

自己組織化学習機械の提案と相互干渉のある英字認識への応用†

鈴木 寿^{††} 有木 卓^{††}

木枠データベースを自己組織化する一手法を提案し、英字認識へ応用する。システムは処理部と、初期状態では空白な記憶部とから成る。処理部は、入力信号に似ている信号を記憶部から探索し、その関連情報を出力する。もし入力信号の随伴情報が出力情報と一致しなければ、新たな情報が記憶部に付加され、その影響は次の一連の処理に反映される。このように標本系列と記憶状態とに依存する手順に従って、データベースが自己組織化されてゆく。データ圧縮の分野では、データ系列中の文脈構造を自己組織的に抽出してゆき、最適な圧縮率を漸的に達成するユニバーサル符号化が知られているが、提案した方法においても同様に、入力信号に対する情報の再現性は標本系列のそれに漸的に近づく。したがってこのシステムは、要求された任意の応答パターンをユニバーサルに学習する学習機械として応用できる。この学習機械を利用し、マイクロ計算機上に、印刷間隔が一定でない英活字の認識器、およびその発展型として連続筆記英字認識器を実現した。良好な実験結果が示される。

1. ま え が き

近年、塩野の辞書構成手法¹⁾、ファジー論理により木探索を行う Wang の手法²⁾、NTT が開発した漢字の構成部品の組合せを調査する手法³⁾ などのように、階書体筆記漢字認識に関する研究が盛んである。また英字認識に関しては、垂直スリットと単語綴り探索の組合せによる Hull の不特定印刷字体の英単語認識手法⁴⁾、特徴抽出による Kurtzberg の単筆記英字認識手法⁵⁾ などが報告されている。本論文では、テンプレートのデータベースを適応的に構成する類の一手法、すなわち自己組織化学習機械⁶⁾ を提案し、相互干渉のある英字 (図 1) 認識への応用例を報告する。

第 2 章で提案する学習機械は、大まかに処理部と記憶部とから構成される。処理部は標本として入力された信号に対し、記憶部に貯蔵されている過去の信号中からこれと似た一つを検索し、それと共に記録されている関連情報を呼び出す。もしこの情報が入力信号に随伴する情報と一致しなければ、記憶部の内容は修正されるが、その影響は学習機械のその後の挙動へ反映される。このようにして、入力信号に直接依存しつつアドレスが増設されてゆくという意味で、データベースが自己組織的に構築される。学習の初期段階ではデータベースの規模、すなわち使用メモリ量は、学習した標本の個数に比例して増大するが、入力信号と、

呼び出される関連情報との関係が徐々に標本系列のそれに近づくとも成長の速度は次第に衰え、やがて停止する。

第 3 章では、この学習機械を相互干渉のある英字認識へ応用するため、文書画像に窓を掛け窓内の画像を入力信号、また窓内の先頭文字の読み、および後続活字との境界線を、関連情報と見る方式 (図 3 (a)、図 4 (a)) について述べる。窓画像から先頭文字の読みと境界線を想起すると同時に、境界線を目安に窓を後方へずらすという作業を繰り返すことにより、文字認識が逐次実現できる。

2. 自己組織化学習機械の構成

学習対象に応じて設定された有限次元モデルにおいて係数ファクタの値を自動調整する類の学習機械は、次の設計・動作段階を踏む：

- ① 標本収集、
- ② モデルの選択と次元決定、
- ③ パラメタの計算。

一般にモデルの次元が高いほど学習対象となる情報構造を精密に再現できるが、同時に、パラメタ計算に要する時間とメモリ量とが増大するので、実際には次元を適度に妥協せざるを得ず、入力信号に対する関連情報の再現性に誤差が残存する結果となる。文字認識においては、関連情報の再現性は認識率という術語に置き換えられ、誤差の残存は認識率が 1 に満たない現象として観測される。これに対しデータベースを構成する類の学習機械は、標本系列を、その情報構造の複雑さの度合いに応じて徐々に次元を上げながら解析してゆく。したがってモデル設定の不備に起因する解析上

† A Proposal of Self-Organizing Learning Machine and Application to Recognition of Interactive English-Letters by HISASHI SUZUKI and SUGURU ARIMOTO (Department of Mechanical Engineering, Faculty of Engineering Science, Osaka University).

†† 大阪大学基礎工学部機械工学科

I been in some meetings where the table is contorted and the chairs knotted and the wall one another till you could of wrung sweat out in meetings where they kept talking about a that the patient materialized in the flesh, nu

(a) Typographic letters.

I been in some meetings where the tab contorted and the chairs knotted and the w one another till you could of wrung swea in meetings where they kept talking about a that the patient materialized in the flesh,

(b) Handwritten letters.

図1 相互干渉のある英字の原文書
Fig. 1 Source documents of interactive English-letters.

の限界は見られず、種々の応用に際し都合の良い性質をもたらすことが期待される。以下では後者の学習機械の方式を定義する。

2.1 忘却なし学習

数学的には、入力信号 x はある有限または無限集合 X の要素、関連情報 y は別の有限または無限集合 Y の要素、標本系列 $\{(x_i, y_i); i=1, 2, \dots\}$ は $X \times Y$ から無作為に選ばれた (x, y) の列である。 x と y との対応が一貫しているとは、 $x_t = x_s$ かつ $t \neq s$ なるすべての t, s の組合せに対し $y_t = y_s$ であることを言う。

以下に、一貫性のある標本系列に対する自己組織化学習機械の動作を、二進木形態のデータベースを構成する手順として述べる。記号 $x[n], y[n]$ は、各々の節 n に割り当てられた記憶用変数を表す。また、 d はある任意に与えられた写像

$$d: X \times X \rightarrow R_+ \quad (1)$$

(ここに R_+ はすべての非負実数の集合を表す) で、 $d(x', x'')=0$ なる各々の $(x', x'') \in X \times X$ と対応付けられた $(y', y'') \in Y \times Y$ に対し、常に $y' = y''$ を満足するものとする。これは広義の距離関数の一種であるが、この種の距離概

念については Goldfarb²⁾ が詳しい。

《手順1》 忘却なし学習

① 《初期設定》 x [根] = x_1, y [根] = y_1 とおき、根のみから成る木を設定する。また $t=1$ とおく。

② t を1だけ増やす。節を表すポイント n を根にセットした後、 n が末端節へ到達するまで以下を繰り返す: n の子節を n_L, n_R とおくと、

$$d(x_1, x[n_L]) \leq d(x_1, x[n_R]) \quad (2)$$

ならば $n = n_L$, それ以外ならば $n = n_R$ とおく。

③ $y[n] = y_s$ ならば②へ戻り、それ以外ならば n に続く新たな子節 n_L, n_R を増設し

$$x[n_L] = x[n], y[n_L] = y[n],$$

$$x[n_R] = x_t, y[n_R] = y_t,$$

$$(3a) - (3d)$$

とおいた後②へ戻る。

例えば、ある標本系列の先頭部分

$$(x_1, y'), (x_2, y''), (x_3, y''), (x_4, y''), (x_5, y''), \dots$$

で $d(x_3, x_1) < d(x_3, x_2)$ かつ $d(x_4, x_1) < d(x_4, x_2)$ ($k=2, 3$) かつ $d(x_5, x_2) > d(x_5, x_1) > d(x_5, x_3)$ なるものに対し、図2のような木が構築される。

また、任意の入力信号 $x \in X$ に対応する関連情報 $y \in Y$ の再生とは、手順1の②で x_t の代わりに x

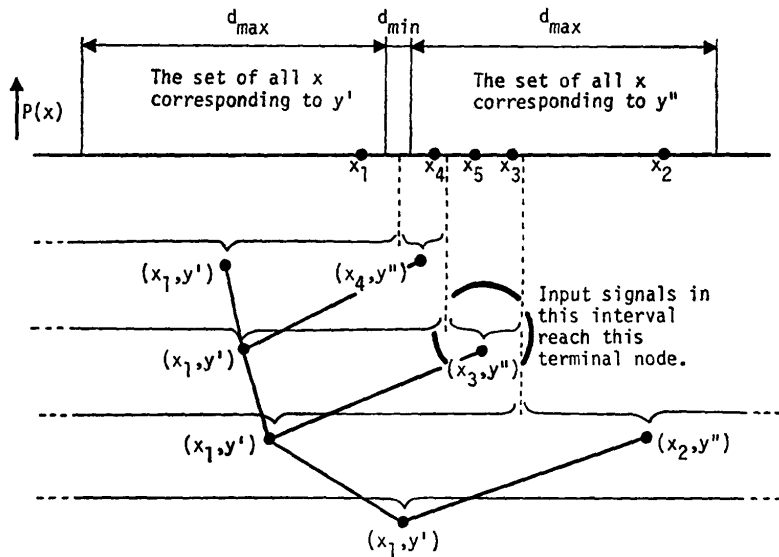


図2 情報構造の例

Fig. 2 An example of information structure.

とおき、到達した末端節 n の記録 $y[n]$ を y と見なす作業であるとする。

さて、手順1により構築される木のレベル数は、計算時間と使用メモリ量とに反映されるので、この量を評価しておくことが重要である。定性的には、学習開始後しばらくの間、節数は学習回数にほぼ比例して増大するが、似た入力どうしが同じ末端節へ到達するようになると成長が停滞する。正確にはレベル数の上界は次のように評価される。

X のある分割 $\{X_k: k \in K\}$ は、任意の k を考えるとき、各々の $(x', x'') \in X_k \times X_k$ と対応付けられた $(y', y'') \in Y \times Y$ に対して常に $y' = y''$ であるような、ある最小個数分割を表すとする。さらに、個々の X_k を囲む最小球の直径の上界を

$$d_{\max} = \max_{k \in K} \max_{(x', x'') \in X_k \times X_k} d(x', x''), \quad (4)$$

また球どうしの間隙の下界を

$$d_{\min} = \min_{\substack{(k', k'') \in K \times K \\ k' \neq k''}} \min_{(x', x'') \in X_{k'} \times X_{k''}} d(x', x'') \quad (5)$$

と表し、

$$r = (d_{\max} + d_{\min}) / d_{\min} \quad (6)$$

とおく。

もし、ある m 次元 Euclidean 空間 R^m 上の距離関数 d_m と、ある写像 $\rho: X \rightarrow R^m$ とが存在し、かつ任意の $(x', x'') \in X \times X$ に対し

$$d(x', x'') / d_m(\rho(x'), \rho(x'')) \leq a \quad (7)$$

なる正定数 $a < 2$ が存在する一すなわち $\rho(X)$ が R^m 上でほぼ線形性を有する一ならば、すべての時刻においてレベル数 l は

$$l \leq (|K| - 1) \cdot (2m - 1) \lceil \log_{2/a} r \rceil \quad (8)$$

(ここに $|K|$ は K の要素数を表す) の範囲にある。この主張は、最初 $a=1$ かつ $m=1$ の場合 (図2) を考え、次いで $m>1$ の場合、最後に $a>1$ の場合を考えることにより得られる。

以上の結果から知られるように、 $r>0$ すなわち球どうしの間隙がわずかでも存在すれば、レベル数は有限である。さらに、 X_k どうしがよく分離されているほど r は小さい値を持ち、このときレベル数も小さい上界によって押えられることがわかる一すなわち直観的に言って、雑音あるいは書き手の不安定性による文字形状のゆらぎが小さいほど r は小さい値を持つので、レベル数も小さくなる。このように r は学習対象の情報構造の複雑さを表す一つの尺度となっている。

なお r が有限のとき、学習が進むにつれ、入力信号

と、再生される関連情報との対応関係が標本系列のそれに近づくことは明らかであろう。

2.2 忘却を伴う学習

学習対象の情報構造が複雑すぎる、すなわち文字形状のゆらぎが激しすぎる場合、使用メモリ量が計算機の最大許容を越える危険がある。それゆえ、あらかじめ規定された限界 N に節数が達したとき、一部の部分木を刈り込み新たな情報の受け入れに備えなければならない。刈り込むべき部分木の選択方法はいくつか考えられるが、ここでは計算の容易さの点から次のミニマックス戦略を採用する：

『ある節 n を刈り込むことにより、最悪の場合、本来 n を訪問すべきであったすべての入力信号が別の節を訪問し、誤った関連情報を再生する可能性がある。それゆえ、その部分木の根への推定訪問確率があらゆる部分木の中で最小なる部分木を刈り込むことにより、その被害を最小限に押える。』

各節への訪問確率は、標本発生源が有限次 Markov 過程であるとき、訪問回数より推定できる (Markov 過程については Feller⁹⁾ が詳しい)。例えば節 n への訪問回数が c 、その子節 n_L, n_R への訪問回数が c_L, c_R のとき、入力信号 x が n を訪問したという条件下で、さらに n_L, n_R を訪問するであろう確率はそれぞれ

$$c_L/c, \quad c_R/c$$

によって推定される。根から測った訪問確率はこれらの条件付き確率の積として推定される。ただし、これらの推定は c が大きくなければ有効でない。したがって、ある正定数 c_0 を定めておき、親節への訪問回数が c_0 に満たないような節を根とする部分木は刈込みの候補から除外するという工夫を施す。

こうして次のような動作手順の自己組織化学習機械が構成される：

《手順2》 忘却を伴う学習

① 《初期設定》 $x[\text{根}] = x_1, y[\text{根}] = y_1, c[\text{根}] = 1$ および $P[\text{根}] = 1$ とおき、根のみから成る木を設定する。また $t=1$ とおく。

② t を1だけ増やす。節を表すポインタ n を根にセットした後、 n が末端節へ到達するまで以下を繰り返す：

1) n の子節を n_L, n_R と表すとき、

$$d(x_i, x[n_L]) \leq d(x_i, x[n_R]) \quad (9)$$

ならば $n = n_L$ 、それ以外ならば $n = n_R$ とおく；

2) $c[n]$ を1だけ増加させ、

$$P[n_L] = p[n] \cdot c[n_L] / (c[n_L] + c[n_R]), \quad (10a)$$

$$P[n_R] = p[n] \cdot c[n_R] / (c[n_L] + c[n_R]) \quad (10b)$$

とおいた後、 $P[n_L] \leq P[n_R]$ ならば n_L に、それ以外ならば n_R に目印を付ける。

③ $y[n] = y_i$ ならば②へ戻る。それ以外ならば、④の忘却を行い、 n に続く新たな子節 n_L, n_R を増設し

$$\begin{aligned} x[n_L] &= x[n], & y[n_L] &= y[n], \\ x[n_R] &= x_i, & y[n_R] &= y_i, \\ c[n_L] &= 1, & c[n_R] &= 1 \end{aligned} \quad (11a)-(11d)$$

とおいた後、②へ戻る。

④ 《忘却》 節数が N 以上ならば、以下を実行する：

1) ある節 n とその兄弟節 n' とに対し $c[n] + c[n'] \geq c_0$ なる n のうち $P[n]$ が最小であるもの一つを、目印を辿ることによって探す；

2) n を根とする部分木を削除し、 n' を根とする部分木を n の親節 n^o に n' が重なるように接木した後、 n' 用の記憶変数 $x[n'], y[n'], c[n'], P[n']$ およ

正しく認識できる製品がある。認識対象を限定するならば、計算機への補助入力手段として利用されるリストリーダのように廉価ながら比較的信頼性の高い製品もある。しかしながら書籍等の英活字 (図 1(a)) 認識に関しては、高密度印刷による活字形状の歪みや印刷間隔の変動などの文字間の相互干渉が、外来雑音と共に状況を複雑化することから、これを読み取る装置の設計手法はいくつかの限定された書体に対し専用開発される類⁶⁾を除きまだ調査し尽くされていない。ここではこの問題に対処できる汎用的な認識手法⁶⁾の一つを提案する。

3.1 自己組織化学習機械の応用法

イメージスキャナを用い水平軸補整を行いつつ収録した文書画像において、活字1個の最大高より大きい H_0 と最大幅より大きい W_0 とに相当する辺を持つ長方形の枠を窓と呼ぶ (図 3(a))—すなわち

$$X = \{0, 1\}^{H_0 \times W_0} \quad (13)$$



また t 回の学習後の、入力信号 $x \in X$ に対応する関連情報の再生 $g_t: X \rightarrow Y$ は、忘却なしの場合と同様に、手順 2 の②で x_i の代わりに x とおき、到達した末端節 n の記録 $y[n]$ を $g_t(x)$ と見なすことにより行う。

さて仮定した有限次 Markov 過程を $\{(X_t, Y_t); t = 1, 2, \dots\}$ と表すとき、 g_t に関する関連情報の再現性—すなわち第 3 章で言うところの認識率—を任意の時点 t_0 における $Pr\{g(X_{t_0}) = Y_{t_0}\}$ によって定義する。学習回数が十分大きいとき、入力信号の訪問回数が c_0 を越える節は少なくとも N/c_0 個以上存在し得るから、もし r が有限ならば、関連情報の再現性は下界

$$1 - O(1/N)$$

によって押えられることがわかる。

節数が異常増大するもう一つの考え得る場合は、標本が一貫性を持たせない、すなわち、ある入力信号に対応する関連情報が複数通り出現する場合である。そのような入力信号の集合を E と表し

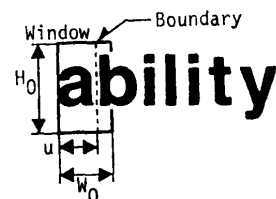
$$e = Pr\{X_t \in E\} \quad (12)$$

とおくとき、 g_t により再生される関連情報のうち率 e は信頼できないから、再現性は下界

$$(1 - e) - O(1/N)$$

字を垂直線で走査し、最初に黒点を検出した位置に左辺が重なるよう窓を掛ける。このとき窓内には、先頭活字とその後続活字の一部とが写っている。もし先頭活字が何であるかが判別できれば、それに伴い活字の終了位置、すなわち後続活字との境界線がわかる。したがって、引き続きこの境界線から走査を開始し以上の一連の作業を繰り返せば、再帰的に活字認識が実現できる。それゆえ論点は、窓内の先頭活字を判別する効率的な方法を見いだすことにある。

既存の方法を改良して得られる初歩的な判別法は次のとおりである：



(a) Windowing.



《判別法1》 テンプレートマッチング

高さ H_0 , 幅 W_0 の窓内に左詰めで当該活字のパターンとその終了位置とを記したものをテンプレートとし, 判別したい記号数分だけこれを準備する。

判別対象の窓画像と, あるテンプレートとを重ね合わせるとき, 終了位置より左側の領域で排他的論理和をとった結果の黒点の面積が最小となるようなテンプレートを探索し, これに対応する読みを判別結果とする。また, 終了位置を境界線と見なす。

一般のテンプレートマッチング法は広く利用されているが, この方法の利点は活字の書体に関係なく応用できる汎用性と, かすれ・にじみ等の雑音への耐性にある。後者については, 漢活字あるいは計算機出力文字のように活字間にある程度すき間がありかつ印刷間隔が一定ならば, 垂直方向への射影の分布を調べるなどの補助手段により窓の位置決めが比較的正確に行えるので, そのような条件下では確かに雑音に強い。だが印刷間隔が一定でない場合, 仮に後続活字が変形していると, 後続活字の開始位置を見誤り, しばしばそれ以降の活字の判別誤りを誘因するので, むしろ雑音に弱くなる。

雑音を考慮に入れた, これに代わる手法は, 第2章で提案した自己組織化学習機械を採用し, 窓画像内の活字の組合せの未学習分を推定する規則を d の定義へ含めた上で, 雑音を含む窓画像の標本系列から直接にテンプレートのデータベースを作成することである。このように設計された機械は, 学習回数が増えた際, 推定規則の不備を雑音の影響と共に自然に吸収する形態となっている。それゆえ, 学習初期における読みと境界線との再現性の立上がり速度は d の定義の成否に依存するが, 学習が繰り返されるにつれ再現性は第2章の意味で確実に上限に近づいてゆくことが保証される。したがって, 残された問題はデータベースができる限り小規模となるような d の定義を実験的に見いだすという作業に帰着される。

こうして後に3.2節で述べる実験を種々の d に対し行い, 最も適切と思われる d の定義を選んだ。

《判別法2》 自己組織化学習機械の応用

それぞれの $(x, x') \in X \times X$ に対し d を次のように定める (図3(b)): 窓画像 x' 上で境界線より左側の領域に注目し, その領域内の黒点を B と記す。各各の B を窓画像 x 上へ投影するとき, B に最も接近している x 上の黒点 B^* との幾何学的距離 δ を測り, すべての B に関する δ の総計を B の個数で割

った結果が $d(x, x')$ であるとする。

読みと境界線に関連情報 y と見なし, これに対応するように Y を定める。文書画像に対し手順1または2を適用し, 窓画像と読み・境界線との関係を学習させる。このデータベースを用いて判別を行う。 ■

3.2 活字認識の実例

自己組織化学習機械を利用した活字認識器の運用形態として次の3通りが考えられる:

《適応方式》 利用者が常時教示を行い, 専用のデータベースを構築する;

《固定方式》 設計者は種々の書体に対するデータベースを準備し, 利用者は希望のデータベースを認識器にセットして学習なしで利用する;

《混合方式》 上記二方式の混合形態で, 設計者が準備したデータベースを基に, 利用者が用途に応じて再教示する。

適応方式が最も基本となる運用形態であるから, 提唱した認識手法の有効性を示すには, この形態における認識器の性能すなわち認識率を実例により確かめればよいだろう。

記憶容量 640 kbyte のマイクロ計算機 NEC-PC 9801 上に自己組織化学習機械を利用した活字認識器を実現し, その性能を調べた。システムのおおまかな仕様は次のとおりである。イメージスキャナの性能を考慮に入れ, $H_0=20$, $W_0=10$ とおく:

$$X = \{0, 1\}^{20 \times 10}. \quad (14)$$

また, 読みを表す 128 個のアスキーコードと境界線座標との対から成る集合を Y とする:

$$Y = \{0, \dots, 127\} \times \{0, \dots, 9\}. \quad (15)$$

学習対象となる $x \in X$ と $y \in Y$ との対応関係は, 与えられた文書画像に対し, 例えば「図3(a)の窓画像 x に対応する y は (97, $u-1$) である (ここにアスキーコード 97 は a を表す)」というように, 利用者の教示によって暗黙のうちに定められる。

システムは次の順序で動作させる。3.1節の初頭で述べたように, イメージスキャナにより文書画像を収録し, 水平補整を行った後, 手順1または2を適用する。窓画像に対し, データベースから再生される当該活字の判別結果が誤っているならば利用者はコンソールから正しい読みを教示し, また境界線位置が誤っているならばマウスを用いて正しい境界線位置を教示する。

典型的な原文書一図1(a)にその一部を示す一に対し, $N=500$, $c_0=5$ において手順2の下で節数の変遷と認識率とを調べた結果を図5(a)に実線で示す。こ

ここに認識率は、過去 1000 回における正答の相対頻度によって算出した。木のレベル数は 8000 時点で 5 から 15 の範囲に分布しており、認識率は約 95% に収束した。(なお、種々の活字の文書画像に対し同様な実験を行ったところ、認識率は 90-100% の範囲の値に収束した。)

同様に $N=\infty$ すなわち手順 1 の下での認識率と節数の変遷を点線で重ね描いた。ただし認識率は $N=500$ の場合と大差なく、図では点線と実線とを区別できない。節数がある程度大きくなると認識率の増加は 100% 未満で停滞し 2.2 節の最後に述べた現象が起こる。誤認識の主な原因は、ランダムな雑音によって学習対象の一貫性が部分的に破壊されることにある。すなわち、雑音により例えば c と e とが同じ窓画像に変形する可能性があるとき、この窓画像を c と教える場合も、また e と教える場合もあり、その度に木が成長するが、誤認識の可能性は依然として残る。したがって、認識率の増加が停滞した時点以降に木の成長を許しても認識率は改善されないので、 N は適当な大きさに妥協しておく方が賢い。

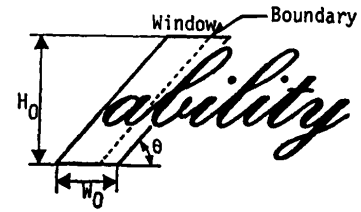
学習対象の一貫性の不成立は、どのような認識器によっても制御できない要因であり、これに由来する誤謬は認識手法の良否に関わらず、高密度印刷された活字認識に生来的につきまとう。したがって大抵の原文書に対し、自己組織化学習機械を利用した認識手法もその他の手法も誤認識を残存する。だが後者においては、誤認識は学習対象の情報構造に関するモデル仮定の不備に起因する分も含むのに対し、自己組織化学習機械は N を十分大きくとりさえすればそのような不備を除去できるので、誤認識は理論上の最小量、すなわち、異なる読みに対し互いに区別できない窓画像が生成される確率、によって押えられることとなる。こうして残存した誤謬に対処するためには、さらに単語の綴りに注目した認識手法⁴⁾へと発展させなければならないが、これは別の機会に論ずることとする。

3.3 連続筆記英字認識への発展

活字と異なり、筆記文字(図 1 (b))は

- ① 書き手に応じて多様な書体が存在する；
- ② 文字の垂直位置が変動する；
- ③ 隣接する文字間の境界線が傾いている；

などの攪乱要因を伴う。自己組織化学習機械を利用すれば多様な書き手の個性に適應できるので、残された問題は後半の二つの攪乱要因へどう対処すればよいかに絞られる。以下にその対処法を述べる。



(a) Windowing.



(b) Calculation of slant angular.

図 4 筆記文字認識の方法

Fig. 4 Method of handwritten letter recognition.

水平線を引いた下敷きの上に紙を載せ、書き手に対し、文字を水平方向に揃えて書くように制約を課す。(この制約は、書き手の個性を没するほど強力ではないだろう。)これにより②の障害は解決される。

イメージスキャナで収録した文書画像(図 1 (b))において、少なくとも一文字を覆えるだけの面積を持つ高さ H_0 、幅 W_0 の平行四辺形状の窓を掛け(図 4 (a))、活字の場合と同様な方法で再帰的に文字認識を行う。その際、各単語の走査に先立ち、次の要領で窓の傾きを決める(図 4 (b)):ある傾き θ のスリットを掛け、長軸方向に見て最も外側の黒点を包含する区間の長さ δ を、スリットを水平方向にずらしつつ測り、一単語にわたってその平均 $\langle \delta \rangle$ を算出する； $\langle \delta \rangle$ を最大化する θ の値をもって所要の傾きとする。この方法は、文字が左に傾いているような場合でも、各行において傾きがほぼ一定ならば有効である。こうして③の障害も解決される。

典型的な原文書一図 1 (b)にその一部を示す一に対し、 $H_0=20$ 、 $W_0=10$ 、また $N=1000$ 、 $c_0=5$ において手順 2 の下で節数の変遷と認識率とを調べた結果を図 5 (b)に実線で示す。この例では θ が 56-79 度の範囲で変動し、木のレベル数は 4000 時点で 6 から 19 の範囲に分布していた。3.2 節と同様、認識率を過去 1000 回における正答の相対頻度によって評価したところ、約 74% に収束した。(なお、いろいろな書き手に対し個別に、同様な実験を行ったところ、認識率は 62-94% の範囲の値に収束した。)また、 $N=\infty$ すなわち手順 1 の下での認識率と節数の変遷を点線で重ね描いた。ただし認識率は $N=1000$ の場合と大差なく、

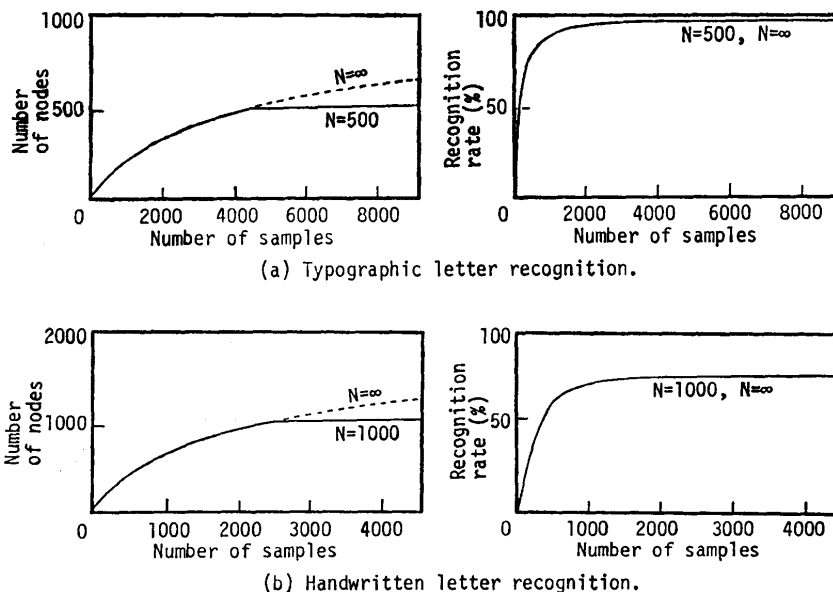


図5 実験結果
Fig. 5 Experiment results.

図では点線と実線とを区別できない。

誤認識の主な原因は、3.2節の活字認識の場合と同様、学習対象の一貫性が成り立たないことにある。しかしながら、一貫性の不成立は主に、筆記時点から画収録時点までの途中で混入する雑音というよりも、むしろ筆記の不安定さに起因する。例えば、乱暴な書き手によって書かれた *b*, *c*, *e*, *l* は、互いにほとんど判別できない。したがって、さらに認識率を上げるには、活字認識の場合と同様、単語の綴りを利用する必要がある。これについても、活字認識の場合とまとめて別稿で論ずることとする。

4. む す び

提案した手法を含むいろいろな文字認識手法において、処理手順は大きく分けて、画像から特徴を抽出する解析過程と、その結果を基に読みを出力する判定過程とから成る。効率の良い解析過程を設計する立場からは、文字画像と読みとを関連付ける特徴抽出モデルの精緻化に主眼を置くが、手書き文字認識への適用は難しい。また、テンプレートマッチングは、特徴抽出モデルがトリビアルな場合に当る。この方法は、解析途中で情報欠損が起こらないので、他の特徴抽出モデルに比べ扱いやすいが、テンプレートを記憶するためにしばしば巨大な記憶容量を必要とする。一方、判定過程に関しては、線形・非線形の積和を利用する方式や、データベースを利用する方式などがある。積和は

各々の特徴を基底として構成されるが、その特別な形態として、ファジー論理やニューラルネットワークに基づく方式がある。後者のニューラルネットワークは学習を行うが、本論文で提案した学習機械とは異なり、学習対象の本質的な構造を利用して学習を推進する形態にはなっていないので、学習速度と記憶素子の利用効率の点で問題が残る。また、データベース方式に関しては、文字画像と読みとの対が一括して多数個与えられたとき、データの探索時間をできる限り少なくするようなデータベースを構築する研究があるが、これは学習とは別の問題意識に基づく。ほかに、AVL木構築法⁹⁾のような動的手法があるが、過去に教示されたことがある文字画像に対する読みしか出力できないので、本用途には向かない。一方、自己組織化学習機械はデータベースを動的に構築するだけでなく、一種の連想能力を持っており、未教示の文字画像に対しても読みを出力する。

提案した文字認識手法は、テンプレートマッチングと自己組織化学習機械とを組み合わせた方式であり、両者のユニバーサル性を活用して手書き英字認識を可能にしている。しかしながら、自己組織化学習機械の利用形態はこの一つに限定されるものでなく、種々の特徴抽出モデルと組み合わせることによって、さらに認識率を改善し得る。例えば3.2節と3.3節の最後の部分で、単語の綴りを考慮に入れる可能性について述べたが、これは窓の範囲を広げると共に特徴抽出モデ

ルを高度化することを意味する。言い換えれば、自己組織化学習機械の特長は、その学習能力を利用して、特徴抽出モデルを最大限に活用する点にある。このような運用上の巨視的観点から見た自己組織化学習機械の有用性は未明であり、今後の研究によって十分な検討を加えなければならない。

まとめとして、木状データベースを自己組織化する学習機械を提案し、相互干渉のある英字認識へ応用した。この学習機械はあらかじめ学習用のモデルが与えられている方式と異なり、標本系列に応じて記憶構造を自己組織化してゆくことにより、要求された任意の応答特性を再現する。本論文では、この学習機械が文字認識へ有効であることを示したが、ほかにも多様な応用が考えられる¹⁰⁾。特に、似た入力信号の集合における局所構造は既知であるが全体構造は容易にはわからないような場合、本方法は威力を発揮する。

提案した学習機械の一つの欠点は、第2章で述べたように、データベースのレベル数の上昇が漸近的な意味でしか保証されていないので、実際には筆記漢字認識などの用途では非現実的なほどデータベースが巨大化する可能性があることである。データベースの規模を真の意味で最小化するための一方法として、 d の値を直接用いた連想規則でなく、その分布関数に基づいた規則を導入することが考えられるが、詳細については、実際のデータと照合しながら検討中である。

参 考 文 献

- 1) 塩野 充: 多重辞書類似度法による手書き漢字認識の基礎実験, *情報処理*, Vol. 27, No. 9, pp. 853-859 (1986).
- 2) Wang, Q.R. and Suen, C.Y.: Large Tree Classifier with Heuristic Search and Global Training, *IEEE Trans Pattern Anal. & Mach. Intell.*, Vol. PAMI-9, No. 1, pp. 91-102 (1987).
- 3) 中林 清ほか: 用語あいまい検索を用いた手書き文字列入力方式, 人工知能学会全国大会(第1回)論文集, pp. 435-438 (1987).
- 4) Hull, J.J.: Hypothesis Generation in a Computational Model for Visual Word Recognition, *IEEE Expert*, Fall, pp. 63-70 (1986).
- 5) Kurtzberg, J.M.: Feature Analysis for Sym-

bol Recognition by Elastic Matching, *IBM J. Res. Develop.*, Vol. 31, No. 1, pp. 91-95 (1987).

- 6) 鈴木 寿: 有本 卓: ユニバーサル学習機械と英活字認識への応用, *電子情報通信学会誌*, Vol. J70-A, No. 10, pp. 1472-1480 (1987).
- 7) Goldfarb, L.: A New Approach to Pattern Recognition, *Progress in Pattern Recognition 2*, North-Holland, Amsterdam (1985).
- 8) Feller, W.: *Probability Theory and Its Applications*, John Wiley & Sons Inc., New York (1950).
- 9) Wirth, N.: アルゴリズム+データ構造=プログラム, 片山卓也(訳), pp. 245-257, 日本コンピュータ協会, 東京 (1976).
- 10) 鈴木 寿, 大西弘之, 有本 卓: 連想データベース自己組織化と自律移動知能ロボットへの応用, *日本ロボット学会誌*, Vol. 5, No. 6, pp. 431-441 (1987).

(昭和62年9月10日受付)

(昭和63年3月3日採録)

鈴木 寿 (正会員)

昭和34年生。大阪大学基礎工学部生物工学科, 昭和58年卒業。同大学大学院基礎工学研究科物理系専攻機械工学分野, 昭和63年修了。工学博士。現在, 同学助手。データ圧縮, 暗号, ロボットの知能化に関する研究に従事。電子情報通信学会, 日本ロボット学会, 人工知能学会などの会員。

有本 卓 (正会員)

昭和11年生。昭和34年京都大学理学部卒業。同年沖電気(株)入社。電子計算機の開発に従事。37年東京大学工学部助手, 42年講師, 43年大阪大学基礎工学部助教授, 48年教授, 42年工学博士。この間, 情報理論, 制御理論, デジタル信号処理の高速アルゴリズムに従事する一方, ロボット工学に興味をもち, そのインテリジェント化の研究開発を行っている。電子情報通信学会, 計測自動制御学会, 日本機械学会, ロボット学会各会員。IEEEのFellow会員。