

資料

パターンの学習識別アルゴリズム*

田 中 幸 吉**

Abstract

Generally speaking, the pattern recognition techniques must perform two basic functions, that is, the process of characterizing a class of the common pattern of inputs that belong to the same class and the process of classifying any input as a member of one of several classes.

The present paper shows the synthesized and coordinated résumé on the pattern classification techniques with learning procedures focusing around the present author's works.

The proposed techniques are expected to be available not only to temporal patterns but also to spatial patterns.

まえがき

パターンには音声波、地震波、脳波、心電図等の時間的パターンと、文字、景色、写真、物体等の空間的パターンとがある。何れにしろパターンを認識するためにはパラメータの集合論的・位相論的記述か、あるいは統計量とかある関数による写像値をもって数量化記述 (hypergeometrical description) するかによってパターンの抽象化された構造記述 (いわゆる前処理・特徴抽出) を行い、どのテンプレートあるいはシンボルとか概念または概念のカテゴリに該当するかを識別決定する必要がある。前者の特徴抽出はパターン認識の最重要問題であり、パターン認識合同国際会議において2年連続してこれに関連するパネル討論¹⁾が行われた。後者は比較的系統立った理論構成が可能である。然しながら本来前処理・特徴抽出、識別決定のように分けられるものではなく全体的に把握しなければならない性質のものである。また特徴抽出においても学習とか対話の技法導入の必要な場合もある***。更にまた空間的パターンにおいてもスペクトル的あるいは時間的特徴を手がかりに認識を進める場合もある²⁾。

画像処理における画像品質改善に学習型フィルタリングの手法を用いることがある³⁾。

本資料においては空間的とか時間的とかの区別なく、一般的パターンの学習識別アルゴリズムについて筆者等が従来から行ってきた研究成果を中心として(多少古い感があるが、リモートセンシング、画像処理等においても応用されようとしているこの種の成果を)系統的・類別的に整理した理論体系を以下に述べる。

1. パターンの学習識別アルゴリズムの分類

学習機能をもったパターン識別アルゴリズムは、教師ありの学習形式と教師なしの学習形式とがある。教師ありの学習形式においては、学習パターンの情報と教師が教える正解、すなわち学習パターンのカテゴリに関する情報を用いて、各カテゴリに属するパターンの確率的構造を推測し、識別機構を逐次学習的に構成する。然しながら教師がない場合には、学習パターンのみから、その属するカテゴリの推測決定をしなければならない。すなわちパターンの形成する空間の構造のみがカテゴリ決定の手がかりとなる。したがって教師なしの学習アルゴリズムについては、パターンの分布形のわかっているパラメトリックな学習形式が主に研究されてきた。

教師なしの学習形式には Decision-directed Algorithm^{4),5)}のように教師の代わりに機械自身の決定を用いる方法、あるいはパターンの混合分布を学習パター

* Pattern Classification Algorithms with Learning Procedure by Kokichi TANAKA (Faculty of Engineering Science, Osaka University)

** 大阪大学基礎工学部情報工学科

*** RADC (Rome Air Development Center) の WPS (Waveform Processing System), OLPARS (On-line Pattern Analysis & Recognition System), Purdue 大学の INTERSPACE, IBM の IBM Interactive Sys. 等がある。

ンより推測し、それを各カテゴリの分布に分解していくことによって識別アルゴリズムを構成していく方法がある。前者の方法は機械自身が正解を与えるならば最適な決定となるという意味で準最適決定である。また後者においては、混合分布から各カテゴリの分布に分解するときの一意性 (identifiability) の仮定が必要とされる⁶⁾。

ノンパラメトリックな学習においては、予め定めておいた識別関数を学習パターンによって逐次修正していくが、教師なしのノンパラメトリック学習アルゴリズムとしては Jakowatz⁷⁾、田中⁸⁾がそのモデルとアルゴリズムを提案している。然しながら学習の収束性に対する証明は特別な場合を除いて未だ解決されていない⁹⁾。また教師なしのクラスタ化の方法¹⁰⁾がある。すなわち同一カテゴリに属するパターンはクラスタをなすという仮定のもとにクラスタ間に導入したある種の距離を用いてクラスタ化の評価関数を作り、その値を最小にするようにパターン空間の分割を行う方法である。

また第1主成分を利用した教師なしのノンパラメトリックな学習アルゴリズムを確率近似法のアプローチにより求める方法¹¹⁾がある。

2. ノンパラメトリック型学習識別アルゴリズム

2.1 確率近似法と相関型学習識別機構

ノンパラメトリック法の思想的底流をなすものは古くからあった逐次近似の考え方である。統計学の分野におけるそれは Robbins, Monro 等により確率近似法として定式化され Wolfowitz, Dvoretzky¹²⁾等により一般化された。確率近似法は特に固定したパラメータ、ひいては固定したパターンの推定に用いられる。

未知パターンを $x(t)$ 、第 n 期間における相加雑音を $n_n(t)$ とし、第 n 期間における観測パターンを

$$y_n(t) = \theta_n x(t) + n_n(t) \quad (2.1)$$

とする。ここに θ_n は 0 または 1 をとる 2 値確率変数である。このとき識別関数を

$$g(y_n(t)) = C_n^T \phi(y_n(t)) \quad (2.2)$$

とする。ここに ϕ は適当な直交関数ベクトルであり、 C_n は重みベクトルである。いま評価関数を次式で与える。

$$J(C_n) = \frac{1}{2} E y_n \cdot \theta_n (g(y_n) - d(y_n, C_n))^2 \quad (2.3)$$

$$d(y_n, C_n) = \begin{cases} 1, & g(y_n) \geq 0 \text{ のとき} \\ 0, & g(y_n) < 0 \text{ のとき} \end{cases} \quad (2.4)$$

このとき確率近似法による学習識別アルゴリズムは次式で与えられる。ここに $E y_n \cdot \theta_n$ は y_n, θ_n に関する集合平均 (期待値) を表わす。

$$C_{n+1} = C_n - \gamma_{n+1} (C_n^T \phi(y_n) - d(y_n, C_n)) \phi(y_n) \quad (2.5)$$

このとき C_n は適当な条件のもとに J を最小にする C^* に平均 2 乗収束、および確率 1 をもって収束する。然しながら (2.3) 式の $J(C_n)$ は教師なし評価基準であるため、 C_n は $\phi(x(t))$ と完全には整合しないが、SN 比が大きいたまにはほぼ整合状態となる¹³⁾。

前記の直交関数ベクトル ϕ として Laguerre 関数を取り、 γ_n 一般に定数とし、非同同期型未知パターンの学習識別装置として実現したものが適心型相関フィルタである⁸⁾。このフィルタは非同同期型であるが、パターンの継続時間が未知の場合でも自動的にその終端を見出す動作をすることは興味深い⁹⁾。

このアルゴリズムは脳波の K-波、心電図の QRS-波¹⁴⁾、地震反射波等の認識に適用可能である。

2.2 教師付きノンパラメトリック学習識別機構

この方法の一般的適用においては識別関数としては線形、双 1 次形、区間的線形などが考えられる。線形学習機構は線形識別関数を用いる場合に相当する。

線形識別関数は増大パターンベクトル Y と重みベクトル W を用いて

$$g(X) = Y \cdot W \quad (2.6)$$

で与えられる。 $g(X) = 0$ となるような W の値の軌跡として重み空間内の 1 つの超平面が得られ、この超平面により $g(X)$ の正、負による X のクラスを 2 分するよう教育パターンを用いて重みベクトルを決定することが教師付きノンパラメトリック線形識別機構である¹⁵⁾。

2 クラスでなく多クラス識別についても提案されているが、いずれも識別するクラスの個数に応じて可変荷重素子の数が増える欠点がある。文字識別のような場合、識別対象はある標準的パターンとその変形とから構成されることを利用し、一種のマスキング機構をもたせることにより荷重素子の数を減らすことができる¹⁶⁾。パーセプトロン、アダライン型の機構についても筆者等の研究があるが、この種の機構は実用的パターン認識の見地からは適当とは思われないので省略する¹⁷⁾。

ポテンシャル関数法は未知の様な有界連続関数の逐次近似として用いられる。ポテンシャル関数としては次の 2 種類が考えられる。

$$(i) \quad K(X, Y) = e^{-AR^2}$$

ただし $R^2 = \|X - Y\|^2, A > 0$

$$(ii) K(X, Y) = \sum_{m=1}^M \lambda_m^2 \phi_m(X) \phi_m(Y)$$

ただし $\{\phi_m\}, m=1, 2, \dots$ は多次元直交関数族であり

$$\phi_m(X) = \phi_{i_1}(x_1) \phi_{i_2}(x_2) \dots \phi_{i_k}(x_k)$$

$$M = \frac{(k+s)!}{k!s!}, K = X \text{ の次元数}$$

$$s = i_j - 1$$

例えば N 個のサンプルパターンを用いた学習により分布関数 $\hat{p}(X|\omega_i)$ を推定するアルゴリズムはポテンシャル関数 (i) を用いて

$$\hat{p}_N(X|\omega_i) = \frac{1}{N} \sum_{n=1}^N K(X, X_n) \quad (2.7)$$

で与えられる。ただし X_n はカテゴリ ω_i に属する第 n 番目の学習パターンを示す。また非線形識別関数

$\phi(X) > 0$ ただし $X \in \omega_1; \phi(X) < 0$ ただし $X \in \omega_2$ の学習アルゴリズムは第 $(r+1)$ 回目学習期間におけるサンプルパターンを X_{r+1} とすれば

$$\left. \begin{aligned} \phi_r(X_{r+1}) > 0 \text{ ただし } X_{r+1} \in \omega_1 \\ \phi_r(X_{r+1}) < 0 \text{ ただし } X_{r+1} \in \omega_2 \end{aligned} \right\} \text{ならば } \phi_{r+1} = \phi$$

$$\left. \begin{aligned} \phi_r(X_{r+1}) < 0 \text{ ただし } X_{r+1} \in \omega_1 \text{ ならば} \\ \phi_{r+1}(X_{r+1}) = \phi_r(X) + K(X, X_{r+1}) \\ \phi_r(X_{r+1}) > 0 \text{ ただし } X_{r+1} \in \omega_2 \text{ ならば} \\ \phi_{r+1}(X) = \phi_r(X) - K(X, X_{r+1}) \end{aligned} \right\}$$

(2.8)

で与えられる。米パデュュー大学の LARS (Laboratory for Applications of Remote Sensing) はリモートセンシングによって得たマルチスペクトル画像データから地上の畑の作物の種類を識別するシステムを開発した¹⁸⁾。

地上の各点から反射する電磁波エネルギーを 13~24 のスペクトルバンドに分けて観測すると、13~24 次元の入力パターンベクトルが得られる。各スペクトルバンドにおけるエネルギーの分布関数を式 (2.7) で与えられるポテンシャル関数法により推定し、その分布関数を用いて最尤推定により識別決定する方式 (方式 I) と式 (2.8) で与えられる学習アルゴリズムにより非線形識別関数を推定する方式 (方式 II) を、分布関数をガウス型と仮定したパラメトリック学習認識アルゴリズムであるベイズ推定による識別方式 (方式 III) と 912 個の学習用サンプルパターンを用いて比較実験した結果が次の Table 2.1 である。

また 205 個の大豆とトウモロコシの混合したサンプルについて行った結果が Table 2.2 である。Table

Table 2.1 Comparison Example of Recognition Rate depending on Classification Algorithms¹⁹⁾

正解率		識別決定の正解率
方 式		
方 式 I		98.3%
" II		97.5%
" III		81.7%

Table 2.2 Experimental Example¹⁹⁾

サンプル数 N	誤り 個 数		蓄積サンプル数		所要計算時間	
	方式 I	方式 II	方式 I	方式 II	方式 I	方式 II
60	44	22	60	9	15.6 sec	13.5 sec
60	23	17	60	11	20.8	13.3
100	16	13	100	15	26.0	16.7
150	14	14	150	16	39.6	25.4
250	0	0	205	33	53.2	34.2
(混合パターン) 98	12	11	98	16	25.5	16.4

2.1, 2.2 からわかるように、方式 II は方式 III より成績はよいが、記憶容量と時間が余計にかかる。また方式 I が最良であるが方式 II 以上に記憶容量と時間を要する。

3. パラメトリック型学習識別アルゴリズム

3.1 時変パターンの教師付き学習識別機構

ガウス雑音 N_n の相加した第 n 期間の観測パターンを、

$$Y_n = \theta_n X_n + N_n$$

とする。ここに 2 値確率変数 θ_n は $k \neq n$ なる θ_k と独立であり、 $\theta_n = 1$ となる事前確率は p で、 $\theta_n = 0$ となる確率は $1-p$ とする。また L 次元相加雑音ベクトル N_n は $k \neq n$ なる N_k と独立なガウス分布に従い、その平均は Q 、分散は Σ_n (正定) とする。すなわち

$$N_n \sim \mathcal{N}(Q, \Sigma_n) \quad (3.1)$$

L 次元パターンベクトル X_n は平均 M_n 、分散 Φ_n なるガウス分布をなす l 次自己帰帰過程に従うものとする。すなわち

$$X_n \sim \mathcal{N}(M_n, \Phi_n) \quad (3.2)$$

$$\sum_{j=0}^l a_{nj} (X_{n-j} - M_{n-j}) = Z_{n-1} \quad (3.3)$$

ここに Z_n はガウス確率変数であり

$$a_{nj} = Q, n \leq j \leq l$$

$$a_{n0} = I_L (L \times L \text{ 単位マトリックス})$$

$$E(Z_n) = Q$$

$$E(X_{n-j} Z_n^T) = Q, j = 0, 1, 2, \dots$$

$$E(Z_n Z_n^T) = Q, n \neq k$$

$$= \Psi_n, n = k$$

なお右肩の T は転置を表わし、 $E(\cdot)$ は期待値をあらわす。

$$\eta_n = [Z_n, Q, \dots, Q]^T \quad (lL \times 1 \text{ ベクトル})$$

$$Y_n = \begin{bmatrix} X_n - M_n \\ X_{n-1} - M_{n-1} \\ \dots \\ X_{n-l+1} - M_{n-l+1} \end{bmatrix} \quad (lL \times 1 \text{ ベクトル})$$

$$G_n = \begin{bmatrix} -g_{(n+1)1} & -g_{(n+1)2} & \dots & -g_{(n+1)l} \\ I_L & Q & \dots & Q \\ Q & & I_L & \dots \\ \vdots & & \vdots & \vdots \\ Q & Q & & I_L \end{bmatrix} \quad (lL + lL \text{ マトリックス})$$

$n=l, l+1, \dots$

とおけば、式(3.3)は

$$Y_n = G_{n-1} Y_{n-1} + \eta_{n-1} \quad (3.4)$$

となる。教育系列 $\theta^{n+1} = \{\theta_1, \theta_2, \dots, \theta_{n+1}\}$ (つまりパターンの有無に関する教師の指示) がこの系に与えられたときの最適学習識別アルゴリズムを求める。

設計基準として誤り率最小を用いると、最適識別系は第 n 期間において条件付確率 $P(\theta_n = 1 | Y^n, \theta^{n-1})$ と $P(\theta_n = 0 | Y^n, \theta^{n-1})$ との尤度比 λ を計算する (ベーズ推定), そして λ と 1 の大小により Y_n を識別する。ここに $Y^n = \{Y_1, Y_2, \dots, Y_n\}$ である。

ベーズの定理を用いると

$$P(\theta_n = 1 | Y^n, \theta^{n-1}) = \frac{P(\theta_n) P(Y_n | Y^{n-1}, \theta_n)}{P(Y_n | Y^{n-1}, \theta^{n-1})} \quad (3.5)$$

となる。式(3.5)の分母は θ_n に無関係の項である。分子の第2項は直接積分により求めることもできるが、ここでは Kalman の理論を用いて求める。いま考察している系の入力パターンは書き改めると

$$Y_n = \theta_n W_n Y_n + N_n \quad (3.6)$$

となる。ここに W_n は

$$W_n = [I_L, Q, Q, \dots, Q] \quad (3.7)$$

なる $L \times lL$ マトリックスである。式(3.4)で推移する系を式(3.6)で観測すると、 Y_n の事後確率は

$$P(Y_n | Y^{n-1}, \theta^n) = Q(Y_n - \theta_n W_n H_n, \Sigma_n + \theta_n W_n \Sigma_n W_n^T) \quad (3.8)$$

で与えられる。ここに $Q(A, B) = |2\pi B|^{-1/2} \exp\{-(1/2)A^T B^{-1} A\}$ 、 H_n は Y^{n-1} および θ^{n-1} が与えられたときの Y_n の事後平均であり、 Σ_n はその事後分散である。これらは次の繰返し形関係式で与えられる¹⁹⁾。

$$\left. \begin{aligned} H_1 &= M_1, \quad \Sigma_1 = \Phi_1 \\ H_{n+1} &= G_n \{H_n + K_n (Y_n - W_n H_n)\} \\ K_n &= \theta_n \Sigma_n W_n^T (W_n \Sigma_n W_n^T + \Sigma_n)^{-1} \\ \Sigma_{n+1} &= G_n (I_{lL} - K_n W_n) \Sigma_n G_n^T + R_n \end{aligned} \right\} \quad (3.9)$$

ここに

$$R_n = \begin{bmatrix} \Psi_n & Q & \dots & Q \\ Q & & & \\ \vdots & & & \\ Q & & & \end{bmatrix} \quad (tL \times tL \text{ マトリックス})$$

$t = \min(l, n+1)$

なお式(3.9)の $W_n H_n$ は $X_n - M_n$ の事後平均であり、 $W_n \Sigma_n W_n^T$ はその分散である。これらの量は式(3.9)を用いて教育系列 $\{\theta_n\}$ を与えることにより逐次学習される。またガウス分布の場合、事後平均は一般に admissible loss に対する最適推定になっている²⁰⁾。特にパターンが時変でなく固定している場合、すなわち $g_{n1} = -I_L, \Psi = Q, \Sigma_n = \Sigma$ および Σ_n が正定 (positive definite) のとき式(3.9)は

$$H_{n+1} = H_n + \theta_n \Sigma_{n+1}^{-1} (Y_n - H_n) \quad (3.10a)$$

$$\Sigma_{n+1} = (\Sigma_n^{-1} + \theta_n \Sigma^{-1})^{-1} \quad (3.10b)$$

となり Scudder の結果⁴⁾と一致する。このアルゴリズムを Scudder は Gremlin アルゴリズムと呼んでいる。

なお式(3.10a)で与えられるアルゴリズムは

$$\begin{aligned} H_{n+1} &= H_n + \nu_{n+1} (Y_n - H_n) \\ &= (1 - \nu_{n+1}) H_n + \theta_n \nu_{n+1} X + \nu_{n+1} Y_n \end{aligned}$$

ただし $\nu_{n+1} = \theta_n \Sigma_{n+1}^{-1} \Sigma^{-1}$

となり $T_{n+1}(H_1, \dots, H_n) = (1 - \nu_{n+1}) H_n \nu_{n+1} X$

$$F_n = 1 - \nu_n$$

とみなせば Dvoretzky の条件¹²⁾を満足し

$$\begin{aligned} \lim_{n \rightarrow \infty} E(\|H_n - X\|^2) &= 0 \\ P\{\lim_{n \rightarrow \infty} H_n = X\} &= 1 \end{aligned}$$

となる。要するにパラメトリック型学習識別アルゴリズムの基礎となるベーズ推定法は確率近似法に他ならない。したがって $n \rightarrow \infty$ になれば識別関数は双1次形式から線形へ推移する。その様子を示したのが Fig. 3.1 (次頁参照) である。

以上のアルゴリズムは時系列パターンのみならず航空写真その他の空間パターンの場合にも適用可能である²¹⁾。

更に識別結果を入力側に帰還することにより、自己バイアス型の場合に比べ (同一の SN 比において) 誤り率を改善することができる²²⁾。その場合の最適帰還量は次式で与えられる。

$$U_n = \rho H_n \quad (3.11)$$

ただし H_n は X_n の事後平均であり、式(3.6)に対応して $Y_n = \theta_n X_n + N_n - U_n$ である。この系はマーク読取機とか画像強調処理等におけるスライスレベルの自

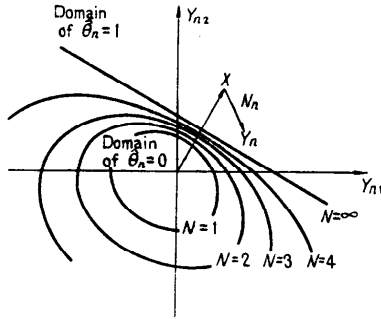


Fig. 3.1 Transition of Decision Domain Along with Learning Steps in Parametric Learning Algorithms with Teacher

動設定アルゴリズムとして応用できる*。

3.2 教師なしの学習識別機構

教育系列を与える教師がない場合、式(3.8)は

$$P(Y_n | Y^{n-1}) = \int \theta^{n-1} P(Y_n | \theta^{n-1}, Y^{n-1}) P(\theta^{n-1} | Y^{n-1}) \quad (3.12)$$

となり、計算すべき項数 n と共に指数関数的に増大する。したがって、このような形式の最適な教師なし学習は非実地的である。そこで前述の Gremlin アルゴリズムは H_n および E_n が近似的に求まれば構成できるという点に着目して以下のようなアルゴリズムを構成することができる²³⁾。

(i) Decision-directed Algorithm (DDA)

H_n および E_n を求めるとき、 θ_n の代りに機械自身の最尤決定の結果 $\hat{\theta}_n$ を使うアルゴリズムである。このとき $\hat{\theta}_n$ には必然的に誤りを含むため、 H_n は X に収束しないで kX に収束する。ここに k は 1 より大きなスカラで、 SN 比の関数である。したがって SN の大きくなるときは DDA の性能は良いが、 SN 比が小さくなると悪くなる。その欠点を改善する試みが以下のアルゴリズムである。

(ii) Consistent Estimator-type DDA (CDDA)

Y_n に SN 比の関数 α をかけることにより (i) の k を打消す方法である。すなわち

$$H_n \triangleq \lim_{n \rightarrow \infty} H_n \cong E(aY_n | \hat{\theta}_n = 1) \cong X \quad (3.13)$$

(iii) Modified DDA (MDDA)

もし大きな記憶容量と長い計算時間が許されるならば MDDA を用いることができる。これは DDA において $\hat{\theta}_n$ の決定を m 期間遅らせる方式である。したがって MDDA は過去 m 期間の可能なすべての系

列 2nd 種類について事後確率を求める。したがって最適なアルゴリズムと DDA との折衷したものである。

(iv) Non DDA (NDDA)

DDA は SN の比の小さい場合誤りが大きくなり性能が低下する。したがって決定によらない推定量

$$\mu_n = \frac{1}{(n-1) \rho} \sum_{i=1}^{n-1} Y_i \quad (3.14)$$

を用いて真のパターンを推定し、これを識別に用いるアルゴリズムである。 μ_n の条件付き期待値 $E(\mu_n | X) = X$ であるから、 μ_n は X の不偏推定量であり、また大数の法則により X の一致推定量でもある。 μ_n の共分散行列は $K/(n-1)$ で与えられる。ここに

$$K = \frac{1}{\rho^2} \{ (\rho - \rho^2)(\Phi + MM^T) + \Sigma \} \quad (3.15)$$

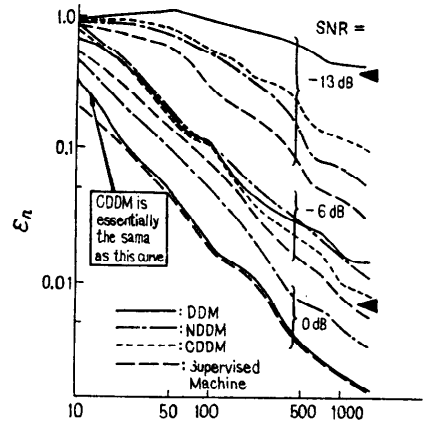


Fig. 3.2(a) Normalized-Squared-Error ϵ_n as a function of Time of Learning Processes for Four Machine Used

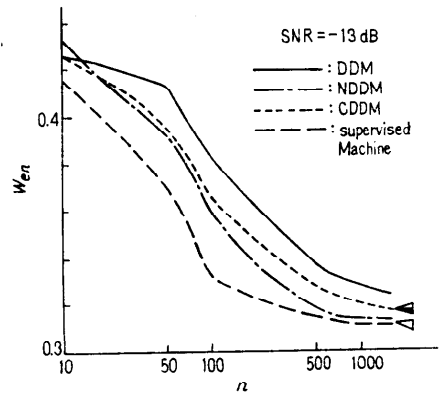


Fig. 3.2(b) Probability of Error W_{en} as a Function of Time in Learning Processes for Four Machines Used

* 筆者等は「蛍光眼底写真における色素漏出現象の解析 (I)」(信学技報 75 巻 148 号 IE 75-78, 1975-11) にこのアルゴリズム適用を目下実験中である。

この場合、学習アルゴリズムは次のようになる。

$$\left. \begin{aligned} H_1 &= M, \quad \varepsilon_1 = \Phi \\ H_n &= H_{n-1} + \varepsilon_n K^{-1} \left(\frac{X_{n-1}}{p} - H_{n-1} \right) \\ \varepsilon_n &= \varepsilon_{n-1}^{-1} + K^{-1} \end{aligned} \right\} \quad (3.16)$$

この NDDA は安定した特性を示し、Gremlin アルゴリズムに収束する。次の Fig. 3.2 (a), (b) (前頁参照) は DDA, CDDA, NDDA, Gremlin アルゴリズムについて、それぞれ平均 2 乗誤差と誤り率を比較したものである。Fig. 3.2(a) の \blacktriangleright は DDA の ε_n は収束値、Fig. 3.2(b) の \triangleleft と \triangleleft はそれぞれ Gremlin アルゴリズムの収束値および DDA の誤り率の収束値を示す。

3.3 パターン系列の教師付き学習同期化機構²⁴⁾

時系列パターンの場合、周期はピリオドグラム等により推定できたとしても、必ずしも同期が始めからとれてはいない場合もある (Fig. 3.3)。したがって観測パターンベクトル Y_n は θ_{n+1} に依存する。s を未知時系列パターンの真の始点とすれば、この場合の事後確率は式(3.17)で与えられる。

$$P(s, \theta_n | \theta^{n-1}, Y^n) = \sum_{\theta_{n+1}=0}^1 P(s, \theta_n, \theta_{n+1} | \theta^{n-1}, Y^n) \quad (3.17)$$

ただし

$$\begin{aligned} P(s, \theta_n, \theta_{n+1} | \theta^{n-1}, Y^n) &= C p(\theta_{n+1}) P(Y_n | s, \theta^{n-1}, \theta_n, \theta_{n+1}, Y^{n-1}) \\ &\times P(s, \theta_{n-1}, \theta_n | \theta^{n-2}, Y^{n-1}) \end{aligned} \quad (3.18)$$

ここに

$$C = 1 / [P(Y_n | \theta^{n-1}, Y^{n-1}) P(\theta_{n-1} | \theta^{n-2}, Y^{n-1})]$$

であり、s, θ_n, θ_{n+1} に依存しない係数である。

s の最尤推定値 \hat{s} は $P(s | \theta^{n-1}, Y^n) = \sum_{\theta_n=0}^1 P(s, \theta_n | \theta^{n-1}, Y^n)$ を最大にする s である。また θ_n の決定 $\hat{\theta}_n$ は

$$p(\theta_n = 1 | \theta^{n-1}, Y^n) = \begin{cases} \geq \frac{1}{2} & \text{ならば } \hat{\theta}_n = 1 \\ < \frac{1}{2} & \text{ならば } \hat{\theta}_n = 0 \end{cases}$$

で与えられる。式(3.18)の第3項は

$$P(Y_n | s = i, \theta^{n-1}, \theta_n, \theta_{n+1}, Y^{n-1}) = Q(Y_n - V_{in} H_{in}, V_{in} \varepsilon_n V_{in}^T + \sigma_i) \quad (3.19)$$

ここに

$$\begin{aligned} V_{in} &= \begin{bmatrix} Q \cdots \theta_n I_{i-1} \\ \theta_{n+1} I_{L-i+1} Q \end{bmatrix}, \quad I_i = l \times l \text{ 単位マトリックス} \\ \Sigma &= \begin{bmatrix} \underline{\sigma}_{11} & \underline{\sigma}_{12} \\ \underline{\sigma}_{12}^T & \underline{\sigma}_{22} \end{bmatrix} L - i + 1, \quad \underline{\sigma}_i = \begin{bmatrix} \underline{\sigma}_{22} & Q \\ Q & \underline{\sigma}_{11} \end{bmatrix} \end{aligned}$$

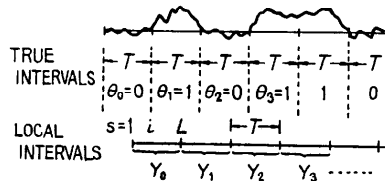


Fig. 3.3 Measurement Sequence

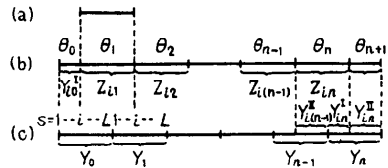


Fig. 3.4 Hypothetical Intervals

- (a) *i*th Hypothetical Interval
- (b) Sequence of *i*th Hypothetical Interval
- (c) Sequence of Local Interval

$H_{in}, i=1, 2, \dots, L$ および ε_n は次のアルゴリズムにより学習される

$$\left. \begin{aligned} H_{i1} &= M, \quad \varepsilon_1 = A \\ H_{in} &= H_{i(n-1)} + \theta_{n-1} \varepsilon_n \Sigma^{-1} (Z_{i(n-1)} - H_{i(n-1)}) \\ \varepsilon_n &= (\varepsilon_{n-1}^{-1} + \theta_{n-1} \Sigma^{-1})^{-1} \end{aligned} \right\} \quad (3.20)$$

ここに第 (n-1) 学習期間の第 i 標本点より、第 n 学習期間の第 (i-1) 標本点までを第 i 仮説期間とし、その第 i 仮説期間に含まれる Y^n の部分を Y_{in}^I 、残りを Y_{in}^{II} とすれば (Fig. 3.4)

$$Z_{in} \triangleq (Y_{i(n-1)}^{II T} Y_{in}^{I T})^T$$

で与えられる。

むすび

まえがきに述べたようにパターン構造が主として hypergeometrical に記述されているものと仮定して、一般的パターンの学習識別に適用可能なアルゴリズムを紹介した。syntactic または linguistic な記述のなされている場合についても学習識別アルゴリズムを求めることができるが紙面の都合上割愛する^{25)~28)}。

参考文献

1) Workshop on "Bridging the Gap between Theory and Implementation in Pattern Recognition" (1st) • Workshop on "Problems in Pattern Recognition Research" (1st) • Panel on "Structure Pattern Recognition" (2nd) • Panel on "Feature Extraction" (2nd) ここに 1st およ

- び 2nd はそれぞれ The First (held on Oct. 30-Nov. 1, 1973) Int'l. Joint Conf. on PR および The Second (held on Aug. 13-15, 1974) Int'l. Joint Conf. on PR を表わす。
- 2) L. A. Gerhardt: Bridging the Gap between Theory and Practice in PR (Workshop, 2nd).
 - 3) N. E. Nahi et al.: Non-linear Adaptive Recursive Image Enhancement (1st).
 - 4) H. J. Scudder: Adaptive Communication Receiver, IEEE Trans. Vol. IT-11 No. 2, pp. 167~174 (1965).
 - 5) K. Tanaka: Some Considerations on A Type of Pattern Recognition Using [Nonsupervised Learning Procedure, Proc. Hawaii Int'l. Conf. on System Sci. pp. 546~549 (Jan. 1968).
 - 6) H. Teicher: Identifiability of finite mixtures, Ann. Math. Stat. Vol. 34, p. 1265 (1963).
 - 7) C. V. Jakowatz et al.: Adaptive Waveform Recognizer, Proc. 4th London Symp. on Inf. Theory, pp. 317~326 (1961).
 - 8) K. Tanaka et al.: An Identification Method of System Characteristics Using A New Type of Adaptive Correlating Filter, Proc. IFAC Tokyo Symp. pp. 245~254 (Aug. 1965).
 - 9) 樋口, 田村, 田中: 未知波形の適応同定, 計自制御論文集, 巻5, 号1 (1969).
 - 10) K. Fukunaga et al.: A Criterion and An Algorithm for Grouping Data, IEEE Trans. Vol. C-19, No. 10, pp. 917~923 (1970).
 - 11) 今井, 志村: 主成分による教師なしの学習パターン識別, 計自制御論文集, 巻9, 号2 (1973).
 - 12) J. Wolfowitz: On the Stochastic Approximation Method of Robbins and Monro, Ann. Math. Stat. Vol. 23 (1952).
 - 13) 田中他: 学習過程の一般的考察, 信学会, オートマトン, インホメーション研究 (1971-10).
 - 14) B. Widrow: The "Rubber Mask" Technique-I, in "Learning Systems and Intelligent Robots" (ed. Fu & Tou), pp. 365~400, Plenum Press, New York (1974).
 - 15) C. A. Rosen: Pattern Classification by Adaptive Machines, Science, Vol. 156, No. 3771, pp. 38~44 (Apr. 7, 1967).
 - 16) M. Shimura: Multicategory Learning Classifiers for Character Reading, IEEE Trans. Vol. SMC-3, No. 1 (Jan. 1973).
 - 17) M. Shimura, K. Tanaka: On Some Problems of Pattern Recognition, Proc. Hawaii Int'l. Conf. on System Sci, p. 807 (Jan. 1968).
志村, 田中: 自己組織化機能をもつ層状学習認識機械, IECE (電子通信学会) Trans. Vol. 54-C, No. 8 (1971-08).
 - 18) K. S. Fu et al.: Information Processing of Remotely Sensed Agricultural Data, Proc. IEEE, Vol. 57, No. 4 (Apr. 1969).
 - 19) S. Tamura, K. Tanaka: On the Recognition of Time-varying Pattern Using Learning Procedure, IEEE, Trans. Vol. IT-17, No. 4 (Apr. 1971).
 - 20) J. S. Meditch: Stochastic Optimal Linear Estimation and Control, p. 174, McGraw-Hill, New York (1969).
 - 21) R. W. McLaren et al.: Learning Texture Information from Singular Photographs and Its Application in Digital Image Classification, in "Learning Systems and Intelligent Robots" (ed. Fu & Tou), pp. 423~435, Plenum Press New York (1974).
 - 22) K. Tanaka et al.: Optimal Learning Recognizer for Unknown Signal Sets in A Channel with Feedback Link, in "Learning Systems and Intelligent Robots" (ed. Fu & Tou), pp. 273~293, Plenum Press, New York (1974).
 - 23) K. Tanaka: Some Studies on Pattern Recognition with Non supervised Learning Procedures, in "Pattern Recognition and Machine Learning" (ed. Fu), pp. 1~17 Plenum Press, New York (1971).
 - 24) K. Tanaka: A New Method of Pattern Recognition and Synchronization for Unknown Signal Sequences, Commission VI-6, URSI XVII-th General Assembly, Warsaw, (Aug. 21~29, 1972).
 - 25) H. Enomoto et al.: Synthesis of Automata that Recognize Given Strings and Characterization of Automata by Representative Sets of Strings, Proc. 1st US-JAPAN Comp. Conf. (October 1972).
 - 26) S. Tamura, K. Tanaka: Learning of Fuzzy Formal Languages, IEEE Trans. Vol. SMC-3, No. 1 (Jan. 1973).
 - 27) K. Tanaka, M. Mizumoto: Fuzzy Programs and Their Execution, in "Fuzzy Sets and Their Applications to Cognitive and Decision Processes" (ed. Zadeh, Tanaka et al.), pp. 41~76 Academic Press, New York (1975).
K. Tanaka: Learning in Fuzzy Machine, NA TO Advanced Study Institute on Computer-oriented Learning Processes, Bonas (France), (Aug. 27-Sep. 6, 1974).
 - 28) Y. Yakimovsky et al.: A Semantic Decision Theory Region Analyzer, Proc. 3rd Int'l. Joint. Conf. on Artificial Intelligence, p. 580 (1973).
(昭和51年2月13日受付)
(昭和51年3月26日再受付)