数理解析研究所講究録
353巻 1979年 228-237

正規表現の直和標準形

山本修一郎[+]　　稲垣 康善[++]　　本多 波雄[+]
（ + 名大工　　++ 三重大工 ）

§1 はじめに

正規表現の標準形を与えようとする試みはいくつかある[1~5]。その中でBRZOZOWSKI, COHEN[4]は，すべての正規表現がその形に帰着できるような直列形の標準形を与えている。これに対して本報告では直和形の標準形として直和標準形並びに主直和標準形正規表現を提案する。また，オートマトンの標準形を定義して標準形オートマトンが受理する語の集合の正規表現と直和標準形並びに主直和標準形正規表現との関連について述べる。

§2 正規表現の直和標準形

定義1　＋を用いないで スター＊，連接・のみで構成される正規表現を＋free正規表現という。 ▨

定義2　$\Gamma_1, \cdots \Gamma_n$を＋free正規表現とする。各$\Gamma_i, \Gamma_j (i \neq j)$に対して，$|\Gamma_i| \cap |\Gamma_j| = \phi$が成立するとき*，正規表現$\Gamma_1 + \cdots + \Gamma_n$は

(*) 正規表現 R に対し |R| で その表現する正規集合をあらわす。

直和標準形正規表現とよばれる。 ▨

<u>定義3</u>　決定性有限オートマトン(DFA)を $A=(\Sigma, Q, q_0, \delta, F)$ であらわす。ここで，Σ は入力記号の有限集合，Q は状態の有限集合，$q_0 \in Q$ は初期状態，$\delta: Q\times\Sigma \to Q$ は遷移関数．$F\subseteq Q$ は最終状態の有限集合である。また，δ は通常の方法で $Q\times\Sigma^+$ 上に拡張されているとする。このとき，$Q' \subseteq Q$ に対して

$$a_{ij}^{Q'} \equiv \{w \mid \delta(i,w)=j,\ \forall \alpha\in\Sigma^+ [w=\alpha x] \to [\delta(i,\alpha)\in Q']\}$$

ただし，$Q=\{1,2,\cdots,n\}$ とする。 ▨

<u>定義4</u>　DFA $A=(\Sigma, Q, q_0, \delta, F)$，$Q=\{1,2,\cdots,n\}$ に対して

$$t_{ij}^{0} \equiv t_1+\cdots+t_m \quad \text{if} \quad a_{ij}^{\phi}=\{t_1,\cdots,t_m\} \quad t_i\in\Sigma$$

$$t_{ij}^{k+1} \equiv t_{ij}^{k}+t_{i\,k+1}^{k}(t_{k+1\,k+1}^{k})^{*}t_{k+1\,j}^{k} \quad (0\le k\le n)$$ ▨

さてDFA, $A=(\Sigma, Q, q_0, \delta, F)$，$Q=\{1,2,\cdots,n\}$ に対して $|t_{ij}^{n}|=a_{ij}^{Q}$ が成立する。従って $F=\{f_1,\cdots,f_m\}$ とすると A の受理する入力系列の集合 $a_{q_0f_1}^{Q}\cup\cdots\cup a_{q_0f_m}^{Q}$ は正規表現 $t_{q_0f_1}^{n}+\cdots+t_{q_0f_m}^{n}$ によってあらわされる。このとき，次の式がなりたつ。

<u>補題1</u>　すべての $t_{q_0f_i}^{n}$ は $+$ free 正規表現 $\tau_{f_i}^{n1},\cdots,\tau_{f_i}^{n\,I(n,f_i)}$ によって $t_{q_0f_i}^{n}=\tau_{f_i}^{n1}+\cdots+\tau_{f_i}^{n\,I(n,f_i)}$ と分解でき各 $\tau_{f_i}^{n\ell}$ に対して $\ell\neq\ell'$ ならば，$|\tau_{f_i}^{n\ell}|\cap|\tau_{f_i}^{n\ell'}|=\phi$ とできる。 ▨

この性質から次の定理が導かれる。

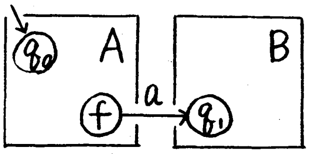
<u>定理1</u>　かってな正規表現 Γ に対して直和標準形正規表現 $\Gamma_1+\cdots+\Gamma_n$ が存在して　$\Gamma=\Gamma_1+\cdots+\Gamma_n$ が成立する。 ▨

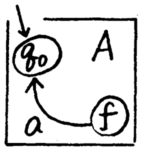
§3 決定性有限オートマトンの標準形

定義5 標準形DFAの帰納的定義

(1) q_0 は標準形DFAである。

(2) A, Bが標準形DFAであるとき，Aの状態fからの入力aに対する遷移先が *dead state* でA, Bの初期状態が q_0, q_1 ならば，



および


は標準形DFAである。

(3) 以上で定義されるもののみ標準形DFAである。 ▨

この定義では最終状態に関する記述がないが標準形DFAの適当な状態を指定して最終状態とすればよい。また，以下では特に断らずに *dead state* を省略している。

定理2 DFA $A=(\Sigma, Q, q_0, \delta, F)$ に関して，Aが標準形DFAであることと $\forall q \in Q$ に対して q_0 から q に導く入力系列で同じ状態を2度以上経由しないものは1つしかないこととは同等である。▨

この性質により，任意のDFA A が標準形DFAであるか否かを簡単に知ることができる。また任意のDFA Aから，Aと等価な標準形DFA A^* を導くこともできる。これをアルゴリズム1により示す。

アルゴリズム1

(入力) DFA $A=(\Sigma, Q, 0, \delta, F)$, $Q=\{0, 1, \cdots, n\}$

(出力) 標準形DFA $A^*=(\Sigma, Q^*, (\lambda, 0), \delta^*, F^*)$

(方法) 初期化 $Q^0 \leftarrow \{(\lambda, 0)\}$, $Q^i \leftarrow \phi$ $(n+2 \geq i > 0)$, $\ell \leftarrow 1$

$PATH[(\lambda,0)] \leftarrow \{(\lambda,0)\}$

```
begin  while  Q^{ℓ-1} ≠ φ  do
  begin  Q* ← Q* ∪ Q^{ℓ-1}
     while  Q^{ℓ-1} ≠ φ  do
     begin  Q^{ℓ-1} の中のかってな元 (w,i) を選ぶ;
            Q^{ℓ-1} から (w,i) を削除する;
            V ← Σ
            while  V ≠ φ  do
            begin  Vの中のかってな元 a を選ぶ;
                   V から a を削除する;
                   j ← δ(i, a)
                   if ∃x∈Σ* (x,j)∈PATH[(w,i)]
                   then  δ*((w,i), a) ← (x,j)
                   else
                     begin  δ*((w,i),a) ← (wa,j)
                            Q^ℓ ← Q^ℓ ∪ (wa,j)
                        PATH[(wa,j)] ← PATH[(w,i)] ∪ {(wa,j)}
                   end
            end
     end
     ℓ ← ℓ+1
```

end

$F^* \leftarrow \{(W, i) \mid i \in F\}$

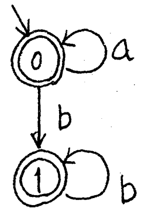
end ▨

<u>定理3</u>　任意のDFA Aに対して，Aと等価な標準形DFA A*が存在する。▨

<u>定理4</u>　標準形DFA $A=(\Sigma, Q, q_0, \delta, F)$ に対して q_0 から $\forall q \in Q$ へ至るすべての入力系列の集合は+free正規表現であらわされる。▨

従って標準形DFAに対して直和標準形正規表現をその各+free正規表現が標準形DFAの各最終状態と1:1に対応するように自然に構成することができる。

ところが，正規集合 $\{a^i b^j \mid i, j \geq 0\}$ に対する1つの直和標準形正規表現 $a^* b^*$ を考えると最終状態1つでこの集合を受理する標準形DFAは存在しないことがわかる。これに対し $a^* + a^* b b^*$ という直和標準形正規表現に対しては次のような標準形DFAが存在して $|a^*| = a_{00}^{\{0,1\}}$, $|a^* b b^*| = a_{01}^{\{0,1\}}$ をみたす。



このように，直和標準形正規表現 $r_1 + \cdots + r_n$ に対して，いつも標準形DFA, $A=(\Sigma, Q, q_0, \delta, F)$ が存在して次の条件を満足するとは限らない。

<u>条件1</u>　$\forall r_i \; \exists q \in F \;\; |r_i| = a_{q_0 q}^{Q}$ ▨

以下ではこの問題について述べる。

§4 主直和標準形正規表現

定義6　正規表現 α が Σ^* の元であるか，または $\alpha=\alpha_0(\alpha_1^*\cdots\alpha_n^*)^*\beta$ の形をしていて $\forall i\,(0\leq i\leq n)\,\forall x\in\Sigma^*\,[\,|\alpha_i|\not\ni\beta\cdot x\,]$ をみたすとき，α はループ可能であるといい，$LC(\alpha)$ とかく。▨

定義7　正規表現 α が $\alpha=\alpha_0\alpha_1$，$\alpha_0\in\Sigma^+$ の形をしているとき，α は先端条件をみたすといい，$PC(\alpha)$ とかく。▨

定義8　正規表現 $\alpha_1,\cdots,\alpha_n$ に対して，$\forall i,j\,(1\leq i\neq j\leq n)\,\exists x,y,z\,\exists a,b\in\Sigma\,[\alpha_i=x\cdot a\cdot y,\ \alpha_j=x\cdot b\cdot z,\ a\neq b]$ がなりたつとき $\alpha_1,\cdots\alpha_n$ は分離可能といい，$Sep(\alpha_1,\cdots,\alpha_n)$ とかく。▨

定義9　正規表現の集合 M_0 の帰納的定義

(1) $\phi,\lambda\in M_0$

(2) $a\in\Sigma$ ならば $a\in M_0$

(3) $\alpha,\beta\in M_0$ かつ $LC(\alpha)$ のとき $\alpha\cdot\beta\in M_0$

(4) $\alpha\in M_0$ かつ $PC(\alpha)$ かつ $LC(\alpha)$ のとき $\alpha^*\in M_0$

(5) $\alpha_1,\cdots,\alpha_n\in M_0$ かつ任意の $i\,(1\leq i\leq n)$ に対し $LC(\alpha_i)$ かつ $PC(\alpha_i)$ かつ $Sep(\alpha_1,\cdots,\alpha_n)$ のとき $(\alpha_1^*\cdots\alpha_n^*)^*\in M_0$

(6) $\alpha_1,\cdots,\alpha_n\in M_0$ かつ任意の $i\,(1\leq i\leq n)$ に対し $LC(\alpha_i)$ かつ $PC(\alpha_i)$ かつ $Sep(\alpha_1,\cdots,\alpha_n)$ のとき $(\alpha_1^*\cdots\alpha_{n-1}^*)^*\alpha_n\in M_0$

(7) 以上により定義されるもののみ M_0 の元である。▨

定義10　M_0 を次のように拡張する。

(1) M_0 の元は M の元である。

(2) $\alpha=\alpha_0(\alpha_1^* \cdots \alpha_n^*)^* \in M_0$ かつ $\alpha_i=\beta\cdot\delta$, $\beta,\delta\in M_0$, $PC(\delta)$ のとき $\alpha\cdot\beta\in M$

(3) 以上により構成されるもののみMの元である。 ▨

<u>定理5</u>　任意の標準形DFA　$A=(\Sigma,Q,q_0,\delta,F)$ に対し $\exists x\in M$ $a^{Q}_{q_0f}=|x|$ $(f\in F)$ が成立する。また任意の $x\in M$ に対しある標準形DFA $A=(\Sigma,Q,q_0,\delta,F)$ が存在して $\exists f\in F$　$a^{Q}_{q_0f}=|x|$ が成立する。 ▨

<u>定義11</u>　正規集合 Γ に対する直和標準形正規表現を $\alpha_1+\cdots+\alpha_n$ とする。このとき，かってな α_i,α_j $(i\neq j)$ に対し

(1) $Sep(\alpha_i,\alpha_j)$

(2) $\exists\beta\in M$　$\exists a\in\Sigma$ ($\alpha_j=\alpha_i\cdot a\cdot\beta$ または $\alpha_i=\alpha_j\cdot a\cdot\beta$)

のいずれか一方が成立するならば　$\alpha_1+\cdots+\alpha_n$ を主直和標準形正規表現という。 ▨

このようにして定義された主直和標準形正規表現は§3の条件1を満足するような直和標準形正規表現である。すなわち，次の定理6が成立する。

<u>定理6</u>　$\alpha_1+\cdots+\alpha_n$ を正規集合 Γ に対する主直和標準形正規表現とするとき標準形DFA　$A=(\Sigma,Q,q_0,\delta,F)$ が存在して $F=\{f_1,\cdots f_n\}$ $|\alpha_i|=a^{Q}_{q_0f_i}$　をみたす。 ▨

このように主直和標準形正規表現は標準形DFAと明確に対応するので，同一の正規集合に対する主直和標準形正規表現の全体は最大元をもつ束となる。

この問題を次に述べる。

定義12　正規集合Γに対する主直和標準形正規表現r_1, r_2に対して関係$\prec$を定義する。$r_1 = L_1 + \cdots + L_\ell$, $r_2 = M_1 + \cdots + M_m$

$$r_1 \prec r_2 \Leftrightarrow \forall L_i \, \exists M_j \; |L_i| \subseteq |M_j|$$

$$r_1 = r_2 \Leftrightarrow [\forall L_i \, \exists M_j \; |L_i| = |M_j|] \,\&\, [\forall M_j \, \exists L_i \; |M_j| = |L_i|]$$ ▨

このとき次の性質が成立する。

補題2　Γを正規集合，r_1, r_2をΓに対する主直和標準形正規表現とすると，ある主直和標準形正規表現r_3が存在して，

$$(|r_3| = \Gamma) \,\&\, (r_3 \prec r_1) \,\&\, (r_3 \prec r_2)$$ ▨

定義13　DFA $A_i = (\Sigma, Q_i, q_0^i, \delta_i, F_i)$ $(i=0,1)$ に対し関係$\leftarrow$をきめる。$A_0 \leftarrow A_1 \Leftrightarrow \forall q^0 \in Q_0 \, \exists q^1 \in Q_1 \; (\alpha^{Q_0}_{q_0^0 q^0} \subseteq \alpha^{Q_1}_{q_0^1 q^1})$ ▨

補題3　DFA A, Bに対し，アルゴリズム1により構成される標準形DFAをA^*, B^*とすると$A \leftarrow B$ならば$A^* \leftarrow B^*$である。▨

また標準形DFA Aには自然な形で主直和標準形正規表現が対応しているからこれを$r_0(A)$とかくことにする。

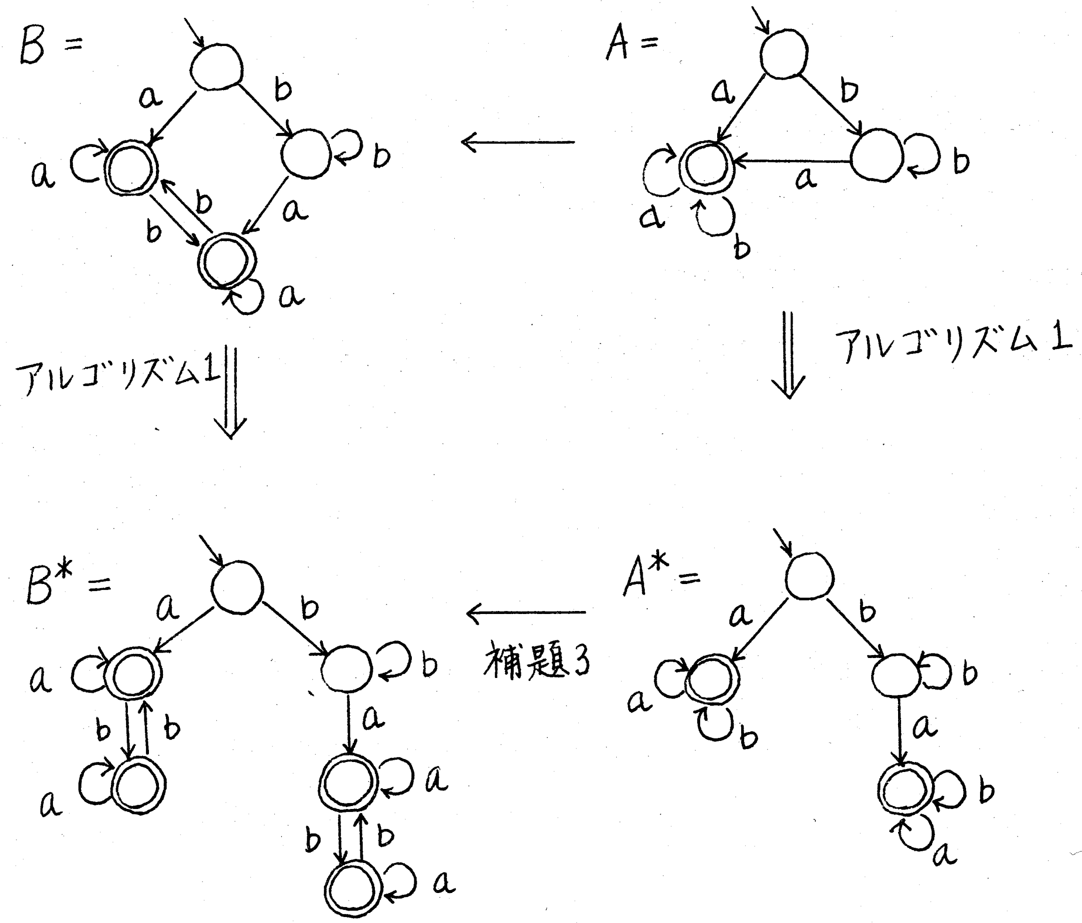
補題4　標準形DFA A, Bに対し$A \leftarrow B$ならば$r_0(A) \prec r_0(B)$ ▨

以上の補題2,3,4を用いて次の定理を示すことができる。

定理7　正規集合Γに対する主直和標準形正規表現の全体は$\prec$の意味で最大の要素をもつ束になる。▨

このことにより，主直和標準形正規表現に対しては$\prec$の意味で最大な要素をとることによって＋の個数が最小な主直和標準形正規表現を求めることができる。

補題3,4に対する例題



自然な主直和標準形 ⇓　　　　　　　　⇓ 自然な主直和標準形

$$g_0(B^*) = a(a^*(ba^*b)^*)^* + a(a^*(ba^*b)^*)^*ba^* + bb^*a(a^*(ba^*b)^*)^* + bb^*a(a^*(ba^*b)^*)^*ba^*$$

$\prec$ 補題4

$$g_0(A^*) = a(a^*b^*)^* + bb^*a(a^*b^*)^*$$

§6 あとがき

正規表現の直和標準形ならびに主直和標準形について調べ，それらと決定性有限オートマトンの標準形との関連について明らかにした。また，主直和標準形正規表現に対して関係≺を導入し，これについて調べた。

謝辞　末筆ながら，日頃御指導賜る名古屋大学福村晃夫教授ならびに研究室の皆様に感謝致します。

参考文献

(1) 五十嵐，本多：一般化された決定性事象とオートマトン　信学論 51C 10（S.43）

(2) 高浪，本多：一般化された準有限確定事象とオートマトン　信学論 54C 9（S.46）

(3) Brzozowski, Cohen: On Decompositions of Regular Events　J.A.C.M, 16, 1 (1969)

(4) Paz, Pelege: On Concatenative Decompositions of Regular Events, I.E.E.E, vol. C-17, 3 (1968)