

テンプレートマッチングによるオフライン手書き文字認識 ニューラルネットワークの作成

篠 沢 佳 久[†] 大 駒 誠 一^{††}

現在さまざまな文字認識の研究が行われているが、最も基本となる方法は認識したい未知文字パターンから抽出した特徴と任意の文字種の代表となる文字パターン(テンプレート)から抽出した特徴との距離を識別関数を使用して求め、最も距離の短いものを認識結果とするテンプレートマッチング方式である。本論文ではこのテンプレートマッチング方式をニューラルネットワークを用いて実現する方法を提案する。テンプレートとして任意の文字種の代表となる特徴と認識したい未知文字パターンの特徴を入力層に同時に入力したとき、この2つの文字種が同じ場合は出力層が発火し、異なる場合は発火しないような挙動をするニューラルネットワークを構築する。そこで問題となるのが、そのニューラルネットワークの出力層を発火させる文字種(教師信号)、発火させない文字種(非教師信号)をどのように選び、どのように学習させていくかが課題となる。そこで本論文ではこのテンプレートマッチング方式を基本としたニューラルネットワーク構築のための学習方法として、他のニューラルネットワークとの協調を考慮して、非教師信号を選択しながら学習を進めていく方法を提案する。そして本手法に従い文字画像データベース ETL9B を利用しニューラルネットワークを構築した結果、認識率において優れていることを示すことができた。

The Neural Networks for Offline Handwritten Character Recognition by Template Matching

YOSHIHISA SHINOZAWA[†] and SEIICHI OKOMA^{††}

Recently character recognition is studied by many researchers. One of basic approaches for character recognition is template matching. In this approach a character recognition system calculates distances between an unknown pattern and all template patterns and searches the template pattern that is the shortest distance between an unknown pattern and decides it as recognition result. We try to construct the neural networks for character recognition based on template matching. A template pattern and an unknown pattern are input into the input units of neural networks at same time. If two patterns are same, an output unit of neural networks is fired. But if these are different, it isn't fired. In case of learning such neural networks, it is important how to select excitement and inhibition samples and learn them. We propose how to learn excitement and inhibition samples as cooperating with the other neural networks and selecting proper samples. In this work we constructed neural networks for ETL9B. The results obtained from our work show an increased recognition rate of character patterns.

1. ま え が き

手書き文字認識の研究は古くから行われており、今日に至るまでさまざまな認識方法が考案されている。字体が書き手によって異なり、筆順情報を利用できないオフライン手書き文字認識の研究はパターン認識の分野において最も困難な課題の1つである。特にわが国において使用されている文字種は JIS 規格のものだ

けでも 6349 個存在するので、1 回の認識処理で文字候補を 1 つに絞り込むのは困難である。そこでまず始めに認識したい未知の文字パターンに対しておおまかに分類を行い、文字候補の数を絞り込む(大分類)。そして次に大分類によって絞り込まれた文字候補の中からさらに詳しく認識を行い、最終的に候補を 1 つに絞り込む(詳細認識)という 2 段階による認識方法がとられている。

基本的な認識方法としては認識したい未知の文字パターンから抽出した特徴と認識対象となるすべての文字種の標準文字パターンから抽出した特徴との距離を 1 つ 1 つ求める。そしてその結果から最も距離の短い

[†] 慶應義塾大学インフォメーションテクノロジーセンター
Information Technology Center, Keio University

^{††} 慶應義塾大学理工学部管理工学科

Faculty of Science and Technology, Keio University

ものを認識結果とする総当たりのテンプレートマッチング方式が基本となっている。この方式では日本語のように認識対象となる文字候補の数が多いと膨大な計算量を必要とする。そのため上述した大分類・詳細認識の手法を利用し、まず大分類では計算量の少ない簡単な特徴およびユークリッド距離のような簡単な識別関数を使用する。一方で詳細認識においては大分類で使ったものよりも複雑な特徴およびより多くの計算量を必要とするが、より精度の良い識別関数を使用することによって総合的に計算量を少なくしながら、認識率を向上させる工夫が試みられている。

またテンプレートマッチング方式のほかにニューラルネットワークを用いた認識方法^{1),2)}も多く考案されている。ニューラルネットワークを用いた場合の利点として学習サンプルを適切に提示することによって、異なるカテゴリをより厳密に判別できる識別関数を学習により柔軟に作成できるということがあげられる。またニューラルネットワークを用いた場合、認識時にテンプレートマッチング方式のように対象となるすべての文字種と比較する必要はないので、高速で精度の良い文字認識システムを構築することができる。しかしこれは比較的、認識対象とする文字種の数が増加するにつれてニューラルネットワークの規模も大きくなり学習が困難になるという問題が生じてしまう。そのため単一のニューラルネットワークで数千種類の文字種を認識対象とするのは困難である。

そこでニューラルネットワークを大規模な文字認識に適用する場合、大分類・詳細認識による方法を応用する。すなわち大分類により文字候補の数をできるだけ減らした後、詳細認識においてそれらの文字候補のみに対してニューラルネットワークを適用する。たとえば CombNET^{3),4)}では大分類部で学習ベクトル量子化を利用して文字候補数を絞り込む。次に詳細認識部では絞り込まれた各カテゴリごとでバックプロパゲーションネットワークを利用して文字候補を1つに絞り込むという2段階の認識方法を行っている。通常ニューラルネットワークで大規模な文字認識を構築する場合、複数のニューラルネットワークに機能を分散させることにより対処している。

このように2つの方法は互いに一長一短の長所を持っている。本論文ではこれら2つの手法の特性を活かしたテンプレートマッチング方式による大規模な文字認識ニューラルネットワークを提案する。提案するニューラルネットワークはテンプレートマッチング方式、すなわちある特定の識別関数はそれと対応した

1種類の文字種のみを認識を行うのと同様に、特定のニューラルネットワークはそれと対応した1種類の文字種のみを認識を行う。基本的に1つのニューラルネットワークにおいて認識対象とする文字種の数が増加するほど、学習が難しくなり認識率は低下していく傾向がある。そのため1つのニューラルネットワークにおいてはできるだけ少ない文字種を認識対象とした方がよい。特に1つのニューラルネットワークの認識対象数を最少の1つのみとする考え方はELNET(排他的学習ネット)^{5),6)}によって実現され、その有効性も確認されているため問題はないはずである。このようにニューラルネットワークを利用する利点として柔軟に識別面を構築でき、認識率の向上も期待できる。また計算量に関しても大分類・詳細認識の方式を採用すれば抑えることができるはずである。

テンプレートマッチング方式の基本は、あらかじめ用意してある任意の文字種を代表する標準文字パターン(テンプレート)の特徴と認識したい未知文字パターンの特徴との距離を求める。そして比較しているテンプレートの文字パターンの特徴との距離が短ければ、認識結果としてそのテンプレートと類似している可能性が高いことになる。この方式をニューラルネットワークで実現するには、テンプレートとして任意の文字種の代表となる特徴と認識したい未知文字パターンの特徴を入力層に同時に入力したとき、この2つが同字種である場合は出力層が発火し、異なる場合は発火しないような挙動をするニューラルネットワークを構築すればよい。すなわちあらかじめ認識対象と同じ個数のニューラルネットワークを用意し、各文字種を個々のニューラルネットワークと対応づける。そして上記のようにテンプレートの標準文字パターンと認識したい未知の文字パターンを同時に入力したとき、2つの文字種が同一な場合のみ発火し、異なる場合は発火しないように学習させる。すると未知の文字パターンを入力した場合、比較しているテンプレートの文字パターンと類似していれば発火し、類似していなければ発火しなくなるものと期待できる。

テンプレートマッチング方式による識別関数の構築は多くの学習用のサンプル文字パターンを使用して統計的に各パラメータを調整することにある。一方ニューラルネットワークによる識別関数の構築は、同じく多くのサンプルを使用して微調整を繰り返しながら適切なパラメータを求めていく。結局のところパラメータの調整とは、文字パターンが分布している特徴空間上においてニューラルネットワークの出力を発火させなければならない文字種(教師信号)と抑制させ

なければならない文字種（非教師信号）との間に適切な識別面を生成することにある．したがって，教師信号と非教師信号の選び方，特に後者の選び方が重要になる．本論文で提案するニューラルネットワークにおいて非教師信号は教師信号の対象以外の文字種すべてを使えばよいと考えがちであるが，学習対象となる文字種の数が増加するにつれて，ニューラルネットワークの規模も大きくしなければならず，学習も容易に進まなくなってしまう．そのため非教師信号の対象となる文字種は適切に選択しなければならない．

そこで本論文では上述したテンプレートマッチング方式によるニューラルネットワークを構築するうえでまずニューラルネットワークの構造について述べ，次に他のニューラルネットワークとの協調を考慮し学習，認識を進めながら適切に非教師信号の文字種を自動的に選択していく学習方法を提案する．そして電子技術総合研究所提供の文字画像データベース ETL9B を用いた認識実験を通してその有効性を評価する．

2. ニューラルネットワークの構造と学習方法

2.1 ニューラルネットワークの構造

提案するニューラルネットワークにおいて特定のニューラルネットワークはそれに対応した1種類の文字種のみを認識を行う．したがって，認識対象数を N 個とすると同数の N 個のニューラルネットワークが必要となる．例として図1に『あ』を認識対象としたニューラルネットワークを示す．1つ1つのニューラルネットワークはフィードフォワード型で入力層，中間層，出力層の3層から構成される．入力層には認識したい未知文字パターンの特徴と比較対象となる文字パターン（この場合は『あ』である）の特徴を同時に入力する．入力層において未知文字パターンの特徴を入力する部分を認識対象部，比較対象となる文字パターンを入力する部分を比較対象部と呼ぶ．入力層のニューロンの個数は2つの文字パターンの特徴の合計数分を必要とする．中間層のニューロンの個数は任意とする．出力層のニューロンの個数は1個と固定する．

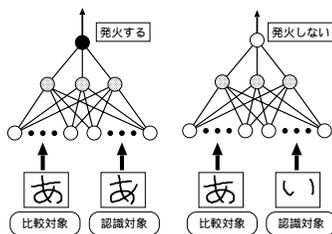


図1 ニューラルネットワークの構造

Fig. 1 The structure of a neural network.

そして入力層の比較対象部にテンプレートとなる『あ』の特徴を入力すると同時に認識対象部には認識したい未知文字パターンの特徴を入力する．入力した2つの文字種が同じものである場合は出力層のニューロンが発火し，異なる場合は発火しないような挙動をさせる．すなわち任意の文字パターン i と対応したニューラルネットワーク F_i において，テンプレートとなる文字パターンの特徴を T_i ，認識したい未知文字パターンの特徴を x としたとき，

$$F_i(x, T_i) = \begin{cases} 1.0 & \cdots x \in i \\ 0.0 & \cdots x \notin i \end{cases} \quad (1)$$

というような挙動をする識別関数 F_i をニューラルネットワークで構築する．

実際に未知文字パターンを認識させる場合，図2に示すようにまず対応したテンプレートの文字パターンのみに対して上記のような挙動をするニューラルネットワークを N 個配置する．そしてそれぞれのニューラルネットワークの認識対象部には未知文字パターンの特徴を入力する．比較対象部にはそのニューラルネットワークと対応したテンプレートの文字パターンの特徴を入力し，出力層の発火が最も大きいものを認識結果とすればよい．

2.2 学習方法

最も重要なことは個々のニューラルネットワークの学習方法である．認識対象数を N として任意のニューラルネットワークを学習する場合，教師信号はそのニューラルネットワークと対応した文字種1個が対象であり，非教師信号は残り $N - 1$ 個の文字種が対象となる．しかし非教師信号として残り $N - 1$ 個の文字種を学習で使用するのは非効率である．なぜならばテンプレートの文字パターン（教師信号）と類似している，すなわち特徴空間上で教師信号と近い位置関係にある文字パターンのみを非教師信号として使用して適切にその識別面を構築することができれば，類似していない，すなわち特徴空間上で遠い位置関係にある文字パターンとの区別がつくものと期待できるからで

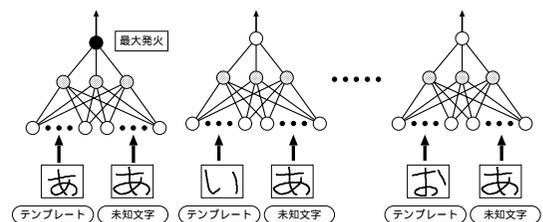


図2 提案するニューラルネットワークでの認識

Fig. 2 Recognition on the neural networks.

ある．またニューラルネットワークの学習時において学習対象となる文字種の数が増加するほど一般的に学習が困難になるので，非教師信号をただ増やすのは得策ではない．そこで教師信号と類似している文字種のみを選ぶためにはあらかじめユークリッド距離などを用いて大分類を行っておく．そして大分類の結果からその上位候補に列挙された文字候補のみを非教師信号の対象として考慮すればよい．

次にそのような類似した文字パターンのみを使用してどのように学習させていけばよいかを考える．大分類の上位に出現する文字候補すべてを非教師信号の対象としてもよいが，本論文では各ニューラルネットワークの挙動から自動的に非教師信号の対象を選択し，適切に教師信号の文字種と非教師信号の文字種との間の識別面を構築する協調学習を提案する．

たとえば図2に示したような『あ』『い』『う』『え』『お』の5文字を認識対象とした5個のニューラルネットワーク群を考える．仮に『あ』を認識させようとして，各ニューラルネットワークの出力層の発火の状況から，第1候補から順に『おえあうい』という認識結果になったとする．この場合『おえ』を対象としたニューラルネットワークの発火を抑制する必要がある．すなわち，

- 『お』および『え』を対象としたニューラルネットワークは『あ』を入力すると発火しやすいことから，非教師信号として『あ』は使用する．
 - 『う』および『い』を対象としたニューラルネットワークは『あ』を入力しても発火しにくいことから，非教師信号として『あ』は使用しない．
- といったように『あ』を非教師信号として使用するニューラルネットワークを決める．他の4文字についても同様に認識を行い，それぞれの文字パターンをど

のニューラルネットワークで非教師信号として使用するかを定める．すなわち他のニューラルネットワークすべての挙動を考慮したうえで適切な非教師信号を決めていく．

そして各ニューラルネットワークで使用する非教師信号を決定したならば，図3に示すように比較対象部には教師信号と同じ文字種の特徴をあらかじめ入力しておき，認識対象部に教師信号と同じ文字種の特徴を入力した場合は出力層を発火させるように学習させる．一方で認識対象部に非教師信号の文字種の特徴を入力した場合は出力層を発火させないように各ニューラルネットワークを学習させる．個々のニューラルネットワークの学習にはバックプロパゲーションアルゴリズム（誤差逆伝播アルゴリズム）を用いた．

このような学習を行った後ではニューラルネットワークの挙動は変化しているはずである．たとえば学習後，先程の『おえあうい』という認識結果から『うあいおえ』というような認識結果になってしまったとする．この場合『おえ』のニューラルネットワークは抑制されるようになったが『う』のニューラルネットワークが発火しやすくなってしまっている．このように新たに別の文字種が発火しやすくなってしまいう場合もある．この場合『う』を対象としたニューラルネットワークにおいて『あ』を入力しても出力層が発火しないように新たに学習しなければならない．そのため学習後，再び認識を行い非教師信号としてまだ他の文字種が必要かどうかを調べる必要がある．

すなわち他のニューラルネットワークの挙動を考慮しながら認識，非教師信号の決定そして学習というループを非教師信号の個数が一定（学習文字パターンの認識率が100%に達する）になるまで繰り返せばよい．学習方法は以下のとおりである．認識対象となる文字種の数 N 個，したがってニューラルネットワークの個数も N 個である．

step1

学習文字パターンを用いて大分類を行い，各文字パターン i ($i = 0, 1, 2, \dots, N$) ごとに第 P 候補までに出現する文字候補の集合 C_i を調べる．そして文字候補の集合 C_i の中から，第 i ニューラルネットワークで最初の学習時において使用する非教師信号をランダムに選び，その集合を T_i とする．その際ニューラルネットワーク1個あたりの非教師信号の平均個数を t_{old} とする．

step2

各ニューラルネットワークにおいて非教師信号 T_i を使用して教師信号と非教師信号の文字パターン

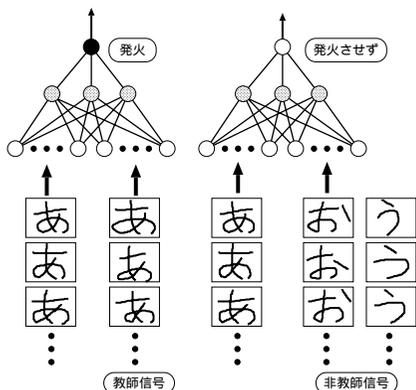


図3 提案するニューラルネットワークでの学習
Fig. 3 Learning the neural networks.

の判別ができるまでバックプロパゲーションアルゴリズムによって学習を行う。

step3

文字種 i に対してはその文字候補の対象を集合 C_i として認識を行い、それぞれのニューラルネットワークにおいて新たな非教師信号の集合 T'_i を求める。非教師信号の求め方は以下のとおりである。

文字種 s ($s \in C_i$) の特徴を x_s とする。任意のニューラルネットワーク F_i ($i \neq s$) において特徴 x_s を入力したときの出力値を $F_i(x_s, T_i)$ とする。一方、文字種 s と対応したニューラルネットワーク F_s にその特徴 x_s を入力したときの出力値を $F_s(x_s, T_s)$ とする。そこで出力値 $F_i(x_s, T_i)$ が $F_s(x_s, T_s)$ よりも大きければニューラルネットワーク F_i の出力を抑制する必要があるので、文字種 s をニューラルネットワーク F_i の非教師信号として用いる。

$$F_s(x_s, T_s) < F_i(x_s, T_i) \rightarrow s \in T'_i \quad (i \neq s)$$

一方、出力値 $F_i(x_s, T_i)$ が $F_s(x_s, T_s)$ より小さければ、抑制する必要はないので文字種 s を非教師信号として用いる必要はない。

$$F_s(x_s, T_s) > F_i(x_s, T_i) \rightarrow s \notin T'_i \quad (i \neq s)$$

また以前非教師信号として登録したものはそのまま残しておく。このように全文字種についてどのニューラルネットワークの非教師信号として用いるかを調べた後、ニューラルネットワーク 1 個あたりの非教師信号の平均個数を t_{new} とする。

step4

学習後の非教師信号の平均個数 t_{new} と学習前の非教師信号の平均個数 t_{old} との差が小さく、非教師信号の個数に変化がなければ非教師信号 T'_i によって学習はせずに終了する。一方で非教師信号の個数が増加している場合、非教師信号 T'_i によって再び学習する必要があるので、 $t_{new} \rightarrow t_{old}$ 、 $T'_i \rightarrow T_i$ とした後に step2 に戻る。

以上のようにニューラルネットワーク 1 個あたりの非教師信号の平均個数が一定になるまで学習を続ける。

3. 認識実験

3.1 実験条件

提案手法の有効性を評価するために認識実験を行う。実験には電子技術総合研究所提供の ETL9B, 3036 文字を利用した。実験には最初の 1 セット目と次の 2 セット目を使用した。表 1 に示すようにニューラルネット

表 1 実験方法

Table 1 The way to use ETL9B database.

	学習文字パターン	未学習文字パターン
方式 1	1 セット	2 セット
方式 2	2 セット	1 セット

ワーク構築のため学習文字パターンとして 1 文字あたり最初の 1 セット目 40 パターンを利用し、認識実験で利用する未学習文字パターンとして 2 セット目の 40 パターンを利用する方法 (方式 1) および学習文字パターンと未学習文字パターンを方式 1 とは逆に交換して利用する方法 (方式 2) の 2 種類を行う。

各文字パターンには前処理として非線形正規化処理⁷⁾を施し 48×48 の大きさにした後で 196 次元の輪郭線特徴⁸⁾を抽出した。まず輪郭線特徴を利用して大分類を行い、各文字種 i ごとで第 P 候補まで出現する文字種の頻度を調べた後、その集合 C_i を非教師信号の候補として、前述した学習方法によってニューラルネットワークの構築を行う。

各ニューラルネットワークの構造は次のとおりである。入力層のニューロンの個数は使用する特徴の次元数が 196 なので 2 倍の 392 個となる。また中間層は 16 個、出力層は 1 個と固定した。学習の高速化のため入力層は 8 個の局所領域 (それぞれ 7×7 の領域) に分割し、入力層と中間層は局所結合⁹⁾とした。

ニューラルネットワーク 1 個あたりのリンク数は入力層と中間層の間が 784 ($= \frac{392}{8} \times 16$) 個、中間層と出力層の間が 16 個で合計総リンク数は 800 個である。学習方法はモーメント法¹⁰⁾を使用した。学習時のパラメータの設定はすべての実験で同一とした。また step2 のバックプロパゲーションアルゴリズムにおける 1 回あたりの学習回数は 100 回と固定した。本来学習については非教師信号の個数が一定になるまで step2 から step4 を無限に繰り返すのだが、実験においてはこれを固定して 3 回繰り返すことにした。

実験は大きく分けて 2 通り、大分類部で使用する識別関数を変え非教師信号の候補集合を変化させ、大分類で異なる識別関数によって生成されるニューラルネットワークの違いについて調べてみた。大分類には計算量が比較的少なく済むユークリッド距離と重み付けユークリッド距離を使用した。未知文字パターンの特徴 $x^t = (x_1, x_2, \dots, x_i, \dots, x_{196})$ と標準文字パターン j の特徴 $a_j^t = (a_{j1}, a_{j2}, \dots, a_{ji}, \dots, a_{j196})$ としたとき、2 つの文字パターン間のユークリッド距離 d_j は、

表 2 使用した文字パターンの大分類の結果

Table 2 The classification rate of character patterns.

大分類の方法	ユークリッド距離の場合の認識率 (%)				重み付けユークリッド距離の場合の認識率 (%)			
	学習文字		未学習文字		学習文字		未学習文字	
	方式 1	方式 2	方式 1	方式 2	方式 1	方式 2	方式 1	方式 2
候補数								
第 1 候補	94.95	94.74	92.17	92.34	97.92	97.71	93.10	93.52
第 10 候補	99.36	99.39	99.00	98.97	99.83	99.83	98.99	99.08
第 20 候補	99.62	99.64	99.42	99.36	99.91	99.91	99.35	99.41

表 3 大分類にユークリッド距離を使用した場合

Table 3 The recognition rate of character patterns by Euclidean distance.

識別方法	非教師信号の数 (個/字種)		認識率 (%)			
	方式 1	方式 2	学習文字		未学習文字	
			方式 1	方式 2	方式 1	方式 2
ユークリッド距離	—	—	94.95	94.74	92.17	92.34
提案方法 ($M = 0$)	43.62	44.43	99.03	99.00	92.52	92.65
提案方法 ($M = 1$)	54.32	60.20	99.02	99.11	92.80	93.21
提案方法 ($M = 2$)	61.93	67.77	99.05	99.16	92.97	93.24
$P = 10$	81.06	82.14	99.08	99.07	93.00	93.25
$P = 20$	148.87	150.88	99.35	99.30	93.04	93.29

$$d_j = \sum_{i=1}^{196} (x_i - a_{ji})^2 \quad (2)$$

のように求めることができる。また標準文字パターン j の分散 $v_j^t = (v_{j1}, v_{j2}, \dots, v_{ji}, \dots, v_{j196})$, $V_j = \sum_{i=1}^{196} v_{ji}$ とすると 2 つの文字パターン間の重み付けユークリッド距離 wd_j は,

$$wd_j = \sum_{i=1}^{196} \frac{V_j}{v_{ji}} \cdot (x_i - a_{ji})^2 \quad (3)$$

のように求めることができる。まずこの 2 つの識別関数による大分類の結果を表 2 に示す。

実験には大分類の結果から

- 第 10 候補 ($P = 10$)
- 第 20 候補 ($P = 20$)

までに現れるすべての文字パターンを無条件に非教師信号として学習する方法と提案方法を比較する。しかし提案方法のままだと多少問題がある。第 1 候補が正解文字の場合、非教師信号として何も選択されなくなってしまう。学習が収束に向かいつつある段階で、第 1 候補が正解文字であっても問題ではないが、学習の初期過程においてこうしたことがあると学習に必要な非教師信号が十分に選択されない場合が想定される。

そこで非教師信号として第 10 候補内で正解となった順位からさらに下位 M 個の候補までを用いることにする。これについては前述した『あいうえお』を認識対象としたニューラルネットワークの例で説明する。たとえば『あ』を認識させた場合、認識結果が『おあ

表 4 ユークリッド距離の場合の認識率の向上

Table 4 The changes of recognition rate by Euclidean distance.

学習回数	非教師信号の数 (個/字種)		認識率 (%)			
	方式 1	方式 2	学習文字		未学習文字	
			方式 1	方式 2	方式 1	方式 2
乱数	—	—	27.12	8.65	18.04	6.42
1 回目	49.07	55.55	95.28	98.21	90.68	92.46
2 回目	59.50	66.66	98.49	99.08	92.75	93.19
3 回目	61.93	67.77	99.05	99.16	92.97	93.21

えうい』となったものとする。そこで $M = 2$ とするとこれまでどおり『お』を対象としたニューラルネットワークのみで非教師信号として『あ』を利用するのではなく、正解となった順位からさらに下位 2 つの候補『え』と『う』を対象としたニューラルネットワークも『あ』を非教師信号として利用するという方法である。 M については正解の文字候補までを含めた方法 ($M = 0$)、および正解の文字候補より下位第 1, 第 2 候補までを用いた方法 ($M = 1, 2$) について調べた。

その他細かいことでは最初に学習する際 (step2)、認識対象部には非教師信号の対象として集合 C_i からランダムに文字パターンを選択しているのではなく、認識対象部には乱数値を入力している。また認識時において比較対象部に入力するテンプレートの文字パターンは、40 個ある学習文字パターンの中からランダムに選ぶことにしている。

3.2 大分類にユークリッド距離を使用した場合

大分類にユークリッド距離を使用した場合の認識結果を表 3 と表 4 に示す。各表に提示する情報は表 3

表5 大分類に重み付けユークリッド距離を使用した場合

Table 5 The recognition rate of character patterns by weighted Euclidean distance.

識別方法	非教師信号の 数(個/字種)		認識率(%)			
			学習文字		未学習文字	
	方式1	方式2	方式1	方式2	方式1	方式2
重み付けユークリッド	—	—	97.92	97.71	93.10	93.52
提案方法($M=0$)	42.86	46.02	99.36	99.46	93.18	93.54
提案方法($M=1$)	53.65	64.17	99.41	99.48	93.40	93.83
提案方法($M=2$)	61.10	71.32	99.44	99.48	93.52	93.91
$P=10$	82.18	83.64	99.44	99.42	93.52	93.92
$P=20$	151.17	153.48	99.57	99.53	93.57	93.95

には1字種あたりの非教師信号の平均個数, 学習文字パターンの認識率および未学習文字パターンの認識率を示す. 表4には $M=2$ の場合の提案方法において学習アルゴリズム中のstep2からstep4までの繰返し回数の経過とともに, 1字種あたりの非教師信号の平均個数, 学習文字パターンおよび未学習文字パターンの認識率が上昇していった過程を示す.

表3からユークリッド距離を利用した場合と比較して提案方法($M=2$)の場合, 方式1において学習文字パターンでは4.1%(4981文字)未学習文字パターンでは0.8%(972文字), 方式2において学習文字パターンでは4.42%(5368文字), 未学習文字パターンでは0.9%(1093文字)認識率が向上した.

表4より認識対象部に非教師信号の文字パターンの特徴ではなくただの乱数値を入力した学習方法ではまったく認識できておらず, 各学習段階が進むにつれ認識率が向上していく経過が分かる. 一方で学習の繰返し回数の2回目と3回目では非教師信号の個数にそれほど差がなく, それにともない認識率の上昇も飽和状態に達していることが分かる.

また表3からニューラルネットワークを用いた場合, いずれの方法でも識別関数にユークリッド距離を用いた場合の認識結果を上回った. 表3において提案方法の $M=0$ と $M=2$ の場合を比較すると, $M=0$ の場合にはそれほど認識率が向上していない. その原因は前述したように $M=0$ の場合には正解が第1候補にあると, 非教師信号として何も選ばれなくなってしまうことがあげられる. この場合非教師信号なしで学習しなければならなくなってしまう. また教師信号の文字パターンとそれとわけて類似している非教師信号の文字パターンのみを識別する境界線を厳密に生成すると, 学習文字パターンに対しては問題ではないが, 未学習文字パターンの認識率がそれほど向上しない. すなわち汎用性がそれほど向上しないことが分かる. 大分類の結果において正解文字のすぐ上位・下位にある文字候補はその正解文字と類似していると判断でき

表6 重み付けユークリッド距離の場合の認識率の向上

Table 6 The change of recognition rate by weighted Euclidean distance.

学習 回数	非教師信号の 数(個/字種)		認識率(%)			
			学習文字		未学習文字	
	方式1	方式2	方式1	方式2	方式1	方式2
乱数	—	—	26.93	7.55	17.54	5.23
1回目	48.81	61.80	96.94	98.75	92.08	93.33
2回目	59.11	70.45	99.10	99.48	93.34	93.91
3回目	61.10	71.32	99.44	99.48	93.52	93.91

る. そのためニューラルネットワークが誤りであると判断しなかった文字パターンについてもテンプレートの文字パターンと類似している, すなわち正解文字のすぐ下位候補にある文字パターンについても, 非教師信号として学習した方が未学習文字パターンに対する認識率の向上を図れることが分かる.

その一方で表3より提案方法($M=2$)と $P=20$ の結果を見ると, 非教師信号の個数を2倍以上に増やしても未学習文字パターンにおいて認識率はほとんど向上しない. これは逆に多くの下位候補を非教師信号として利用しても大きく認識率の向上には影響しないということである. すなわち非教師信号としては, テンプレートの文字種と類似しているものだけを選択し学習した方が効果的であり, 提案する学習方法では少ない個数の非教師信号を利用して効率的に認識率の向上に役立つことが分かった.

3.3 大分類に重み付けユークリッド距離を使用した場合

次に大分類に重み付けユークリッド距離を使用した場合の認識結果を表5, 表6に示す. 提示する情報は先程の表と同様である. 表5, 表6からユークリッド距離の結果と同じ傾向にあることが分かる. 表6より学習の繰返し回数の2回目と3回目を比較すると, 非教師信号の個数が一定になるにつれて認識率も向上しなくなっていることが分かる. 表5から提案方法($M=2$)では, 重み付けユークリッド距離を利用した場合と比較して, 方式1において学習文字パター

表7 大分類・詳細認識(ユークリッド距離)

Table 7 The result of rough classification and detail recognition by Euclidean distance.

テンプレート の入力方法	認識率(%)			
	ランダムに入力		平均値を入力	
	方式1	方式2	方式1	方式2
文字候補数(個)				
2	93.29	93.42	93.30	93.42
3	93.33	93.53	93.34	93.53
4	93.27	93.50	93.28	93.51
5	93.21	93.46	93.21	93.46
ユークリッド 距離のみの場合	方式1		方式2	
	92.17		92.34	

表8 大分類・詳細認識(重み付けユークリッド距離)

Table 8 The result of rough classification and detail recognition by weighted Euclidean distance.

テンプレート の入力方法	認識率(%)			
	ランダムに入力		平均値を入力	
	方式1	方式2	方式1	方式2
文字候補数(個)				
2	93.93	94.23	93.93	94.23
3	93.96	94.24	93.96	94.24
4	93.89	94.19	93.91	94.20
5	93.83	94.13	93.84	94.14
重み付けユークリッド 距離のみの場合	方式1		方式2	
	93.10		93.52	

ンでは1.52%(1851文字)未学習文字パターンでは0.42%(510文字),方式2において学習文字パターンでは1.77%(2150文字)未学習文字パターンでは0.39%(473文字)認識率が向上した.またユークリッド距離の場合と同様に,ニューラルネットワークを利用した方法ではいずれの方法でも重み付けユークリッド距離の結果を上回った.また非教師信号の個数を増やし下位候補まで非教師信号として利用しても,大きく認識率の向上に影響していない.下位候補 M については以上2つの実験結果より $M=0$ とするとそれほど認識率の向上が望めないことから $M=2$ と固定し引続き実験を行う.

3.4 大分類・詳細認識

次に大分類部と協調した大分類・詳細認識型の認識方法を行う.まず大分類でユークリッド距離または重み付けユークリッド距離を識別関数として使用して文字候補を絞り込む.そして次に絞り込まれた文字候補と対応したニューラルネットワークのみを使用して詳細認識を行う.大分類で使用する識別関数の値 $Dist_i$ と詳細認識部で使用するニューラルネットワークの出力値 Out_i は尺度が異なる($Dist_i$ は0から無限大までの値をとり0に近い値ほど良く, Out_i は0から1の値で1に近い値ほど良い)ので,

$$Sim_i = Dist_i \times (1.0 - Out_i)$$

という協調計算を行い, Sim_i を最小とする文字候補 i を最終的な認識結果とする.実験条件は前述したものと同様である.大分類部においてユークリッド距離および重み付けユークリッド距離を用いて絞り込む文字候補数を変化させたうえで,詳細認識部においては提案方法の学習結果を利用して認識結果を1つに絞り込む.また先程は入力層の比較対象部にはテンプレートとしては40個の学習文字パターンの中からランダムにどれか1つを選ぶだけであったが,今回は40個の学習文字パターンの特徴の平均値を入力させる試みも行った.その結果を表7,表8に示す.表7,表8

表9 認識結果の詳細

Table 9 The number of right answers and errors on each result.

大分類での 識別関数	実験方式	正解数 (個)	誤認識数 (個)	増加数 (個)
ユークリッド 距離	方式1	2340	911	1429
	方式2	2207	761	1446
重み付け ユークリッド距離	方式1	1962	917	1045
	方式2	1732	857	875

に提示した情報は大分類で絞り込む文字候補数を変化させていった場合の未学習文字パターンの認識率を示した.比較対象部にランダムにテンプレートを選んだ場合および特徴の平均値を入力した場合の2種類を調べた.表7には大分類でユークリッド距離を使用した結果を,表8は重み付けユークリッド距離を使用した結果を示す.

表7,表8の結果から,認識方法を大分類・詳細認識型にすることでユークリッド距離の場合,文字候補数が3で比較対象部にテンプレートとして特徴の平均値を入力したときにおいて方式1では1.17%(1429文字),方式2では1.19%(1446文字)認識率の向上が図れた.また重み付けユークリッド距離の場合,同様の条件で方式1では0.86%(1045文字),方式2では0.72%(875文字)認識率の向上が図れた.

また表9にそれぞれの結果による認識結果の向上の内訳を示す.これは認識率の向上の内訳を大分類で使った識別関数では認識できなかったが,ニューラルネットワークを使用したところ正確に認識できた個数(正解数)および反対に大分類で使った識別関数では正確に認識できたのだが,ニューラルネットワークを使用したところ誤認識してしまった個数(誤認識数)の2つに分けた.表9からニューラルネットワークを用いることによって大分類で誤認識した結果を訂正しているだけでなく,正解であった結果を誤認識してしまう場合もそれなりに多いことが分かる.具体的に例をあげると『し』という文字パターンはユークリッド

距離の場合、未学習文字パターン 40 個中、34 個正しく認識していたのだが、提案方法によって認識させたところ 40 個中、29 個しか正しく認識できなくなってしまっていた。この場合「し」と「レ」のニューラルネットワークからの出力値がともに大きかった。学習文字パターンにおいてはユークリッド距離の場合 40 個中、33 個しか正しく認識していなかったのだが、提案方法によって 40 個すべて正しく認識できるようになったので学習ができなかったわけではなく、汎用性がなかったものと判断できる。類似文字についてはこうしたことが多く見受けられた。一般論ではあるがニューラルネットワークにおいて汎用性を高めるためにはより有効な特徴を利用する、学習文字パターン数を増やすといった処置が必要となる。しかし誤認識数と比較して正解数の多いことから、提案方法による再認識の効果はあるものと判断できる。

表 7、表 8 の結果から比較対象部にテンプレートとしてランダムに学習文字パターンを選んで入力した場合と平均値を入力した場合とでは大きな差はない。しかし表 3 と表 5 の結果から分かるように学習文字パターンの認識率が 100%になっていないため、学習できなかった文字パターンが比較対象部への入力として選ばれると認識できないこともあり、わずかではあるが平均値を入力する場合よりも認識率が低かった。

また表 7、表 8 の結果からユークリッド距離および重み付けユークリッド距離の結果とも大分類部での文字候補数を 3 としたときを境にして、文字候補数を増やすと認識率が低下する傾向にあることが分かる。

4. 考 察

本論文で提案するニューラルネットワークの学習について考察を行う。各ニューラルネットワークにおいて認識対象となる文字数は 1 文字であり、ELNET の事例からして、このようなニューラルネットワークを学習するのは比較的容易なはずである。事実ほぼすべてのニューラルネットワークにおいて学習は収束した。これは表 3 および表 5 から分かるように、学習文字パターンの認識率がほぼ 100%に達していることから明らかである。認識対象を 1 文字とするニューラルネットワークの学習は容易であることが分かった。

次にニューラルネットワークへの入力と出力について考察する。たとえば文字種 i と対応したニューラルネットワークの識別関数 F_i について出力層の発火の挙動を式 (1) とは逆にする、

$$F_i(\mathbf{x}, \mathbf{T}_i) = \begin{cases} 0.0 & \cdots \mathbf{x} \in i \\ 1.0 & \cdots \mathbf{x} \notin i \end{cases}$$

としても大きく認識率に変化はなかった。入力に関しても一方の文字パターンの特徴値を負の値にして入力してみるなど、ニューラルネットワークへの入力と出力をいろいろと変えてみたが、特に大きく認識率に影響はなかった。これは学習により入出力に合わせてニューロン間の結合係数を適切に調整していくからであり入力と出力の値を変えても、学習ができれば認識率には特に影響しないからである。

また表 3 および表 5 より大分類でユークリッド距離を使用し、その文字候補から非教師信号を決めた場合と重み付けユークリッド距離から決めた場合とでは、1 字種あたり学習に使用する非教師信号の平均個数がほぼ同数であるにもかかわらず、重み付けユークリッド距離から非教師信号を決めた場合の方が未学習文字パターンに対する認識率が高い。また非教師信号の個数をただ増やしただけでは大きく認識率の向上には至らなかった。したがって、本論文で提案するニューラルネットワークの学習方法においては学習に必要な非教師信号の選び方が重要であることが分かる。すなわち教師信号の文字パターンと特徴空間上できわめて近い位置関係にある文字パターンの集合のみを非教師信号として見つけ学習を行い、教師信号と非教師信号の文字パターン間に適切な識別面を構築できれば、より多くの非教師信号のために文字パターンは必要ではないことが分かる。これは大分類で重み付けユークリッド距離を使用した場合の方が分類率が高い(表 2 よりユークリッド距離の場合、学習文字パターンに対する第 10 候補までの分類率が 99.36%である一方で、重み付けユークリッド距離の場合 99.83%であった)ため、より教師信号の文字パターンと類似している非教師信号の集合を選択でき、より認識率の向上が図れたことから明らかである。結局のところ非教師信号を決める過程は、特徴空間上での文字パターンの分布を調べることにほかならない。すなわち本論文で提案する手法は、まず特徴空間上で文字パターンがどのように分布しているのかを調べ、その分布状況に見合った識別関数をニューラルネットワークを用いて新たに作成するという 2 段階の過程をとっていることになる。そうしたことから考えても本手法では最初の特徴空間上における文字パターンの分布状況の調査過程に依存する部分が大きいことが分かる。

5. む す び

本論文では文字認識において基本的な手法であるテンプレートマッチング方式をニューラルネットワークを用いて実現することを試みた。テンプレートとして任意の文字種の標準文字パターンの特徴と認識したい未知文字パターンの特徴を入力層のニューロンに同時に入力したとき、この2つの文字種が同じ場合は出力層が発火し、異なる場合は発火しないような挙動をするニューラルネットワークを構築した。そしてこのようなニューラルネットワーク構築のための学習方法として、他のニューラルネットワークとの協調を考慮して、非教師信号の対象を自動的に選択しながら学習していく方法を提案した。そして文字データベースETL9B, 3036文字を用いた認識実験を行った結果、本手法では少ない個数の非教師信号を使用して効率良く認識率の向上を図れることを確認できた。

今後は使用した文字パターンの特徴、ニューラルネットワークの構造、学習アルゴリズムそして大分類との協調方法など個々の技術の改良を行い認識率の向上を図るとともに、さらに効率良く認識率を向上させるために学習時に必要な非教師信号の選択方法を改良していく予定である。

謝辞 本研究を行う際に、貴重な手書き文字データベースを提供していただきました通産省工業技術院電子技術総合研究所の皆様にご感謝いたします。

参 考 文 献

- 1) 羅, 四維, 岩根, 山崎: 人工神経網による手書き漢字の認識法, 情報処理学会論文誌, Vol.30, No.8, pp.962-969 (1989).
- 2) 朱, 小燕, 山内, 神保, 梅野: 階層的ニューラルネットワークを用いた手書き文字認識, 電子情報通信学会論文誌, Vol.J73-D-II, No.1, pp.54-61 (1990).
- 3) 岩田, 富麻, 松尾, 鈴村: 大規模4層ニューラルネット CombNET, 電子情報通信学会論文誌, Vol.J73-D-II, No.8, pp.1261-1267 (1990).
- 4) 堀田, 岩田, 松尾, 鈴村: 大規模ニューラルネット CombNET-II, 電子情報通信学会論文誌, Vol.J75-D-II, No.3, pp.545-553 (1992).
- 5) 猿田, 加藤, 安倍, 根元: 排他的学習ネット (ELNET) を用いた手書き文字認識の細分類手

法, 電子情報通信学会論文誌, Vol.J79-D-II, No.5, pp.851-859 (1996).

- 6) Saruta, Kato, Abe and Nemoto.: High Accuracy Recognition of ETL9B Using Exclusive Learning Neural Network-II (ELNET-II), *IEICE Trans.*, Vol.E79-D, No.5, pp.516-522 (1996).
- 7) 山田, 斉藤, 山本: 非線形正規化法の改良, 電子情報通信学会総合大会, D-439 (1988).
- 8) 鶴岡, 栗田, 原田, 木村, 三宅: 加重方向指数ヒストグラム法による手書き漢字ひらがな認識, 電子情報通信学会論文誌, Vol.J70-D, No.7, pp.1390-1387 (1987).
- 9) 大友, 大槻, 石谷, 原: 局所結合型神経回路網モデルによる手書き漢字の効率的認識法, 情報処理学会論文誌, Vol.35, No.6, pp.1091-1100 (1994).
- 10) 中野馨ほか: 入門と実習ニューロコンピュータ, 技術評論社 (1988).

(平成12年2月24日受付)

(平成12年11月2日採録)



篠沢 佳久 (正会員)

1994年慶応義塾大学理工学部管理工学科卒業。1996年同大学大学院理工学研究科管理工学専攻前期博士課程修了。1999年同大学大学院理工学研究科管理工学専攻後期博士課程修了。博士(工学)。現在同大学インフォメーションテクノロジーセンターに勤務。文字認識, パターン認識に興味を持つ。電子情報通信学会会員。



大駒 誠一 (正会員)

1936年生。1959年慶應義塾大学工学部卒業。慶應義塾大学理工学部管理工学科教授。工学博士。アルゴリズム, プログラミング言語, 日本語情報処理, 文字認識に興味を持つ。現在は, 日本の初期のコンピュータのソフトウェアの復刻に取り組む。著書「FORTRAN77」(サイエンス社)、「COBOLの基礎と応用」(サイエンス社)、「文科系のためのC」(サイエンス社)、「入門Cプログラミング」(培風館)等。日本ソフトウェア科学会, 計量国語学会, 日本公益学会, ACM各会員。