

有限オートマトンの特徴集合について

北九州大 商 棚次 奎介
九大 理 情報研 有川 節夫

§1. 有限オートマトンとその特徴集合

Biermann [1] は任意の有限サンプル集合をもとに, 有限オートマトンの対話型学習アルゴリズムを提供した. ここでは, その学習アルゴリズムの中核をなす有限オートマトンの特徴集合と次数についての 2, 3 の性質を調べ, 正則言語族の一つの階層を与える.

$\alpha = (S, \delta, s_0, F)$ を Σ 上の有限オートマトンとする. 長さ k 以下の語 $x \in \Sigma^*$ が存在して $\delta(s, x)$ または $\delta(s', x)$ のどちらか一方だけが最終状態となるとき, 二つの状態 $s, s' \in S$ は k -区別可能 であるという. S の任意の二つの状態が k -区別可能となるような最小の k を α の 次数 といい, $d(\alpha)$ で表わす. $\#S = 1$ のときは $d(\alpha) = 0$ と定める. $d(\alpha)$ は α が既約であるとき, そのときのみ定義される. 以後 α は最小オートマトンとする.

次に示す方法によって α の状態遷移図から対応する木 $T(\alpha)$ を導き, $T(\alpha)$ と $d(\alpha)$ とから α の 特徴集合 とよばれる $L(\alpha)$ の有限

部分集合 $C(\alpha)$ を決定できる [1]:

(1) $T(\alpha)$ の構成

1° Δ_0 を木の root node 名とし, 木の最上位におく. それから $\# \Sigma$ 本の枝を延ばして各枝を Σ の要素で名づけ, 枝 σ につながる node 名を $\delta(\Delta_0, \sigma)$ とする.

2° Δ をある node 名とする. Δ がこの node の左方または上位にある node の名前として既にあらわれているなら, いま着目している node からは枝を張らない. そうでないなら, $\# \Sigma$ 本の枝を延ばして各枝を Σ の要素で名づけ, 枝 σ につながる node 名を $\delta(\Delta, \sigma)$ とする.

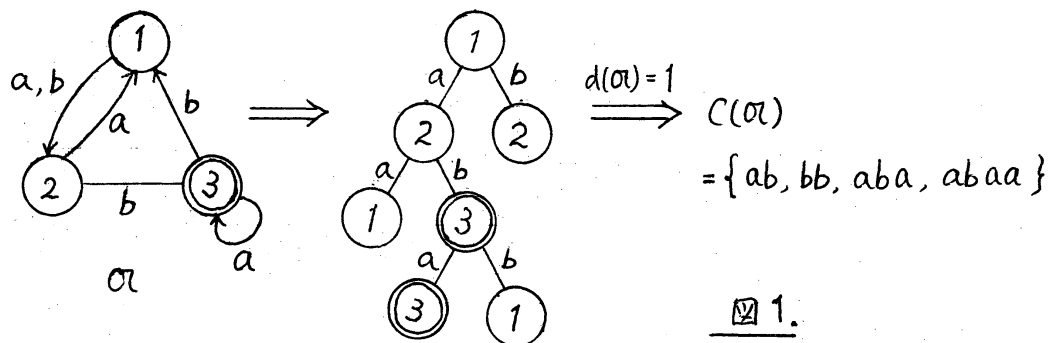
(2) $C(\alpha)$ の決定

$T(\alpha)$ において root node から任意の node Δ 至るまでの枝名を順に連接することによって得られる語の全体を $T^*(\alpha)$ とするとき,

$$C(\alpha) = \{ uv; u \in T^*(\alpha), |v| \leq d(\alpha), \delta(\Delta_0, uv) \in F \}.$$

上記方法の一例を図 1 に示す. ここに $S = \{1, 2, 3\}$,

$\Delta_0 = 1$, $F = \{3\}$ とする.



逆に $C(\alpha)$ と $d(\alpha)$ が既知であれば、もとの α を再構成できる。
これを一般に $C \subseteq \Sigma^*$, 整数 $k (\geq 0)$ に対する方法として次に与える:

いま, $g(w, C, k) = \{x; wx \in C, |x| \leq k\}$ とする。そのとき, 前記方法 (1) の 1° における $\Delta_0 \in g(\varepsilon, C, k)$, $\delta(\Delta_0, \sigma) \in g(\sigma, C, k)$ で, 2° における $\Delta \in g(w, C, k)$, $\delta(\Delta, \sigma) \in g(w\sigma, C, k)$ でおきかえることによって, まず C と k に対応する木 $T(C, k)$ が得られる。 $T(C, k)$ の各 node 名を状態名とすることにより, 最小有限オートマトン $\alpha(C, k)$ が容易に定まる。すなわち $\alpha(C, k) = (S, \delta, \Delta_0, F)$ において

$$S = \{g(x, C, k); x \in T(C, k)\}$$

$$\Delta_0 = g(\varepsilon, C, k)$$

$$F = \{g(x, C, k); \varepsilon \in g(x, C, k)\} \subseteq S$$

$$\delta: \delta(g(x, C, k), a)$$

$$= \begin{cases} g(xa, C, k); & x \in T^*(C, k) - T_*(C, k) \text{ のとき} \\ g(x'a, C, k); & x \in T_*(C, k) \text{ のとき} \end{cases}$$

である。ここに $T_*(C, k)$ は $T(C, k)$ において root node から leaf node へ至るまでの枝名を順に連接することによって得られる語の全体とし, $x' \in T^*(C, k) - T_*(C, k)$ かつ $g(x, C, k) = g(x', C, k)$, $a \in \Sigma$ とする。

上記方法は C が有限であるなしにかかわらず唯一つの $\alpha(C, k)$ を決定する。 $L(C, k) = L(\alpha(C, k))$ とするとき, さらに次の命題が成り

たつ。

命題. L が正則言語ならば $L = L(C, k)$ となるような $C \subseteq \Sigma^*$, 整数 $k (\geq 0)$ が存在する。

この命題は α を $L = L(\alpha)$ なる最小オートマトンとするとき, $C = C(\alpha)$, $k = d(\alpha)$ において成り立つが, 後に示すようにその場合だけに限らない。

§2. 正則言語族の階層化

$C \subseteq \Sigma^*$ と整数 $k (\geq 0)$ を任意に与えたとき, それらが有限オートマトン決定に及ぼす影響を考察する。

補題1. $w \in T^*(C, k)$ のとき $g(w, C, k) = g(w, L(C, k), k)$ 。

(証明) $x \in g(w, C, k)$ とする。 $wx \in T^*(C, k)$ の場合は $\varepsilon \in g(wx, C, k)$ より $wx \in L(C, k)$ 。 $\varepsilon \neq \varepsilon$ 故に $wx \notin T^*(C, k)$ とする。 ε のとき 次の (*) を満たす $w_1, w_2, \dots, w_{l-1} \in T^*(C, k) - T_*(C, k)$ が存在し, $\varepsilon \in g(w_{l-1}x_l, C, k)$ となる。 $\varepsilon \neq \varepsilon$ 故に, $l \geq 1$, $w_{l-1}x_l \in T^*(C, k)$ 。

$$(*) \begin{cases} x = x_1 x_2 \cdots x_l \quad (x_1, x_2, \dots, x_l \neq \varepsilon) \\ x_1 \in T_*(C, k), \quad x_2 \cdots x_l \in g(w_1, C, k) = g(x_1, C, k) \\ w_1 x_1 \in T_*(C, k), \quad x_3 \cdots x_l \in g(w_2, C, k) = g(w_1 x_1, C, k) \\ \vdots \\ w_{l-2} x_{l-1} \in T_*(C, k), \quad x_l \in g(w_{l-1}, C, k) = g(w_{l-2} x_{l-1}, C, k) \end{cases}$$

そのとき $\mathcal{O}(c, k)$ において

$$\begin{aligned}\delta(g(\varepsilon, c, k), wx) &= \delta(g(wx_1, c, k), x_2 \cdots x_\ell) \\ &= \delta(g(wx_2, c, k), x_3 \cdots x_\ell) = \cdots = \delta(g(w_{\ell-2}x_{\ell-1}, c, k), x_\ell) \\ &= g(w_{\ell-1}x_\ell, c, k) \in F.\end{aligned}$$

したがって $wx \in L(c, k)$, すなわち $x \in g(w, L(c, k), k)$.

以上から $g(w, c, k) \subseteq g(w, L(c, k), k)$. 逆も同様. $\square$

補題 1 により $T(c, k)$ と $T(L(c, k), k)$ とは全く同一の木となるので たちちに次の定理を得る.

定理 1. $L(c, k) = L(L(c, k), k)$.

補題 2. $w, w' \in T^*(c, k)$, $L(c, k) = C$ のとき

$$g(w, c, k+1) = g(w', c, k+1) \Leftrightarrow g(w, c, k) = g(w', c, k).$$

(証明) $(\Rightarrow)$ 明らか.

$(\Leftarrow)$ $g(w, c, k) = g(w', c, k)$ であって $x \in g(w, c, k+1) \Delta g(w', c, k+1)$ なる x が存在すると仮定する. 一般性を失うことなく $wx \in C$, $w'x \notin C$ とできる. $wx \in C = L(c, k)$ より $\mathcal{O}(c, k)$ において

$$\begin{aligned}\delta(g(\varepsilon, c, k), w'x) &= \delta(g(w', c, k), x) \\ &= \delta(g(w, c, k), x) = \delta(g(\varepsilon, c, k), wx) \in F.\end{aligned}$$

したがって $w'x \in L(c, k) = C$ となり矛盾を生ずる. $\square$

定理 2. $L(c, k) = C \Rightarrow L(c, k+1) = C$

(証明) 補題 2 より $T(c, k+1)$ は $T(c, k)$ と同型となる.

また, 任意の $w \in \Sigma^*$ に対して

$$\varepsilon \in g(w, C, k) \Leftrightarrow \varepsilon \in g(w, C, k+1)$$

だから, $\mathcal{L}(C, k+1)$ と $\mathcal{L}(C, k)$ とは 同等 になる. $\square$

上の定理は, C を固定して k を 0 から 1 ずつ 増やして いくとき, 一度 $\mathcal{L}(C, k)$ が C に 一致 すれば それ以降 は つねに C で あり 続ける ことを 示している.

さて $\mathcal{L}_k = \{ \mathcal{L}(C, k); C \subseteq \Sigma^* \}$ と しよう.

定理 3. $\mathcal{L}_k \subsetneq \mathcal{L}_{k+1}$

(証明) 任意の $\mathcal{L}(C, k) \in \mathcal{L}_k$ に対して 定理 1 から $\mathcal{L}(C, k) = \mathcal{L}(\mathcal{L}(C, k), k)$, 定理 2 から $\mathcal{L}(C, k) = \mathcal{L}(\mathcal{L}(C, k), k+1)$ を 得る. すなわち, $\mathcal{L}(C, k) = \mathcal{L}(C', k+1)$ なる $C' \subseteq \Sigma^*$ が 存在 する. ゆえに $\mathcal{L}_k \subseteq \mathcal{L}_{k+1}$.

次に $\mathcal{L}_k \neq \mathcal{L}_{k+1}$ を 示す. いま $C = \{ \sigma^{k+2} \} \ (\sigma \in \Sigma)$ と しよう. 明らかに $\min \{ |x|; x \in \mathcal{L}(C, k+1) \} = k+2$ である.

$\mathcal{L}(C, k+1) = \mathcal{L}(C', k)$ なる $C' \subseteq \Sigma^*$ が 存在 すると 仮定し, C' の 最小長 の 語 を w と しよう. $|w| \leq k+1$ ならば $w \in \mathcal{L}(C', k)$ だから $|w| < k+2$ となり 矛盾. したがって $|w| \geq k+2$.

そのとき $g(\varepsilon, C', k) = \emptyset$, 任意の $\sigma \in \Sigma$ に対して $g(\sigma, C', k) = \emptyset$ より $\mathcal{L}(C', k) = \emptyset$ となる. これまた 矛盾 を 導く. $\square$

この定理 3 によって 図 2 のような 正則 言語族 の 非負整数 k による 無限の 階層 が 与えられる.

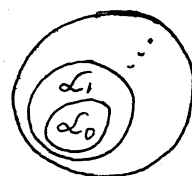


図 2.

§3. $\mathcal{L}_k$ の性質

定理3によって得られた階層に関連して以下の命題が成り立つ。

命題1. $\forall k (z_0), \exists n_k : \# \mathcal{L}_k \leq n_k$

(証明) $\Sigma^{(k)} = \{x; x \in \Sigma^*, |x| \leq k\}$ とするとき, 任意の $w \in T^*(C, k)$ に対して $g(w, C, k) \subseteq \Sigma^{(k)}$. すなわち, $\mathcal{O}(C, k)$ の状態の個数はたかだか $2^{\#\Sigma^{(k)}}$ である. $\square$

例1. $\Sigma = \{a\}$ のとき $\mathcal{L}_0 = \{\phi, \varepsilon, a^*, aa^*, (a^2)^*, a(a^2)^*\}$, $\#\Sigma = 2$ のとき $\#\mathcal{L}_0 = 26$.

命題2. $L \subseteq \Sigma^{(k)} \Rightarrow L \in \mathcal{L}_k$

(証明) $g(\varepsilon, L, k) = L$ より $L = L(L, k) \in \mathcal{L}_k$. $\square$

命題3. (1) $\mathcal{L}_k$ は complementation に関して閉じている.

(2) $\mathcal{L}_k$ は $\cup, \cap$, 連接に関して閉じていない.

(証明) (1) $g(w, \bar{C}, k) = \Sigma^{(k)} - g(w, C, k)$. したがって

$$g(w, \bar{C}, k) = g(w', \bar{C}, k) \Leftrightarrow g(w, C, k) = g(w', C, k).$$

すなわち, $T(\bar{C}, k)$ と $T(C, k)$ とは同型である. また

$$\varepsilon \in g(w, \bar{C}, k) \Leftrightarrow \varepsilon \notin g(w, C, k).$$

ゆえに $\overline{L(C, k)} = L(\bar{C}, k) \in \mathcal{L}_k$.

(2) 演算 $\circ$ に関して閉じているなら

$$\forall C_1, C_2 \subseteq \Sigma^*, \exists C : L(C_1, k) \circ L(C_2, k) = L(C, k).$$

定理1より $L(L(C_1, k) \circ L(C_2, k), k) = L(C_1, k) \circ L(C_2, k)$ となる.

(i) $\cup$ について: $C_1 = \{\varepsilon\}$, $C_2 = \{a^{k+1}\}$ とすれば $a^{2k+2} \notin L(C_1, k) \cup L(C_2, k)$, $a^{2k+2} \in L(L(C_1, k) \cup L(C_2, k), k)$.

(ii) $\cap$ について: $C_1 = \{a^{k+1}, a^{k+2}\}$, $C_2 = \{\varepsilon, a^{k+2}\}$ とすれば $a^{k+2} \in L(C_1, k) \cap L(C_2, k)$ だが, $L(L(C_1, k) \cap L(C_2, k), k) = \phi$.

(iii) 連接について: $C_1 = C_2 = \{a^{k+1}\}$ とすれば $a^{2k+2} \in L(C_1, k) \cdot L(C_2, k)$ だが, $L(L(C_1, k) \cdot L(C_2, k), k) = \phi$. □

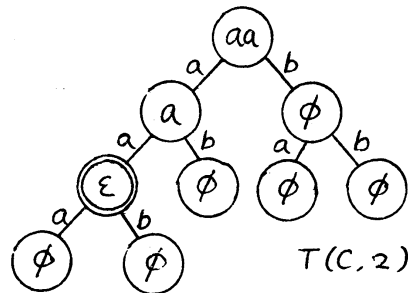
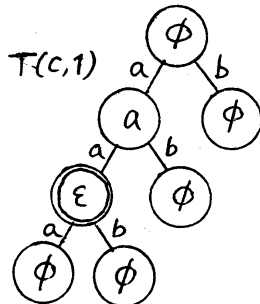
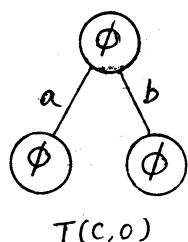
正則言語 (有限オートマトン) の学習という立場からすれば, L からの いくつかの サンプル 集合と k の値の 組み合わせによってもとの L を同定できるかという問題がある. この問題に関連して以下の命題が成り立つ.

いま, $m(C) = \min \{|x|; x \in C\}$, $M(C) = \max \{|x|; x \in C\}$, $d(C) = \min \{k; C \subseteq L(C, k)\}$, $D(C) = \min \{k; C = L(C, k)\}$ としよう.

命題 4. (1) $m(C) - 1 \leq d(C) \leq M(C) - 1$

(2) $D(C) \leq M(C)$

例 2. $C = \{aa\}$ のとき 図 3 より $L(C, 0) \neq C$, $L(C, 1) \supsetneq C$, $L(C, 2) = C$. したがって $d(C) = m(C) - 1 = M(C) - 1$, $D(C) = M(C)$.



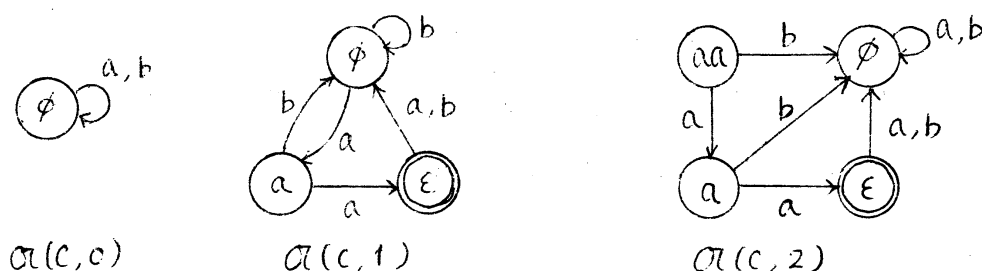


図 3. $C = \{aa\}$ による学習

命題 5. $\forall k (\geq 0), \exists C \subseteq \Sigma^* : d(C) = D(C) = k$

命題 6. σ を最小有限オートマトンとし, $L = L(\sigma)$ とする. $C \subseteq L$

のとき,

$$C(\sigma) \subseteq C \iff \exists k : L(C, k) = L$$

(証明) ($\Rightarrow$) $L(C(\sigma), k) = L$ とする. $C(\sigma) \subseteq C \subseteq L$ より

$$\forall w \in \Sigma^* : g(w, C(\sigma), k) \subseteq g(w, C, k) \subseteq g(w, L, k).$$

$k=3$ で $w \in T^*(C(\sigma), k)$ のとき 補題 1 から

$$g(w, C(\sigma), k) = g(w, L(C(\sigma), k), k) = g(w, L, k).$$

すなわち, $g(w, C(\sigma), k) = g(w, C, k)$. したがって $T(C, k)$ と

$T(C(\sigma), k)$ とは同一である. ゆえに $L(C, k) = L$.

($\Leftarrow$) $L(C, k) = L, L(C(\sigma), k) = L$ とする. $\sigma(C, k)$ と $\sigma(C(\sigma), k)$

とは同等な最小オートマトンとなるから $T(C, k)$ と $T(C(\sigma), k)$

とは同型で 最終状態も同じ位置にある. したがって

$$C(\sigma) = \bigcup_{x \in T^*(C, k)} x \cdot g(x, C, k) \subseteq C.$$



<注> $C(\sigma)$ は $L(\sigma)$ を復元できる $L(\sigma)$ の部分集合のう

ちで 最小のものといえる。

命題 7. 任意の 整数 $k (≥ 0)$, $C ⊆ Σ^*$ に対して, $L(C, k)$
 $= L(C', k)$ なる $C' ⊆ Σ^*$ は 無数に 存在する。

参考文献

- [1] Biermann, A. W., An interactive finite-state language learner, 1st USA-JAPAN Comp. Conf. Proc. (1972), 13-20.