_1 \cup R_1 R_2 R_1 \cup R_1 R_2 R_1 R_2 R_1 \cup R_1 R_2 R_1 R_2 R_1 R_2 R_1 \cup \dots \\ &= R_1 \circ (\varepsilon \cup R_2 R_1 \cup R_2 R_1 R_2 R_1 \cup R_2 R_1 R_2 R_1 R_2 R_1 \cup \dots) \\ &= R_1 \circ (R_2 \circ R_1)^*\end{aligned}$$

11. $(R_1 \circ R_2)^* = \varepsilon \cup R_1 \circ (R_2 \circ R_1)^* \circ R_2$

説明:

$$\begin{aligned}(R_1 \circ R_2)^* &= \varepsilon \cup (R_1 R_2) \cup (R_1 R_2)^2 \cup (R_1 R_2)^3 \cup \dots \\ &= \varepsilon \cup R_1 R_2 \cup R_1 R_2 R_1 R_2 \cup R_1 R_2 R_1 R_2 R_1 R_2 \cup \dots \\ &= \varepsilon \cup R_1 \circ (R_2 \cup R_2 R_1 R_2 \cup R_2 R_1 R_2 R_1 R_2 \cup \dots) \\ &= \varepsilon \cup R_1 \circ (\varepsilon \cup R_2 R_1 \cup R_2 R_1 R_2 R_1 \cup \dots) \circ R_2 \\ &= \varepsilon \cup R_1 \circ (\varepsilon \cup R_2 R_1 \cup R_2 R_1 R_2 R_1 \cup \dots) \circ R_2 \\ &= \varepsilon \cup R_1 \circ (R_2 \circ R_1)^* \circ R_2\end{aligned}$$

12. $(R_1 \cup R_2)^* = R_1^* \circ (R_2 \circ R_1^*)^* = (R_1^* \circ R_2)^* \circ R_1^* = (R_1^* \circ R_2^*)^*$

説明:

$$\begin{aligned}
(R_1 \cup R_2)^* &= \varepsilon \cup (R_1 \cup R_2) \cup (R_1 \cup R_2)^2 \cup (R_1 \cup R_2)^3 \cup \dots \\
&= \varepsilon \cup (R_1 \cup R_2) \cup \bigcup_{i,j \in \{1,2\}} R_i R_j \cup \bigcup_{i,j,k \in \{1,2\}} R_i R_j R_k \cup \dots \\
&= \bigcup_{i_1, \dots, i_k \in \{1,2\} (k \geq 0)} R_{i_1} \dots R_{i_k} \\
R_1^* \circ (R_2 \circ R_1^*)^* &= R_1^* \circ (\varepsilon \cup R_2 R_1^* \cup R_2 R_1^* R_2 R_1^* \cup R_2 R_1^* R_2 R_1^* R_2 R_1^* \cup \dots) \\
&= R_1^* \cup R_1^* R_2 R_1^* \cup R_1^* R_2 R_1^* R_2 R_1^* \cup R_1^* R_2 R_1^* R_2 R_1^* R_2 R_1^* \cup \dots \\
&= \bigcup_{i_1, \dots, i_k \in \{1,2\}} \overbrace{R_{i_1} \dots R_{i_k}}^{R_2 \text{ は } 0 \text{ 回}} \cup \bigcup_{i_1, \dots, i_k \in \{1,2\} (k \geq 0)} \overbrace{R_{i_1} \dots R_{i_k} (k \geq 0)}^{R_2 \text{ は } 1 \text{ 回}} \cup \dots \\
&= \bigcup_{i_1, \dots, i_k \in \{1,2\} (k \geq 0)} R_{i_1} \dots R_{i_k} \\
(R_1^* \circ R_2)^* \circ R_1^* &= (\varepsilon \cup R_1^* R_2 \cup R_1^* R_2 R_1^* R_2 \cup R_1^* R_2 R_1^* R_2 R_1^* R_2 \cup \dots) \circ R_1^* \\
&= R_1^* \cup R_1^* R_2 R_1^* \cup R_1^* R_2 R_1^* R_2 R_1^* \cup R_1^* R_2 R_1^* R_2 R_1^* R_2 R_1^* \cup \dots \\
&= \bigcup_{i_1, \dots, i_k \in \{1,2\} (k \geq 0)} \overbrace{R_{i_1} \dots R_{i_k}}^{R_2 \text{ は } 0 \text{ 回}} \cup \bigcup_{i_1, \dots, i_k \in \{1,2\} (k \geq 0)} \overbrace{R_{i_1} \dots R_{i_k}}^{R_2 \text{ は } 1 \text{ 回}} \cup \dots \\
&= \bigcup_{i_1, \dots, i_k \in \{1,2\} (k \geq 0)} R_{i_1} \dots R_{i_k} \\
(R_1^* \circ R_2^*)^* &= \varepsilon \cup R_1^* R_2^* \cup R_1^* R_2^* R_1^* R_2^* \cup \dots \\
&= R_1^* R_2^* \dots R_1^* R_2^* \\
&= \bigcup_{i_1, \dots, i_k \in \{1,2\} (k \geq 0)} R_{i_1} \dots R_{i_k}
\end{aligned}$$

4. スター演算に関する恒等式 (2)

$S \leq T \stackrel{\text{def}}{\iff} S \cup T = T$ と定義すると，以下の推論が成立する．

1.

$$\frac{S \cup R \circ T \leq T}{R^* \circ S \leq T}$$

2.

$$\frac{S \cup T \circ R \leq T}{S \circ R^* \leq T}$$