分散法による英数字の認識能力

正員 林 陽 一 (茨 城 大)
非会員 坂 田 正 輝 (日立エンジニアリング)
非会員 中 尾 隆 司 (日立エンジニアリング)
非会員 大 橋 新 悟 (茨 城 大)

Recognition Capability of Alphanumeric Characters by Distributed Method

Yoichi Hayashi, Member (Ibaraki University), Masateru Sakata, Non-member, Takashi Nakao, Non-member (Hitachi Engineering Co., Ltd.), Shingo Ohhashi, Non-member (Ibaraki University)

Various learning algorithms for neural networks have recently appeared and have stimulated applications of neural networks in new problem areas. This paper presents new results obtained by applying the distributed method proposed by Gallant to problems in alphanumeric character recognition. This method uses a distributed neural network with the Pocket Algorithm which is a modification of Perceptron Learning. The authors evaluate the recognition (classification) capability of the distributed method for 62 and 93 alphanumeric characters of a single font having different kinds of typeface quality and 76 alphanumeric characters of multiple fonts (12 types of fonts) having the same typeface quality. Furthermore, we propose a useful technique to distinguish between characters that closely resemble each other by using the structure information of characters. We also propose an activation criterion of output cells for character recognition.

In the recognition of alphanumeric characters having different kinds of typeface quality, a markedly high degree of recognition accuracy (99.96% maximum, 99.74% on average) of individual font quality was attained. In the 76 alphanumeric character recognition of the multiple font, a very high degree of recognition (99.64% maximum) was also achieved. The relations among the recognition rate and the number of items of training data, the number of intermediate cells, and the number of training iterations are considered in this paper.

キーワード：ニューラルネットワーク、文字認識、パーセプトロン、学習、パターン認識

1. はじめに

文字認識はパターン認識⁽¹²⁾の最も典型的な応用分野の一つであり、OCR、なつ印文字および刻印文字の認識などの産業応用で盛んに利用されている。現在までに、文字認識技術として複合類似度法⁽¹²⁾、構造解析法⁽¹²⁾などの方法が実用化されているが、人間の認識能力に比べると一部に不十分な面をもっている。近年、このギャップを埋める一つ的手段として人間の脳のニューラルネットワークの情報処理様式に学んだ新

しいタイプの並列分散情報処理が注目されており⁽¹⁾、ニューラルネットワークを用いた文字認識の研究が盛んに行われている。それらのなかでも、誤差逆伝搬法(バックプロパゲーション)⁽¹⁴⁾を用いて学習を行ったニューラルネットワークによる手書き数字⁽²⁾⁽¹⁵⁾、英数字⁽¹⁰⁾およびマルチフォント英数字の認識⁽¹⁵⁾がかなり高い性能を有することが報告されている。また、Fukushimaら⁽³⁾は視覚パターン認識能力と学習能力をもつ神経回路モデル“ネオコグニトロン”を用いた手書き数字認識システムを開発している。

一方, Gallant は, 3 層パーセプトロンの学習⁽¹³⁾を改良したポケットアルゴリズム⁽⁴⁾を提案しており, ポケットアルゴリズムを用いた判別結果は判別関数分析による結果と比較して, 学習データを誤って分類する割合が平均で約 22 % 少ないことを示す数値実験の結果を報告している⁽⁴⁾。更に, Gallant⁽⁵⁾ は, 3 層パーセプトロンのネットワークを変形した“分散ニューラルネットワーク”においてポケットアルゴリズムによる学習を行う“分散法(Distributed Method)”を提案し, Minsky ら⁽¹¹⁾が指摘した 3 層パーセプトロンの学習における欠点を大幅に改善できることを数値実験を用いて示した。

一般に, 印字した文字の品質が良好であれば, それらの文字の認識は容易である。しかし, 実用的な観点からみると明瞭に印字された品質の良い文字だけでなく, コピー, プリンタのリボンカートリッジの消耗などによって印字品質の劣化した文字も高い認識精度で認識しなければならない。また, マルチフォントの英数字の認識では, イタリック, スクリプトなどの特殊なフォントを含む様々なフォント内の 0 (数字の零), O (大文字), o (小文字) などの酷似した文字を認識する必要がある。従来のニューラルネットワークを用いた文字認識の研究では, 上記の観点に立った英数字の認識能力の評価と問題点の考察が十分に行われていなかったと思われる。

本論文では, 線形識別ネットワーク⁽⁶⁾の一種と考えられる分散ニューラルネットワーク⁽⁵⁾⁽⁷⁾に対してポケットアルゴリズム⁽⁴⁾を適用する分散法を用いて行った種々の英数字認識実験の結果を報告し, 分散法を用いたニューラルネットワークの認識能力の評価および問題点の考察を行う。実用的な観点に立つて行われた分散法を用いた文字認識能力の評価は, 著者らの口頭発表⁽⁹⁾以外公刊されていない。最近, Gallant は, 計算量に関する学習理論(Computational Learning Theory)⁽⁸⁾の枠組みにおいて, 分散法での学習を多項式時間で実現する方法を BRD(Bounded Randomized Distributed)アルゴリズム⁽⁷⁾として示している。そこでは, 文献(9)の結果を BRD アルゴリズムの特別な場合の一つの実例的な応用例として示している。

本論文では, ポケットアルゴリズムを用いて学習した分散ニューラルネットを, 異なる印字品質をもつ 62 文字と 93 文字の英数字の認識およびマルチフォントの 76 文字の英数字認識に用いた場合の性能評価を行う。特に, 学習データ数, 中間セル数および学習回数と認識率との関係について考察する。また, 文字の構造情報を用いて分散法をより効果的に働かせること

により, 非常に類似した文字を分類する簡便な方法を提案している。更に, 文字認識を行うための分散ニューラルネットワークの出力セルの活性度(出力値)に関する判定基準についても述べる。

異なる印字品質をもつ英数字の認識では, ドットプリンタにより出力した印字の濃いフォントおよび薄いフォント。更にそれらをコピーしたフォントの 4 品質のフォントを学習データとして用い, いずれのフォントの未学習データに対してもかなり高い認識率(最高 99.96 %, 平均 99.74 %)を得た。

更に, マルチフォントの英数字の認識では, IBM 社のタイプライタを用いて打ち出した 12 種類(イタリック, スクリプトなどの特殊なフォントを含む)のフォントを学習データとして用い, 未学習データに対してかなり高い認識結果(最高 99.64 %)を得た。

2. 分散ニューラルネットワーク

一般に, ニューラルネットワークは, セルと呼ばれる自律的な処理ユニット(頂点)とコネクション(アーク)と呼ばれるセルどうしを結合する数値的な重みのついた辺から構成される。ニューラルネットワークの形式には各種のものがある⁽¹⁾が, 本論文では図 1 に示す分散ニューラルネットワークを用いる。入力層, 中間層および出力層の各層に含まれるセルをそれぞれ入力セル, 中間セル, 出力セルと呼ぶ。入力層は一つの文字を表現するために必要なセル数, 出力層は判別したい文字数のセル数からなる。各セル間のコネクションはセル j におけるセル i の影響に関する整数の重み w_{ji} をもつ。正の重みは強化を, 負の重みは抑制を表す。

分散ニューラルネットワーク⁽⁵⁾⁽⁷⁾の構造は以下の三

図 1 分散ニューラルネットワーク
Fig. 1. A distributed neural network.

つの点で3層パーセプトロンと異なっている。

(1) 入力セルと中間セルがコネクションを介して有向完全2部グラフ $K_{p,q}$ を構成している。

(2) 入力層から中間層へのコネクションの重み w_{ji} は適当な範囲の乱数 ($|w_{ji}| \leq L$; L は任意の正の整数であり、本論文の実験では $L=5$ としている) を用いて生成され固定されている。

(3) 入力層から出力層へ学習可能なコネクションをもつ。

(1)項により、入力パターンの一つの特徴は、一つの間層セルの活性度によって符号化(局所表現)されるのではなく、すべての中間セルの活性度パターンによって符号化(分散表現)される。従って、少数の中間セルを用いて十分効率良く入力パターンの特徴が保持される。

上記に加え、(2)項を採用することにより、中間層が入力パターンを汎化する能力も高められる。以上から、類似性の高い入力パターン E と E' に関しては各々の中間セルの活性度が等しくなる可能性が高くなる。このため、出力セルへの E と E' に対応する入力も類似性が高くなり、学習がより容易になる。

(3)項は出力層へのより分離性の高い入力を実現している。3層パーセプトロンにおいても、原理的には、中間セルの活性度パターンを用いて、入力パターンは分離可能になり得る。しかし、実際の文字認識などでは、入力パターンの次元が高く、多数の認識対象が存在するため、入力パターンの分離は困難である。そこで、入力層からのコネクションを加え、出力層への入力をより高い次元空間(p 次元から $p+q$ 次元)に拡張することにより分離性を高めている。

入力セル $I_i (i=1, \dots, p)$ の活性度(出力値)はネットワークに呈示された文字画像のパターンによって $+1$ または -1 となる。中間セル $H_j (j=1, \dots, q)$ と出力セル $O_k (k=1, \dots, r)$ の活性度は、それぞれ(1a)~(1d)式によって求められる。従って、中間セルおよび出力セルの活性度は、 $-1, 0, +1$ の3値のいずれかの値をとる。ここでは簡単のため、セルの名称とその活性度を同じ記号で表す。

$$SH_j = \sum_{i=1}^p w_{ji} I_i \dots\dots\dots (1a)$$

$$H_j = \begin{cases} +1 & (SH_j > 0) \\ 0 & (SH_j = 0) \\ -1 & (SH_j < 0) \end{cases} \dots\dots\dots (1b)$$

$$SO_k = \sum_{i=1}^p u_{ki} I_i + \sum_{j=1}^q v_{kj} H_j \dots\dots\dots (1c)$$

$$O_k = \begin{cases} +1 & (SO_k > 0) \\ 0 & (SO_k = 0) \\ -1 & (SO_k < 0) \end{cases} \dots\dots\dots (1d)$$

重み u および v は任意の整数値をとり得る。セル間にコネクションがないときは、その重みを0とする。また、セル I_0 の活性度は $+1$ であり、入力セルを除く他のすべてのセルとの間にコネクションをもつ。これらのコネクションの重みは一般にバイアスと呼ばれる。

3. 分散法による英数字の認識

分散法では、分散ニューラルネットワークにおける出力セルのコネクションの最適な重みを学習によって生成するために、ポケットアルゴリズム⁽⁴⁾を用いる。以下にその概要を示す。

〈3・1〉 ポケットアルゴリズム ポケットアルゴリズムは3層パーセプトロンの学習を改良したものであり、パーセプトロンによる学習で計算された重みがそれまでの中で最良のものであれば、“ポケット”と呼ばれる重みベクトルに一時的に保管し、また計算を続ける方法である。各出力セルに関する学習は独立しているため、ここでは簡単のために、一つの出力セルについて学習アルゴリズムを述べる。そのため、出力セルの添字 k を除き、出力セル O へのコネクションの重みベクトル π について記述する。

出力セル O について、 $\{E^n\} (n=1, \dots, N; N$ は学習データの数) を学習データの活性度の集合とし、 $\{C^n\}$ をそれに対応する出力セル O についての正しい活性度の集合とする。ここで、 C^n は $\{\text{True}, \text{False}\}$ に対応する値 $\{+1, -1\}$ を、 E^n の各要素は $\{\text{True}, \text{False}, \text{Unknown}\}$ に対応する値 $\{+1, -1, 0\}$ をとり得る。 $E_0^n \equiv +1$ とすると π_0 は出力セル O のバイアスとなる。

出力セル O に関する最適な重みベクトル π を計算するためのアルゴリズムは以下ようになる。

〔ステップI〕 P を零ベクトルとする (P は π と同次元のベクトルとする)。

〔ステップII〕 P を現在のパーセプトロンの重みとする。ランダムに一つの学習データ E^n を、対応する正しい出力セル O の活性度 C^n と共に選ぶ。

〔ステップIII〕 重み P と選ばれた学習データ E^n の内積を計算する。

$$IP = P \cdot E^n$$

〔ステップIV〕

(1) 重み P が正しく E^n を分類する、すなわち

$$\begin{cases} \{IP > 0 \text{ and } C^n = +1\} \text{ または} \\ \{IP < 0 \text{ and } C^n = -1\} \end{cases}$$

ならば、[ステップIV](1-a)を実行し、それ以外は[ステップIV](2)へ行く。

(1-a) 重み P が連続して正しく E^n を分類した回数が、ポケット内の重みベクトル π が連続して正しく E^n を分類した回数より多いならば、ポケット内の重み π を P で置き換え、その回数を覚えておく。

(2) 重み P が正しく E^n を分類しなければ、新しい重み P' を以下のように生成し、連続して正しく分類した回数を 0 にもどす。

$$P' = P + C^n E^n$$

[ステップV] [ステップII]へ行く。

ポケットアルゴリズムにおける学習回数は、各出力セルに関して[ステップII]から[ステップV]までを繰返す回数として定義される。Gallant⁽⁴⁾ は学習回数が増加するにつれて、ポケット内の重みベクトル π が最適である確率が 1 に近づくことをポケットアルゴリズムの収束定理として証明している。

〈3・2〉 文字認識における出力セルの活性度の判定基準 未学習データの認識の際、(i)複数個の出力セルの活性度が+1となる場合および、(ii)すべての出力セルの活性度が-1となる場合がある。本論文では上記の場合に対して、(1d')式を用いて呈示された未学習データに最も似ていると考えられる文字を各文字に対応している r 個の出力セルの中から一つだけ選んでいる。

$$O_k = \begin{cases} +1 & SO_k = \max_{k \in \{1, 2, \dots, r\}} (SO_k) \\ -1 & SO_k \neq \max_{k \in \{1, 2, \dots, r\}} (SO_k) \end{cases} \dots\dots\dots (1d')$$

雑音の多いデータあるいは不完全な情報を用いる文字認識においては、上記(i)の場合に対応する混読(優劣のつけにくい候補文字が複数個存在する)の発生が多いと考えられる。このような場合、(1d')式は有効な判断基準となると思われる。

一方、誤読に伴う損失が非常に大きいと考えられる認識対象に対しては、上記(ii)の文字は読取不能と判定し、保留として最終決定を後処理にゆだねるのが妥当である。この際、 SO_k の値は有益な評価値として利用することができる。更に、損失などの程度によっては上記(i)の文字も読取不能と判定する場合が考えられる。

4. 異なる印字品質をもつ英数字の認識

本章では、分散法を用いた異なる印字品質をもつ英数字の認識能力の評価および問題点の考察を行う。

〈4・1〉 異なる印字品質をもつ英数字 認識実験に用いた文字は、24ドットのドットインパクトプリンタを用いて濃く(Dark)印刷したフォント(以下、Dフォントと呼ぶ)、薄く(Light)印刷したフォント(Lフォント)およびそれらをコピーしたCDフォント(Copied-Dark)、CLフォント(Copied-Light)の4種類である。ここで、フォントとは認識対象となる文字種類数を基本単位とした文字セットのことである。これらのフォントは、300 dpiのイメージスキャナにより画像データとして計算機内に読込まれる。読込まれた画像データに対し、横方向および縦方向の黒画素のヒストグラム(頻度分布)をとり、文字画像を切出している。イメージスキャナより取込んだ文字画像には文字線の切れ、かすれなどが生じるため、修復処理を行った後、文字の位置および大きさに関する正規化と最近傍法による濃度の正規化を行い、図2に示すような二値の16×16画素のフォントを作成した。4種類のフォントには、図2に示すような“つぶれ”、“切れ”などの文字画像の品質の差が見られる。

本論文では、1フォントが62文字の英数字、および93文字の英数字および記号から構成される2種類のフォントを対象に認識実験を行った。93文字のフォントにおける31個の記号文字は以下のとおりである。

! # \$ % & ' () - =

図2 四種類の印字品質をもつ“s”の二値画像

Fig. 2. Binarized image of four kinds of typeface quality “s”

図 3 文字位置データの例
Fig. 3. An example of character position data.

— ¥ | @ [] ' { }
: * ; + / _ ? > < ,

62 文字の認識実験での学習データは文字の画素パターン情報だけである。しかし、93 文字の認識実験においては“—”や“_”など正規化すると全く同じ画素パターンになってしまう文字もあるので、文字の画素パターンに加えて、イメージスキャナから文字画像を切出した際に得られる文字位置データを付加情報として与えている。文字位置データを与えることにより、学習データに関する矛盾を避けることができる。文字が図 3 に示すように切出されたとき“*”を例にとると、点 A を原点(0, 0)として点 C と点 D の Y 座標が文字位置データとなる。本論文ではこれらの値を 0 から 7 (点 B の Y 座標を 7 とする) の 8 値の位置情報として用いる。文字位置データを表現するには 8×2 個の入力セルが必要となる。従って、本章の実験に用いる学習データの 1 文字あたりのデータ量は 32 または 34 バイトである。

入力セルの数 (p の値) は、(i) 認識対象となる文字の印字品質 (単位長さあたりのドット数, 文字画素の構成など) と形状の特徴 (縦長, 正方形など), (ii) イメージスキャナの性能 (解像度など) およびイメージスキャナから切出された文字画像の画素数および, (iii) 許容されるニューラルネットワークを表現する情報量 (コネクションの数など) を考慮して決められる。また、中間セルの数 (q の値) の決定に関しては、少ない学習データあるいは学習回数を用いて予備実験を行い、中間セルの数についてある程度の目安をつけた後、実験を行うのが現実的である。

4.2 異なる印字品質をもつ英数字の認識実験

本章の実験では、62 文字認識実験、93 文字認識実験各々について、学習データ数および中間セル数と、認識率の関係を考察し、分散法を用いた英数字の認識能力を評価する。また、62 文字の実験においては、学習回数と認識率の関係も考察する。本章の実験で用いるネットワークの入力セル数は、62 文字認識実験

図 4 62 文字認識における学習データ数に対する認識率の変化

Fig. 4. Percent correct recognitions versus number of items of training data for 62 characters.

図 5 93 文字認識における学習データ数に対する認識率の変化

Fig. 5. Percent correct recognitions versus number of training data for 93 characters.

では 256 (16×16) 個, 93 文字認識実験では 272 ($16 \times 16 + 16$) 個である。

(1) 学習データ数による認識率の変化 学習データ数が分散法を用いた文字認識能力に与える影響について検討するために、学習データ数を変化させて、(i) 学習データ認識実験 [学習用データとテスト (評価) 用データが同じ] と、(ii) 未学習データ認識実験 (学習用データとテスト用データが異なる) を行った。一般には、学習データの認識率と未学習データの認識率の差は学習後のネットワークが学習データの性質に

依存しているほど大きく、中間セル数が十分多く、学習データを非常に多く用いれば両者は同一値に近づく傾向をもつと考えられる。

学習データ数は 62 文字認識実験、93 文字認識実験両者共、各々、各品質のフォントから 9 から 20 フォントとし、未学習データ認識実験における評価用データ数はどの実験でも各品質のフォントから 25 フォントとした。各出力セルに関する学習回数は学習データのフォント数（文字セット数）を 1,000 倍した回数、中間セル数は入力セル数の半分とした。図 4、図 5 に学習データと未学習データに対する実験結果を示す。

学習データの認識率は、学習データ数に関係なく、62 文字実験ではすべて 100 %、93 文字実験でもほとんど 100 % である。これは、ランダムに生成されたコネクションによって学習データの重要な特徴が適切に中間セルの活性度として分散表現されたことを示している。

一方、未学習データの認識では、学習データ数を増加させるにつれて両実験共に認識率が向上している。学習データが 80 フォントのとき、62 文字実験では 4 品質のフォントの認識率の平均は 99.63 % であった。また、93 文字実験では CL フォントの認識率は 99.96 % と非常に高く、4 品質のフォントの認識率の平均は 99.74 % であった。正規化によって全く同じ画素パターンあるいは酷似した画素パターンになる文字も含めて未学習データを高い精度で認識しており、文字位置データが少ない情報量で効果的に働いていると考えられる。

図 6 62 文字認識における学習回数に対する認識率の変化

Fig. 6. Percent correct recognitions versus number of training iterations for 62 characters.

実験結果からみると、L フォントのような印字品質の悪い文字に対しては、学習データを増やすことによる認識率の改善が特に大きい。得られたデータの都合上、最高 80 フォントまでしか実験を行っていないが、更に多くのデータを学習させれば、認識率が上がる余地は十分あると思われる。

(2) 学習回数による認識率の変化 学習回数と認識率との関係を調べるために、62 文字実験において各出力セルに関する学習回数を 3 万から 8 万回まで変化させ認識率の変化を調べた。学習データ数は、(1) 項の実験において認識率が最も高かった 80 フォント、中間セル数は入力セル数の半分の 128 個とした。学習回数を上記のように変化させたときの認識結果を図 6 に示す。

学習データについては、学習回数が多いほど認識率が高くなっている。これはポケットアルゴリズムでは、学習回数が多いほどコネクションの重みを最適化できることを示している。今回の実験では、学習回数が 4 万回のときの認識率は 100 % であり、それ以降もほぼ 100 % となって収束していることがわかる。

また、未学習データについても同様の傾向がみられ、4 万回の学習でほぼ収束し、4 品質のフォントの認識率は平均で 99.61 % であった。フォント別にみると、D フォントおよび CD フォントの認識率は学習回数が 3 万回のときには既にそれぞれ 99.87 %、99.80 % に収束している。これに対して、認識率の低い L フォントおよび CL フォントの認識率は、学習データと同じく学習回数が 4 万回のときに収束し始めている。学習回数が 3 万回の時点では L フォントおよび CL フォントについて十分な学習ができていなかったと考えられる。本実験では、学習回数が学習データのフォント数の 500 倍（学習した文字数の約 8 倍）で学習データについて認識率は飽和しており、この時点でネットワークの動作はほぼ最適化されたと思われる。

(3) 中間セル数による認識率の変化 中間セル数の認識率への影響を考察するために、中間セル数を変化させて認識率の変化を調べる。中間セル数は、62 文字実験では 110 から 170 個、93 文字実験では 120 から 150 個とする。また、学習データ数は各 80 フォント、学習回数は 8 万回とした。(2) 項の実験と同様、学習データの認識率は常に 100 % であったので、未学習データ認識実験について考察する。中間セル数と認識率の関係を図 7 および図 8 に示す。

62 文字実験および 93 文字実験での認識率は共に最高 99.96 % であり非常に高い。62 文字実験の L フォ

ントの認識率は中間セルの数が150個のとき最高になり、以後低下している。一方、他のフォントの認識率の変化に顕著な特徴はみられない。これはLフォントの特徴が他の3種類のフォントの特徴と大きく異なっており、多数の中間セルを用いた分散表現が適切に働かないためと思われる。

93文字実験では、どの品質のフォントも同じよう

図7 62文字認識における中間セル数に対する認識率の変化

Fig. 7. Percent correct recognitions versus number of intermediate cells for 62 characters.

図8 93文字認識における中間セル数に対する認識率の変化

Fig. 8. Percent correct recognitions versus number of intermediate cells for 93 characters.

な認識率の傾向が見られる。これは、入力セルに与えられた文字位置の情報がフォントの品質に関係なく個々の文字にほぼ共通しているため、フォントの品質の差による認識率への影響がある程度小さくなっていると考えられる。本実験の結果から、分散法を用いた文字認識では、中間セル数は入力セル数の半数程度でよいことがわかる。

5. マルチフォント英数字の認識

本章ではマルチフォントの英数字および記号を対象にして、分散法を用いた文字の認識能力の評価および問題点の考察を行う。

〈5・1〉 マルチフォントの英数字および記号 実験に用いた12種類の英数字および記号のフォントは、図9に示すように同一文字であっても形状の変化が大きい。1フォントは76文字の英数字および記号から構成されている。1フォントあたり76文字としたのは、12種類のフォントすべてに共通する文字を選んだためである。また、数字の“1”は英小文字の“l”と同じ字体を用いているフォントがほとんどであったので、認識対象から除いた。英大文字、英小文字および数字の61文字以外の15文字は以下のとおりである。

% & ' () =
@ : ; + / ? , "

一つの文字は8×11画素で表現される。フォントの作成方法は、第4章と同様である。入力データは4章の93文字の認識実験と同様、文字の画素パターンだけでなく、文字位置データを付加情報として与えている。従って、本章の実験に用いる学習データの1文字あたりのデータ量は13バイトである。

〈5・2〉 マルチフォント英数字の認識実験 本章

SCRIBE	A O a o 7 0 & @
ADVOCATE	A O a o 7 0 & @
BOOKFACE ACADEMIC72	A O a o 7 0 & @
COURIER72	A O a o 7 0 & @
DELEGATE	A O a o 7 0 & @
ADJUTANT	A O a o 7 0 & @
LARGE ELITE72	A O a o 7 0 & @
PRESTIGE ELITE72	A O a o 7 0 & @
COURIER12	A O a o 7 0 & @
COURIER12 ITALIC	A O a o 7 0 & @
SCRIPT	A O a o 7 0 & @
LIGHT ITALIC	A O a o 7 0 & @

図9 12種類のフォントの例

Fig. 9. Some examples of 12 types of fonts.

図 10 マルチフォントの 76 文字認識における学習データ数に対する認識率の変化
Fig.10. Percent correct recognitions versus number of items of training data for multiple-font 76 characters.

の実験では、学習データ数、学習回数および中間セル数と、認識率との関係を考察し、分散法を用いたマルチフォント英数字に対する認識能力を評価する。本章の実験で用いるネットワークの入力セル数は、 $104(8 \times 11 + 16)$ 個である。

(1) 学習データ数による認識率の変化 学習データ数が未学習データの認識率に与える影響について検討するために、学習データの認識実験と未学習データの認識実験を行った。学習データ数は、各フォントより 10 から 25 フォントとし、未学習データ認識実験における評価データ数は各フォントから 20 フォントとした。学習回数は学習データのフォント数を 1,000 倍した回数、中間セルの数は 70 個とした。図 10 に学習データ数と認識率の関係を示す。

第 4 章の実験では、入力データが 1 種類のフォントであったため直接的に矛盾するデータが存在しなかったが、今回は認識対象がマルチフォントであるため BOOKFACE ACADEMIC 72 の“0” (数字の零) と PRESTIGE ELITE 72 の“O” (英大文字) が酷似しているなど、他の種類のフォントとの関係で矛盾するデータが多く存在する。そのため、学習データ認識実験においても認識率 100% は得られず、学習データが 180 (15 × 12) フォントのときの 99.83% が最高であった。

一方、未学習データの認識率は学習データ数を増加させるにつれて向上しており、学習データ数が合計 300 フォントのとき 99.64% であった。このことは、

図 11 マルチフォントの 76 文字認識における学習回数に対する認識率の変化
Fig.11. Percent correct recognitions versus number of training iterations for multiple-font 76 characters.

分散法を用いたネットワークは多くの学習データを用いることにより、印字品質の差だけでなく、各種の文字形状の変化をもつ学習データの重要な特徴を適切に分散表現することが可能であることを示している。

(2) 学習回数による認識率の変化 学習回数の認識率への影響を調べるために、学習回数を 10 万から 30 万回へと変化させた。学習データ数は 240 フォント、中間セルの数は 70 個とした。学習回数と認識率の関係を図 11 に示す。

学習データの認識では、学習回数が増加するにつれて認識率は向上しているが、学習回数が 20 万回るときはほぼ収束している。

また、未学習データの認識についても同様の傾向が見られるが、10 万回の学習での認識率が 99.55% で、それ以降はほとんど変化していない。マルチフォントの認識実験では矛盾するデータが存在するため、学習データについて最適となったネットワークが必ずしも未学習データに対しても最適になるとは限らない。実験結果を見ると、10 万回学習した時点で既にネットワークの認識能力はほぼ飽和していたと考えられる。学習回数が 20 万回るときと 30 万回ときの認識率の差はわずかに 0.01% であるので、今後の実験では学習回数を 20 万回とする。今までの実験の結果から、学習回数は学習文字数の 10 倍程度で十分であると考えられる。

(3) 中間セルの数による認識率の変化 中間セル数の認識率への影響を考察するために、中間セル数を変化させて認識率の変化を調べる。中間セル数は、30 から 80 個とする。また、学習データ数は 240 フォント、学習回数は 20 万回とした。中間セル数と認識率の関係を図 12 に示す。学習データの認識率は中間

図 12 マルチフォントの 76 文字認識における中間セル数に対する認識率の変化
Fig. 12. Percent correct recognitions versus number of intermediate cells for multiple-font 76 characters.

セル数を増加させるにつれて向上しており、中間セル数が 80 個のときに最高の 99.83 % を示している。

また、未学習データの認識では中間セル数が 60 個のとき、認識率が 99.58 % となり、それ以後は大きな変化はない。中間セル数は入力セル数の半分よりやや多い程度で十分であると考えられる。

6. 終 り に

分散法は 3 層パーセプトロンに対して、ネットワークの計算メカニズムをその構造の変更によって強化し、更に学習メカニズムもポケットアルゴリズムによって改良した方法と考えられる。一般に、3 層パーセプトロンには汎化の能力はなく、認識対象となるパターンの正規化や特徴抽出が必要である。本論文で提案した文字位置に関する情報は後者を簡便に実現する方法の一つである。

分散法を用いた英数字の認識において以下のことが言える。

(i) ポケットアルゴリズムの各出力セルに関する学習回数は学習文字数の 10 倍程度で十分である。

(ii) 中間セル数は入力セル数の半分よりやや多い程度で十分である。

(iii) 文字位置の情報を付加することにより、正規化によって全く同じあるいは酷似した画素パターンになる文字も判別可能になる。

近年、その有効性が広く認められている誤差逆伝搬法と分散法を比較した場合以下のことが考えられる。

(1) 学習速度 ポケットアルゴリズムは、線形の整数演算だけで実行できる。また、分散ニューラルネットワークのコネクションの重みを表現するのに要

する情報量は誤差逆伝播法の 40 % 程度 (本論文の実験の場合) である。更に、分散ニューラルネットワークの入力層に与えられた入力データを表現する情報量は、誤差逆伝播法の情報量の 8 分の 1 程度で十分な場合が多い。このため、分散法は浮動小数点演算を伴う誤差逆伝播法に比べ、原理的には高速であると考えられる。

分散法による 62 文字の英数字認識における学習および認識の速度はそれぞれ、930 万コネクション/s、180 文字/s (CPU time)、また、マルチフォントの英数字認識では、880 万コネクション/s、510 文字/s であった。これらは、東京大学大型計算機センターの Hitac M-682 H で上記の実験を行ったときの測定結果である。

(2) 学習時のパラメータ設定 ポケットアルゴリズムを用いて学習を行う場合、学習効率および認識率に影響するパラメータは主に学習回数と中間セル数であるが、誤差逆伝播法では更に、学習係数 η 、安定化係数 α などの値も影響を与える。これらのパラメータの最適値を求める一般的な規則は、現在のところ確立されていない。

一方、ポケットアルゴリズムは、誤差逆伝播法における二乗誤差に対応するコネクションの重みの最適性を示す指標がない。そのため、学習回数の上限が理論的にわからないことが欠点である。

ポケットアルゴリズムは容易に並列化ができるため、スーパーコンピュータ、ニューラルネットチップ、データフロープロセッサなどを用いて学習速度を飛躍的に向上させることが可能である。

(平成元年 7 月 31 日受付, 同元年 10 月 23 日再受付)

文 献

- (1) 合原: 「総論 ニューラルコンピューティングの展望」, 電学誌, 109, 427 (平成-6)
- (2) D. J. Burr: "A Neural Network Digit Recognizer", Proc. of the 1986 IEEE Int. Conf. on Systems, Man and Cybernetics, Vol. 2, p. 1621 (1986)
- (3) K. Fukushima, S. Miyake & T. Ito: "Neocognitron: A Neural Network Model for a Mechanism of Visual Pattern Recognition", IEEE Trans. Syst. Man & Cybernetics, 13, 826 (1983)
- (4) S. I. Gallant: "Optimal Linear Discriminants", Proc. of the 8th Int. Conf. on Pattern Recognition, p. 849 (1986)
- (5) S. I. Gallant & D. Smith: "Random Cells: An Idea Whose Time Has Come and Gone... And Come Again?", Proc. of IEEE Int. Conf. on Neural Networks, Vol. 2, p. 671 (1987)
- (6) S. I. Gallant: "Connectionist Expert System", Communications ACM, 31, 152 (1988)
- (7) S. I. Gallant: "A Connectionist Learning Algorithm with Provable Generalization and Scaling Bounds", TR NU-CCS-89-13, College of Computer Science, Northeastern University (also to be published in Neural Networks, Vol. 3) (1989)

- (8) D. Haussler & L. Pitt (eds.): "COLT '88, Proc. of the 1988 Workshop on Computational Learning Theory" (1988) Morgan Kaufmann
- (9) Y. Hayashi, M. Sakata, T. Nakao, T. Ohno & S. Ohhashi: "Alphanumeric Character Recognition Using A Connectionist Model with the Pocket Algorithm", Proc. of Int. Joint Conf. on Neural Networks, Vol. 2, p. 606 (1989)
- (10) 香田・高木・鈴木:「ニューラルネットによる手書き英数字認識 —モデル規模及び学習データ数に関する評価検討—」, 信学技報, PRU88-57, 71 (昭63)
- (11) M. Minsky & S. Papert: "Perceptron: An Introduction to Computational Geometry", MIT Press (1969)
- (12) 森: パターン認識 (昭63) 信学会
- (13) F. Rosenblatt: Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms (1961) Spartan, Washington D. C.
- (14) D. E. Rumelhart, G. E. Hinton & R. J. Williams: "Learning Representations by Back-Propagating Errors", *Nature*, 323, 533 (1986)
- (15) 山田・上・溝口・天満:「ニューラルネットを用いた文字認識」, 信学技報, PRU88-58, 79 (昭63)

林 陽 一 (正員)

昭和32年1月1日生。59年3月東京理科大学大学院理工学研究科経営工学専攻博士課程修了。工学博士。61年4月茨城大学工学部情報工学科専任講師, 平成2年4月同助教授, 現在に至る。ファジィ推論, ニューラルネットワーク, エキスパートシステムおよびパターン認識の研究に従事。電子情報通信学会, 情報処理学会, 日本ファジィ学会, 人工知能学会, IEEE, ACM, 国際神経回路網学会(INNS)会員。

坂 田 正 輝 (非会員)

昭和23年7月5日生。51年東北大学大学院工学研究科応用物理学専攻修士課程修了。同年4月日立エンジニアリング(株)入社。59年アメリカ・プリンストン大学計算機科学科修士課程修了(MSE)。現在同社システム開発センタ主任技師。情報処理学会, 人工知能学会, 電子情報通信学会, IEEE, ACM および INNS 会員。

中 尾 隆 司 (非会員)

昭和31年5月22日生。57年3月広島大学理学部物性学科卒業。同年4月日立エンジニアリング(株)入社。現在同社システム開発センタ技師。平成元年8月より, アメリカ・アラバマ州立大学バーミングハム校計算機・情報科学科修士課程に在学中。情報処理学会会員。

大 橋 新 悟 (非会員)

昭和41年1月5日生。平成2年3月茨城大学大学院工学研究科情報工学専攻修士課程修了。同年4月日本電気(株)入社, 現在に至る。