数理解析研究所講究録
229巻 1975年 182-192



# 整合関係を考慮した記号列パターン識別オートマトンの構成および適応的修正

東京工業大学　工学部　槇本　肇
富田悦次

## 1. まえがき

系列パターン認識システムに関する研究の一環として，我々はこれまでに，カテゴリ名の付与された有限のサンプル記号列が与えられたときに，これらを正しく識別するようなオートマトンを構成し，さらに必要に応じて，それを教師との質問応答によって，適応的に修正してゆくアルゴリズムを考えてきた。本稿においては，簡単のために，識別機構を決定性有限オートマトンに限って，これらの方法をまとめて記した。

## 2. 諸定義・記法

[1] 決定性有限オートマトン（以下では，単にFSAと記す）を $M=(K, \Sigma, \delta, S_0, F)$（$K$：状態の集合，$\Sigma$：入力記号の集合，$\delta: K\times\Sigma\to K$への写像，$S_0$：初期状態，$F$：最終状態の集合）で表す。

〔2〕 FSA $M$ において、$\delta(S_1, a) = S_2 \in \delta$ であるとき、この写像規則による推移を、$(S_1) \xrightarrow[M]{a} (S_2)$ と表す。さらに、$i = 1, 2, \cdots, m$ に対して $(S_i) \xrightarrow[M]{a_i} (S_{i+1})$ なる推移が可能であるときには、$(S_1) \xrightarrow[M]{x} (S_{m+1})$ ただし、$x = a_1 a_2 \cdots a_m$ と表す。さもなくば、$(S_1) \xrightarrow[M]{x} (S_{m+1})$ と表す。

入力記号列 $u$ に対し、$M$ の決定性推移 $(S_0) \xrightarrow[M]{u} (P)$ によって一意的に定まる状態 $P$ を、$C_M(u)$ で表す。

〔3〕 入力記号には、適当にアルファベット順序を設ける。入力記号列の間には、短いものから長いものへと順に、また同じ長さのものどうしでは、初めて入力記号の異る位置における入力記号のアルファベット順に従ってその辞書式順序を定める。このような入力記号列集合 $\pi$ のうちで、最も辞書式順序が先行する記号列を $First(\pi)$ で表す。

## 3. 整合関係を考慮した記号列パターン識別FSAの構成[1]

### 3-1 サンプル記号列集合に対する前提

ある正規集合をなすパターン全体の各カテゴリから、いくつかの有限長、有限個のサンプル記号列が抽出されているとき、それらのサンプルから元のパターン集合全体を推論する方法を考える。ここで、元のパターン集合の記号列は終止記号(#)付きのものであり、異ったカテゴリ集合間の重なりはなく、その集合を支配する規則は、最終状態名の区別によって

カテゴリを識別するようなFSA $M_D=(K_D, \Sigma_D, \delta_D, S_{0D}, F_D)$ として記述されているとする。

このような推論の第一近似として，先ず少くともそれらのサンプルだけは正しく識別するようなFSA $M'$の構成を計る。ところで，このようなサンプルの間には，その取得過程に応じて何らかの関係が生じていることが普通であり，それに関する予備知識も導入することによって，より正しい推論を達成することができる。ここで本節においては，サンプル記号列集合は，次の条件を満足するようなものとなっていることが予めわかっている場合を扱うことにする。

［サンプルに対する前提条件］カテゴリ$i$に属するサンプル記号列の集合を$\pi_{ai}$としたとき，$u\nu \in \pi_{ai}$であるならば，必ず$u \cdot First\{\nu' | C_{M_D}(u) \xrightarrow[M_D]{\nu'} (F_i), F_i \in F_D$はカテゴリ$i$用の最終状態$\} \in \pi_{ai}$である。すなわち，ある状態からの推移を示す記号列としては，最も辞書式順序の先行するものをさしおいて，それよりも順序が後の記号列だけがサンプル中に採られているというようなことはない。

### 3-2 部分記号列の整合関係を考慮したFSA構成．

前項の前提条件を満足するようなサンプル記号列の集合に対して，それらの部分記号列間の整合関係を解析しながらFSAの構成を行う方法を示す。

先ず、サンプルに対する前提条件のいかんにかかわらず、次の関係が成立する。

[命題1] (i) $M_D$において、$F_i, F_j$をそれぞれカテゴリ $i, j$ ($i \neq j$) 用の最終状態としたとき、$F_i \neq F_j$。(ii) サンプル記号列中の2箇所で共通な部分記号列 $v$ が存在し、$(S_h) \overset{v}{\underset{M_D}{\Longrightarrow}} (S_k)$, $(S_{h'}) \overset{v}{\underset{M_D}{\Longrightarrow}} (S_{k'})$であるとしたとき、$S_k \neq S_{k'}$ ならば、$S_h \neq S_{h'}$」

さらに、サンプルに対する前提条件のもとでは、

[命題2] サンプル記号列の prefix $u_1, u_2$ に対して、$First\{v_1 | u_1 v_1 \in \pi_{ai}\} \neq First\{v_2 | u_2 v_2 \in \pi_{ai}\}$ for some $i$ であるならば、$C_{M_D}(u_1) \neq C_{M_D}(u_2)$」

この2命題中の $M_D$ を $M'$ で置き換えて得られるような関係は、$M_D$ とできるだけ近いものとして構成しようとする FSA $M'$ においても当然満足されるべき関係であり、FSA構成におけるこのような条件を、<u>suffix整合条件</u>と名付ける。特に、命題2に対する条件は、<u>前提付きsuffix整合条件</u>と呼ぶ。

FSAの構成の仕方としては、サンプル記号列において、その先端位置よりの辞書式順序が後の方の位置から順に、suffix整合条件に反しない限りにおいて逐次状態の統合を計ってゆくという方法をとる。このような統合の可能性は一般に幾通りか考えられるが、ある段階でその可能性を調べるのは $k$ 個先の範囲内のものまでに留め、そこで統合が不可能で

あるようなものに対しては新しい状態を割り当てる。[2] なお、前提付きsuffix×整合条件を考慮せずに，$k=\infty$ あるいは $k=0$ とおいた場合は，文献(3)の最小状態法，および(単純)階層法に相当する。

前述のサンプルに対する前提条件は，いつの場合においても完全に満たされているという保証は勿論ないが，ほぼ満足されているような場合に対しては，この前提におけるのと同様にしてFSAの構成を行い，然る後にその前提の妥当性をチェックし，必要な修正を行うようにしてもよい。(次節参照) また，サンプルの母集団が正規集合でない場合にも，この構成法は，推論の近似的解を与える。

## 4. FSAの適応的修正

前節で得られるようなFSAは，一般にはまだ誤りを含んでいる可能性があるため，本節では，それをさらに適応的に修正して目的のものへと到達させる方法を示す。なお，以下では簡単のために対象とするFSAは非冗長であり，パターンの(受理)カテゴリは単一であるとする。ここで，教師に対してある特定の記号列がそのカテゴリに属すものであるか否かの質問を発したときには，その答は誤りなく与えられるものとする。

### 4.1 FSAの代表記号列集合[4],[5]

受理または非受理の指示（それぞれ，○印．×印で表す）が付けられた有限のサンプル記号列集合$\pi_c(M)$に対して．それらを正しく識別するようなＦＳＡは一般にいくつも構成されるが．構成において満足されるべきある基準を設けたとき．そのもとでは特定のFSA M（あるいは．それと同形のもの）が一意的に対応付けられるとき．記号列集合$\pi_c(M)$は．その基準のもとではMの特性を完全に代表して表現しているとみなすことができ．このような$\pi_c(M)$をMの代表記号列集合と名付ける。ここで、FSA M′の構成の基準は次のように定める。

［ＦＳＡ構成基準］　○$w \in \pi_c(M)$ならば．$(S_0) \overset{w}{\underset{M'}{\Rightarrow}} (S_f')$，$S_f'$はM′の最終状態。×$\overline{w} \in \pi_c(M)$ならば．$(S_0) \overset{\overline{w}}{\underset{M'}{\Rightarrow}} (S_f')$であり．かつ．前節の前提付きsuffix整合条件を満足するような．状態数および写像規則数が最少のFSA M′を構成する。（たとえば．前節で$k=\infty$とした場合の構成法によればよいが．具体的構成法は問わない）」

以下に．代表記号列集合の求め方を示す。

［基本記号列集合　$\pi_F(M)$］　$\pi_F(M)=\{u\,a\,v \mid (S_0) \overset{u}{\underset{M}{\Rightarrow}} (P) \overset{a}{\underset{M}{\rightarrow}} (Q) \overset{v}{\underset{M}{\Rightarrow}} (S_f),\ \delta(P,a)=Q \in \delta,\ u=First\{u' \mid (S_0) \overset{u'}{\underset{M}{\Rightarrow}} (P)\},\ v=First\{v' \mid (Q) \overset{v'}{\underset{M}{\Rightarrow}} (S_f)\}\}$

［副次記号列集合　$\pi_S(M)$］　$u_i v'$，$u_j v'' \in \pi_F(M)$，$First\{v' \mid C_M(u_i) \overset{v'}{\underset{M}{\Rightarrow}} (S_f)\} = First\{v'' \mid C_M(u_j) \overset{v''}{\underset{M}{\Rightarrow}} (S_f)\}$であり．か

つ．$C_M(U_i) \neq C_M(U_j)$ であるとする。$(\exists w_0)[w_0 = First\{w \mid C_M(U_i) \overset{w}{\underset{M}{\Rightarrow}} (S_f), C_M(U_j) \overset{w}{\underset{M}{\Rightarrow}} (S_f)\}]$ であるならば、$\circ U_i w_0 \in \pi_{Sa}(M)$, $\times U_j w_0 \in \pi_{Sr}(M)$。さらに、$C_M(U_i) \overset{w'}{\underset{M}{\Rightarrow}} (Q) \overset{w''}{\underset{M}{\Rightarrow}} (S_f)$, $w_0 = w'w''$ なる任意の状態 Q に関して、$w_0'' = First\{w'' \mid (Q) \overset{w''}{\underset{M}{\Rightarrow}} (S_f)\}$ としたとき、$\circ U_i w' w_0'' \in \pi_F(M) \cup \pi_{Sa}(M)$ とする。$i, j$ の立場を逆にした場合も同様。以上の操作を、前記関係にある基本記号列中のあらゆる prefix $U_i, U_j$ に対して適用した結果得られる記号列集合 $\pi_{Sa}(M), \pi_{Sr}(M)$ の和集合を副次記号列集合 $\pi_S(M)$ とする。」

$\pi_F(M) \cup \pi_S(M) = \pi_C(M)$ とおいたとき、この $\pi_C(M)$ は、前節の，サンプル記号列に対する前提条件を満足している。

［定理］　$\pi_C(M) = \pi_F(M) \cup \pi_S(M)$ から前記構成基準に従って得られる FSA M′ は、もとの FSA M と同形である。それ故、この $\pi_C(M)$ は、前記構成基準のもとで、M の代表記号列集合をなす。

［例題 1］　図 1 の FSA $M_D$ に対して、

$\pi_F(M_D) = \{\circ aab\#, \circ aacb\#,$
$\circ aabbb\#, \circ aabdb\#\}$

$\pi_S(M_D) = \{\circ aac \cdot acb\#, \times aabb \cdot acb\#,$
$\times aabd \cdot acb\#\}$

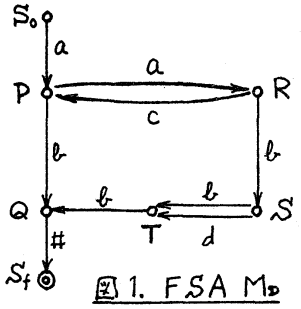

図 1. FSA $M_D$

## 4-2 代表記号列集合を利用したFSAの適応的修正[6]

最初に与えられたFSAを$M_0$としたとき、前項に示したように、代表記号列集合$\pi_C(M_0)$はその構成基準のもとにおいて$M_0$を完全に表現しており、$M_0$が誤りをもっている場合には、目的とするFSA $M_D$に対する代表記号列集合$\pi_C(M_D)$とは異ったものとなっている。従って、FSAの適応的修正の過程は、先ず教師への質問によって、$\pi_C(M_0)$の記号列に対する受理・非受理の指示を全て正しい応答に訂正したような集合$\pi_{C_D}(M_0)$を求め、それでも$\pi_C(M_D) \neq \pi_{C_D}(M_0)$である場合には、さらに積極的に$\pi_C(M_D) = \pi_F(M_D) \cup \pi_S(M_D)$の要素として不足している可能性のある記号列を、$\pi_C(M_D)$の各要素の定義に応じて生成し、その応答を教師に質問して新しい確認情報を得、$\pi_C(M_D) - \pi_{C_D}(M_0)$を逐次的により小さくしてゆく過程として考えることができる。

以下、具体例に沿って、このような考えに従った適応的修正の過程を示す。（一般アルゴリズムは、文献(6)参照）

【例題2】 最初に与えられたFSAは、図2の$M_0$，目的とするFSAは図1の$M_D$とする。

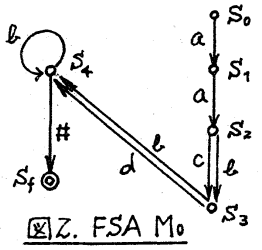

図2. FSA $M_0$

[0] $\pi_F(M_0) = \{$oaabb#, oaacb#, oaabd#,
oaabbb#$\}$

$\pi_S(M_0) = \phi$

従って、$\pi_{C_D}(M_0) = \pi_{C_{DA}}(M_0) \cup \pi_{C_{DR}}(M_0)$

ただし、$\pi_{C_{DA}}(M_0) = \{\circ aacb\#, \circ aabbb\#\}$,

$\pi_{C_{DR}}(M_0) = \{\times aabb\#, \times aabd\#\}$

$\pi_{C_{DA}}(M_0)$の記号列だけでは、まだ$(S_3) \xrightarrow[M_0]{d} (S_4)$の路が未通過であるので、これを通過するような受理記号列を探すと、aabd・#, aabb・b#のprefix, suffixの連接より成るaabd・b#が見出せ、これを基本記号列の補充分$\pi_F'(M_0)$とする。

さらに、以後のFSA構成時においては、前提付きsuffix整合条件を常に考慮に入れることにするため、その妥当性チェックを行う。すなわち、$\pi_{C_{DA}}(M_0) \cup \pi_F'(M_0)$中の記号列で、そのsuffixをより辞書式順序の先行するsuffixで置換したものが新たな受理記号列とならないかを調べると、oa・ac・b#に対して、a・b#が受理であることがわかり、これを$\pi_A(M_0)$とおく。

以上の$\pi_{C_D}(M_0) \cup \pi_F'(M_0) \cup \pi_A(M_0)$より前項の構成基準に従ってFSAを構成すると、図3の$M_1$を得る。

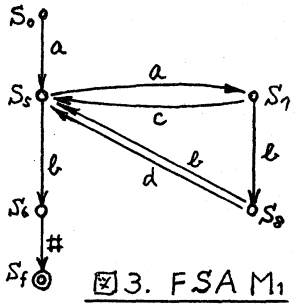

図3. FSA $M_1$

[1] $\pi_{C_D}(M_1) = \{\circ ab\#, \circ aacb\#, \circ aabbb\#,$

$\circ aabdb\#\}$

複数の入力推移のある状態$S_5$において、aac・b#, aabb・b#, aabd・b# $\in \pi_F(M_1)$の共通suffix b#における推移が同じとなっていることが正しいかどうかのチェックとして、副次記号列の定義に応じて、記号列aac・acb#,

aabb·acb#, aabd·acb# を生成して，その正しい応答を質問すると，これらはそれぞれ，受理，非受理，非受理である。従って，$\pi_S'(M_1)=\{\circ aacace\#, \times aabbacb\#, \times aabdacb\#\}$ を副次記号列集合の補充分とする。ここで，$\pi_{CD}(M_1)\cup\pi_S'(M_1)$ は $\pi_C(M_D)$ と一致し，基準に従った構成によって図1のFSA $M_D$ が得られ，修正は完了。

## 5. あとがき

以上の方法は，ある種の決定性プッシュダウンオートマトンに対して拡張することも可能である。[1],[4],[6] また，これまでのオートマトンでは，最終的な受理または非受理の応答を考えているだけであるが，さらに推移の途中における出力応答も考えて，それによって自己の応答，あるいは他の部分構造のふるまいを制御し合うような機構への拡張も重要と考えられる。[7]~[9]

最後に，本研究に対して熱心に御討論，協力をいただきました堂下修司教授（現，京都大学），米崎直樹君はじめ，当研究室の皆様に感謝いたします。また，本研究の一部は，文部省科学研究費の援助を受けていることを記し，謝意を表します。

## 文　献

(1) 榎本，富田，米崎；"サンプル記号列の整合関係を考慮したオートマトンの一構成法"，信学会オートマトン研資（1973-11）

(2) 榎本,富田,堂下,牧野;"サンプル記号列を識別する決定性FSAの構成法-k階層法-" 昭48.信学全大 1313 (1973-03)

(3) 榎本,堂下,富田,山口;"句構造サンプルパターンを識別するオートマトンの構成について" 信学会オートマトン研資 (1970-03)

(4) H.Enomoto, E.Tomita and S.Doshita; "Synthesis of automata that recognize given strings and characterization of automata by representative sets of strings." 1-st USA-Japan Computer Conf. Proc. P.21 (Oct. 1972)

(5) A.W.Biermann; "An interactive finite-state language learner": ibid. P.13

(6) 榎本,富田,堂下.; "オートマトンのチェック修正用記号列の生成", 信学会オートマトン・パターン研資 (1973-01)

(7) 榎本,富田,小谷野; "一般化順序機械の代表入出力記号列集合" 昭49.信学全大 1491 (1974-07)

(8) 榎本,富田,根岸; "相互作用有限オートマトンシステムのテキスト表現" 昭49 信学全大 1492 (1974-07)

(9) 阿部; "記号系列パターンを識別するあるオートマトンの構成法について." 信学会パターン研資 (1974-05)