数理解析研究所講究録
179巻 1973年 102-115

可変構造オートマトンと言語

九大理　有川節夫

§1. はじめに

通常のオートマトンは時間に影響されない一定の構造をもつものである。これに対して，Gill〔1〕，Agasandjan〔2〕は有限オートマトンの拡張として，時間の経過につれて，構造が変化し得る有限オートマトンを考えている。このようなオートマトンを可変構造オートマトンという。可変構造オートマトンは学習機械のモデルとしてだけでなく，一種の神託を使用する計算装置としても興味深い。とくに，可変構造の系列機械，有限オートマトンは状態数を固定して，付加情報の使用を無制限に許すときに実時間で何が処理できるか，という面からも意味のあるモデルであるように思える。

この種のオートマトンについては，その後 Salomaa〔3〕，Teixeira〔4〕等によって，認識装置の面から研究され，通常の有限オートマトンとの関係について多くの成果が得られている。また，文脈自由文法の拡張のために導入された，制御集合をもつ文法に

関する近年の研究 Ginsburg [5], Salomaa [6] 等は可変構造オートマトンへの文法の側からの接近とみることもできる。

本稿では，種々の可変構造オートマトン，可変構造文法について考え，通常のオートマトン・文法との違いを指摘する。

§2. 可変構造有限オートマトン

定義2・1. 可変構造有限オートマトン（$tvfa$ と略記）とは $\mathfrak{A}=(S, A, f, s_0)$ のことをいう，ここに，f は N を自然数の全体とするとき，$f: N \longrightarrow S^{S\times A} \times 2^S$ であり，$f(n)=(\delta_n, F_n)$ とするとき，各 n に対して，$\mathfrak{A}_n=(S, A, \delta_n, s_0, F_n)$ は通常の有限オートマトンである．なお上の $\mathfrak{A}$ を $\mathfrak{A}=(S, A, \{\delta_n\}, s_0, \{F_n\})$ と書くこともある。

定義2.2. $tvfa$ $\mathfrak{A}=(S, A, \{\delta_n\}, s_0, \{F_n\})$ に対して，$\delta: S\times A^* \longrightarrow S$ を $\delta(s, \varepsilon)=s$,

$$\delta(s, a_1\cdots a_n)=\delta_n(\cdots\delta_2(\delta_1(s, a_1), a_2), \cdots), a_n)$$

によって定義する．語 $x\in A^*$ は $\delta(s_0, x)\in F_{|x|+1}$ なるときに $\mathfrak{A}$ によって受理されるという。受理される語の全体を $L(\mathfrak{A})$ で示す。言語 $L\subseteq A^*$ に対して $L(\mathfrak{A})=L$ となる $tvfa$ $\mathfrak{A}$ が存在するとき L を t-正則という。

$C(L, n)$ を同値関係 E_L：

$$xE_Ly \Longleftrightarrow (\forall z)(xz\in L \longleftrightarrow yz\in L)$$

による，長さnの語の属する同値類の集合とする。このとき，

定理2.1.（Salomaa [3]）

Lがt－正則 $\Longleftrightarrow (\exists K)(\forall n)(|C(L,n)| \leq K)$.

4－tuple (S, A, f, s_0)において，$f(n) \subset S^{S \times A} \times 2^S$であるとき$\mathfrak{A} = (S, A, f, s_0)$を非決定性$tvfa$という。これに関して，

定理2.2.〔3〕 非決定性$tvfa$ $\mathfrak{A}$に対して，$L(\mathfrak{A}) = L(\mathfrak{A}')$となる決定性$tvfa$が存在する。

定理2.3.〔3〕 任意の$tvfa$ $\mathfrak{A}$に対して$L(\mathfrak{A}) = L(\mathfrak{A}')$となるconstant final statesをもつ$tvfa$ $\mathfrak{A}'$が存在する。

しかし，constant transition tableをもつ$tvfa$によって受理される言語族はt－正則集合族の真部分族になる〔3〕.

記号．今後つぎの記号を用いる。

$|X|$：集合Xの元の個数，　$|w|$：語wの長さ，

$|w|_a$：語wの中のaの個数，$x^{(n)}$：語xの左からn番目の記号．

§3. 可変構造右線形文法

本節では$tvfa$に対応する可変構造文法について考える。

定義3.1. 可変構造右線形文法($tvrlg$)とは5－tuple $G = (V, V_T, P, f, \sigma)$のことをいう。ここに$(V, V_T, P, \sigma)$は通常の右線形文法であり，$f$は各$n$に対して$f(n) \subseteq P$なる関数である。$\alpha, \beta \in V^*$に対してつぎの(i)または(ii)が成立するときに，

$\alpha \xrightarrow{n} \beta\,(G)$ または $\alpha \xrightarrow{n} \beta$ と書く：

(i) $\alpha = x\eta$, $\beta = xu\xi$ & $\eta \to u\xi \in f(n)$,

(ii) $\alpha = x\eta$, $\beta = xu$ & $\eta \to u \in f(n)$.

また，$\alpha, \beta \in V^*$ に対して，(iii)のような $\gamma_0, \gamma_1, \cdots, \gamma_k \in V^*$ が存在するとき $\alpha \underset{k}{\overset{n}{\Rightarrow}} \beta\,(G)$ または $\alpha \underset{k}{\overset{n}{\Rightarrow}} \beta$ と書く：

(iii) $\gamma_0 = \alpha \xrightarrow{n} \gamma_1 \xrightarrow{n+1} \cdots \xrightarrow{n+k-1} \gamma_k = \beta$.

そして，G によって生成される言語を $L(G)$ と書く。すなわち，

$$L(G) = \{w\,;\,(\exists n)\ \sigma \underset{n}{\overset{1}{\Rightarrow}} w\,(G)\ \&\ w \in V_T^*\}.$$

L に対して $L = L(G)$ となる tvrlg G が存在するとき L を可変構造右線形言語(tvrll)という。

通常の右線形文法は有限オートマトンと同値であるが，可変構造の場合はそうではない。実際つぎのことが簡単にいえる：

定理3.1. すべての言語は tvrll である。

証明　$L \subseteq V_T^*$ に対して，$L - \{x\,;\,|x| \le 2\} = \{x_1, x_2, \cdots\}$ とする．（L は無限集合と仮定してよい．）tvrlg $G = (V, V_T, P, f, \sigma)$ を つぎのように定める：$V = \{\sigma, \eta\} \cup V_T$,

$$P = \bigcup_{a \in V_T} \{\sigma \to a\sigma,\ \sigma \to a\} \cup \{\sigma \to \eta,\ \eta \to \eta,\ \eta \to \sigma\},$$

$$\begin{cases} f(\tau(r)+1) = \{\sigma \to x_r^{(1)}\sigma,\ \sigma \to \eta\}, \\ f(\tau(r)+k) = \{\sigma \to x_r^{(k)}\sigma,\ \eta \to \eta\}, \\ f(\tau(r+1)) = \{\sigma \to x_r^{(|x_r|)},\ \eta \to \sigma\}, \end{cases}$$

ここに，$r = 1, 2, \cdots$; $1 < k < |x_r|$ であり τ はつぎのような

関数である:

$$\begin{cases} \tau(1) = 0 \\ \tau(n+1) = \sum_{i=1}^{n} |x_i| \qquad (n \geq 1). \end{cases}$$

そうすると明らかに, $L(G) = L - \{x ; |x| \leq 2\}$ である. そこで,

$$f(1) := f(1) \cup \{\sigma \to x ;\ x \in L \ \& \ |x| \leq 2\}$$

とすると, $L(G) = L$ となる。

この文法が強過ぎた原因は $\eta \to \xi$ の形の規則を許した点にある。そこで, この規則の使用を禁止した文法を考える。さらにオートマトンの側から拡張した文法について考える。

定義3.2. tvrlg $G = (V, V_T, P, f, \sigma)$ において,

(i) $(\forall \eta)(\forall \xi)(\eta \to \xi \notin P)$ なるとき G を α 型文法といい,

(ii) $(\forall n)(\exists k)(\eta \to u\xi \in f(n) \Rightarrow |u| = k)$ なるとき G を β 型文法という。 また

(iii) β 型文法 G が $\eta \to a\xi$ または $\eta \to \varepsilon$ の形の規則だけをもつとき, G は normal であるという。

補題3.1. 任意の β 型文法 G に対して $L(G) = L(G')$ となる normal な β 型文法 G' が存在する。

定理3.2. 言語 L が t-正則 $\Longleftrightarrow$ L は β 型言語である。

補題3.2. t-正則でない α 型言語が存在する。

証明. 言語 $L = \bigcup_{n=1}^{\infty} \{ab, abb\}^{3^n}$ は α 型であるが t-正則でないことを示そう。α 型であることは, つぎのような f を考えれば

明らかである:

$$f(n)=\begin{cases}\{\sigma\to ab\sigma,\ \sigma\to abb\sigma,\ \sigma\to ab,\ \sigma\to abb\} \\ \qquad (\text{ある}\ m\ \text{に対して}\ n=2^m\ \text{のとき}) \\ \{\sigma\to ab\sigma,\ \sigma\to abb\sigma\} \qquad (\text{そうでないとき}).\end{cases}$$

各 k に対して,

$$x_{kp} = (ab)^{3(k-p)}(abb)^{2p} \qquad (0\leq p\leq k)$$

とすると, $|x_{kp}| = 6k$, $|x_{kp}|_a = 3k-p$ である. いまLが t-正則であるとすると, 定理2.1.によって, $|C(L,n)|\leq K$ となるKが存在する。そうすると各 $k=3^n>K$ に対して, E_L による同値類で語 x_{ki}, x_{kj} $(0\leq i<j\leq k)$ を共に含むものが存在する. z を $x_{ki}z\in L$ となる最短の語とすると, $x_{kj}z\in L$ となる. ところが, 任意の m に対して,

$$\begin{aligned}|x_{kj}z|_a &= |x_{kj}|_a + |(ab)^i|_a \\ &= 3k-j+i \\ &\neq 3^m.\end{aligned}$$

したがって, Lは t-正則でない.

α 型, β 型の言語族間にはつぎの関係がある:

定理3.3. A 上の α 型, β 型の言語の全体をそれぞれ $C_\alpha(A)$, $C_\beta(A)$ とするとき, つぎが成立する:

(i) $C_\beta(A) = C_\alpha(A) = 2^{A^*}$ $\quad(|A|=1$ のとき$)$,

(ii) $C_\beta(A) \subsetneqq C_\alpha(A) \subsetneqq 2^{A^*}$ $\quad(|A|\geq 2$ のとき$)$.

証明. (i)は1個の記号からなるアルファベット上の言語はすべて木-正則である，という事実による. (ii) 左の キ は上の補題により, 右の キ は次節の定理4.2 による。

§4. 準同型写像と例.

木-正則言語族に関する多くの性質は判明しているが，準同型写像によって木-正則性が保存されるか否かについては知られていなかった。ここではα型, β型言語に関して, 簡単な結果を述べる。

定理4.1. (i) Lをβ型言語とし, hを $(\forall a)(\forall b)[|h(a)|=|h(b)|]$ なる準同型写像とすると, $h(L)$ は β型言語である。

(ii) Lをα型言語とし, hを ε-なし準同型写像とすると $h(L)$ はα型である。

定理4.2. (i) 任意の言語Lに対して, $L=h(L')$ となる木-正則言語 L' と準同型写像hが存在する。

(ii) 任意のα型言語Lに対して, $L=h(L')$ となる木-正則言語 L' と ε-なし準同型写像 h が存在する.

定理3.3.(ii) と上の定理によって;

系4.1. 木-正則性は(ε-なし)準同型写像によって必ずしも保存されない。

定理4.3. α型でない言語が存在する。

証明. 言語 $L=\{a^n b^{n^2} ; n \geq 1\}$ がα型でないことを示そう。

いまLがλ型であると仮定する。定理4.2.(ii)によって，k−正則言語$L' \subseteq A^*$とε−なし準同型写像hが存在して$L = h(L')$となる。

$$(4.1) \qquad k = \max_{c \in A,\, n} \{ |C(L', n)|, |h(c)| \} + 1$$

とする。(L'がk−正則であるからこのようなkは必ず存在する).
このkに対してつぎのような語の列$x_0, x_1, \cdots, x_k$ ($x_i \in A^*$)
を選ぶことができる：

$$(4.2) \qquad \begin{cases} |x_i| = k^{2k+2}, & \\ |h(x_i)|_a = k^{2k+2-i}, & (i = 0, 1, \cdots, k) \\ \text{ある } w_i \text{ に対して } x_i w_i \in L'. & \end{cases}$$

L'がk−正則であるから$E_{L'}$による同値類でx_i, x_j ($i<j$)を共に含むものが存在する．(4.2)より$x_j w_j \in L$とすれば$h(w_j) \in b^*$であり，

$$(4.3) \qquad |h(x_j)|_b \geq k^{2k+2} - k^{2k-j+2}$$

であるから，

$$(4.4) \qquad |h(w_j)|_b \leqq k^{2(2k-j+2)} - (k^{2k+2} - k^{2k-j+2})$$

となる．ところで

$$(4.5) \qquad |h(x_i w_j)|_b = |h(x_i)|_b + |h(w_j)|_b \,,$$

$$(4.6) \qquad |h(x_i)|_b \leqq k^{2k+3} - k^{2k-i+2}$$

であるから

$$(4.7) \qquad |h(x_i w_j)|_b \leq k^{2k+3} - k^{2k-i+2} + k^{2(2k-j+2)} - k^{2k+2} + k^{2k-j+2}$$

となる．一方，(4.2) および L の定義によって

(4.8) $$|h(x_i w_j)|_b = k^{2(k-i+2)}$$

となる。ところが $k^{2(k-i+2)} >$ "(4.7) の右辺" となり矛盾する。したがって $x_i w_j \notin L'$ となる。このことは x_i と x_j が同じ同値類に属すということに反す．したがって，L は α型でない。

例.

(1) $\{c^n a^n b^n ;\ n \geq 1\}$ は t-正則でない.

(2) $\{c^{n(n-1)} a^n b^n ;\ n \geq 1\}$ は t-正則である。

(3) $\{a^{n^2} b^n ;\ n \geq 1\}$ は t-正則である。

(4) $\{a^n b^{n^2} ;\ n \geq 1\}$ は t-正則でも，α型でもない。

(5) $\{a^n b^n ;\ n \geq 1\}$ は t-正則でない。

(6) $\{a^{2^n} b^{2^n} ;\ n \geq 1\}$ は t-正則である。

(注)：(1)，(5) は α型でもないと予想されるが証明は未だできていない。

§5. 可変構造二方向有限オートマトン

可変構造オートマトンに関する興味ある問題として，'どの種のオートマトンを可変構造にすればすべての言語を受理できるようになるか？' という問題がある．この問題に関しては，1個の記号からなるアルファベット上の言語はすべて tvfa で受理できることが〔3〕で示されている．一般の言語については〔7〕において Turing 機械について試

みられている。筆者は〔8〕において線型有界オートマトンで可能であることを示したが，ここではさらに二方向有限オートマトンで十分であることを示す。

定義 5.1. 可変構造二方向有限オートマトン（2tvfa）とは 4-tuple $\mathfrak{A} = (S, A, f, s_0)$ のことをいう，ここに f は各 n に対して値 $f(n) = (\delta_n, F_n)$ をとる関数であり，$(S, A, \delta_n, s_0, F_n)$ は通常の二方向有限オートマトンである。

2tvfa による計算，受理される言語等に関する定義は有限オートマトンに対する可変構造有限オートマトンの場合と同様にできるから省略する。

定理 5.1. すべての言語は 2tvfa で受理できる。

証明. $L \subseteq A^*$ として $L - \{\varepsilon\} = \{x_1, x_2, \cdots\}$ $(|x_n| \leq |x_{n+1}|)$ とする．（L が無限の場合だけを考えればよい）．L を受理する constant final states をもつ 2tvfa $\mathfrak{A} = (S, A, \{\delta_n\}, s_0, F)$ を作る．ここでは δ_n を $\{sarm\,;\, \delta_n(s, a) = (r, m)\}$ で表わすことにする．関数 τ： $\tau(1) = 0$, $\tau(p+1) = \tau(p) + 2|x_p|$ $(p \geq 1)$ を用いて δ_n をつぎのように定める： 各 p に対して，

$$\delta_{\tau(p)+k} = \left\{ \begin{array}{ll} s_0 x_p^{(k)} s_1 1, \ s_0 c s_2 1 & \\ s_1 x_p^{(k)} s_1 1, \ s_1 c s_2 1 \ ; & \begin{array}{l} a, c \in A, \ c \neq x_p^{(k)} \\ i = 2, 3 \end{array} \\ s_i a s_i 1 & \end{array} \right\}$$

$$k = 1, \cdots, |x_p| - 1, \quad (|x_p| > 1 \text{ のとき}),$$

$$\delta_{\tau(p)+|x_p|} = \left\{\begin{array}{ll} s_0\, x_p^{(|x_p|)} s_3\, 1,\ s_0\, c\, s_2\, 0 & a, c \in A,\ c \neq x_p^{(|x_p|)} \\ s_1\, x_p^{(|x_p|)} s_3\, 1,\ s_1\, c\, s_2\, 0\ ; & i = 2, 3 \\ s_i\, a\, s_1\, 0 & \end{array}\right\}$$

$$\delta_{\tau(p)+|x_p|+1} = \{ s_3\, a\, s_1 -1,\ s_i\, a\, s_1\, 0\ ;\ \ a \in A,\ 0 \leq i \leq 2 \}$$

$$\delta_{\tau(p)+|x_p|+1+k} = \{ s_i\, a\, s_i -1\ ;\ \ a \in A,\ 0 \leq i \leq 3 \},$$

$$k = 1, \cdots, |x_p| - 1,\ (|x_p| > 1 \text{ のとき}).$$

そして，$S = \{s_i\ ;\ 0 \leq i \leq 3\}$，$F = \{s_3\} \cup \{s_0\ ;\ \varepsilon \in L\}$とおくと，2tvfa $\mathfrak{A}$ は L を受理する．実際，$\varepsilon \in L \Leftrightarrow \mathfrak{A}$ は ε を時間1において状態 s_0 で受理する。$x \in L$ & $x \neq \varepsilon$ ならば $x = x_n$ となる x_n が存在し，時間 $\tau(n) + |x_n|$ において状態 s_3 で x を受理する。$x \notin L$ & $x \neq \varepsilon$ ならば，$|x| < |x_n|$ となる最小の n に対して $\mathfrak{A}$ は x が x_n の *initial subword* であるか否かに従って状態 s_1 または s_2 で時間 $\tau(n) + |x|$ にテープ右の方からとびだす．したがって $L(\mathfrak{A}) = L$.

系5.1. 任意の帰納的に可算な集合 L に対して $L = L(\mathfrak{A})$ となる帰納的な関数 f をもつ 2tvfa が存在する。

定理5.2. 2tvfa $\mathfrak{A} = (S, A, f, s_0)$ において f が帰納的であれば，$L(\mathfrak{A})$ は帰納的に可算である。

上の定理 5.1 の証明により，2tvfa においては最終状態の集合を

一定にすることができるが，状態関数を一定にすることはできない〔4〕．

§6. 可変構造系列機械

可変構造完全系列機械（tvcsmと略す）とは，各nに対して$f(n)=(\delta_n, \lambda_n)$とするとき5-tuple $M=(S, A, B, f, s_0)$のことをいう．ここに各nに対して$(S, A, B, \delta_n, \lambda_n, s_0)$は通常のcsmである。$\bar{\delta}: S\times A^* \to S$をtvfaの場合と同様に定義し，$\bar{\lambda}: S\times A^* \to B^*$を$\bar{\lambda}(s, \varepsilon)=\varepsilon$，$\bar{\lambda}(s, a_1\cdots a_n)=\lambda_1(s, a_1)\cdots\lambda_n(\bar{\delta}(s, a_1\cdots a_{n-1}), a_n)$として，各$x\in A^*$に対して，$M(x)=\bar{\lambda}(s_0, x)$とする．

定理6.1. 任意のtvcsm Mに対して$M\equiv M'$となる一定の出力関数をもつtvcsm M'が存在する。

証明. tvcsm $M=(S, A, B, \{\delta_n\}, \{\lambda_n\}, s_0)$に対して，一定の出力関数をもつtvcsm $M'=(S', A, B, \{\delta_n'\}, \lambda', s_0')$をつぎのように定める：

$S'=S\times B^{S\times A}$, $s_0'=\langle s_0, \lambda_1\rangle$，各$n, s, \lambda, a$に対して

$$\delta_n'(\langle s, \lambda\rangle, a)=\langle\delta_n(s, a), \lambda_{n+1}\rangle,$$

$$\lambda'(\langle s, \lambda\rangle, a)=\lambda(s, a).$$

そうすると明らかに，すべての$x\in A^*$に対して$M(x)=M'(x)$となる。

さて，$L\subseteq A^*$に対してtvcsm $M=(S, A, B, f, s_0)$と$F\subset B$が存在し，$L(M, F)=\{x;\ M(x)=yb\ \&\ b\in F\}$とするとき，$L=L(M, F)$となるならば，$L$はtvcsmによって受理されるという。このときつぎのことが簡単に証明できる：

定理6.2. Lが$t$vcsmによって受理される$\Longleftrightarrow$ Lはt-正則.

定理6.3. Lが一定の状態関数をもつ$t$vfaによって受理される.

$\Longleftrightarrow$ Lは一定の状態関数をもつ$t$vcsmによって受理される.

この定理と§2の注意により$t$vcsmの状態関数は有限オートマトンの場合と同様に必ずしも一定にできないことが分る。

§7. おわりに

可変構造の世界では通常のオートマトン間,オートマトンと文法間で成立する多くの関係が成立しないことをみてきた。本文では触れなかったが,$t$vcfgは,α型$t$vrlgの場合と同様に$\eta \to \xi$型の規則を禁止しても,$\eta \to \varepsilon$型の規則を許せばすべての言語が生成される.そして非決定性の$t$vpdaと$t$vcfgとは同等であることを証明できる.したがって,すべての言語を非決定性の$t$vpdaは受理できる。しかし,決定性の$t$vpdaの認識能力は弱いように思える。決定性の通常のpdaで受理できないcfl $\{a^i b^j a^k ; i=j \text{ or } i=k\}$は決定性$t$vpdaで受理できるがその鏡像は不可能のように思われる。cfgに関しては$\eta \to \varepsilon$, $\eta \to \xi$の形の規則を禁止したときにどのような言語族が得られるか,という問題がある.また,t-正則言語と通常のcflに関しては,補集合もcflであるcflはt-正則でないように思われるが,証明は未だできていない。

参考文献

1. Gill,A., Time-varying sequential machines. J. Franklin Inst. 276 (1963), 519-539.
2. Agasandjan,G.A., Automata with a variable structure. Soviet Physics Doklady 12 (1967), 420-421.
3. Salomaa,A., On finite automata with a time-variant structure. Inform. Control 13 (1968), 85-98.
4. Teixeira,S.R.P., Time-varying automata. Doctor's thesis. Univ. of California, Berkeley, (1970).
5. Ginsburg,S. & Spanier,E.H., Control sets on grammars. Math. System Theory 2(1968), 159-177.
6. Salomaa,A., Periodically time-variant context-free grammars. Inform. Control 17 (1970), 294-311.
7. 水本,豊田,田中,中田., Time-variant チューリング機械. 信学誌 54-C No.11 (1971), 1062-1063.
8. Arikawa,S., On linear bounded automata with a time-variant structure. RIFIS-RR 29 (1972), 1-12.
9. Arikawa,S., Time-variant automata and time-variant grammars. (In prep.).

(1973-3-1 京大数研)