

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よる評価検討

高木, 英行
松下電器産業株式会社中央研究所

香田, 敏行
松下電器産業株式会社中央研究所

久木, 泰治
松下電器産業株式会社中央研究所

<https://hdl.handle.net/2324/4479726>

出版情報：1988. 電子情報通信学会
バージョン：
権利関係：Copyright(C)2020 IEICE



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—

Recognition of Hand-written Alphanumeric Characters by Artificial Neural Network

香田 敏行 高木 英行 木 泰治

Toshiyuki KODA Hideyuki TAKAGI Yasuharu SHIMEKI

松下電器産業 中央研究所

Central Research Laboratories Matsushita Electric Industrial Co., Ltd.

We've applied a artificial neural network (NN) to recognition of thirty six hand-written alphanumeric characters. Two points are especially considered in these experiments. (1) relationship between recognition accuracy and connection number of NN model: it has been found that connection between NN layers can be reduced. (2) relationship between recognition accuracy and number of training data for NN model: recognition accuracy is saturated when more than forty persons data are used for training of NN.

(1) はじめに

最近、神経回路網（ニューラルネット）を用いた情報処理方式をさまざまな応用分野に適用しようという研究が盛んになってきている。特に、文字認識の分野では逆伝搬法（バックプロパゲーション学習則）⁽¹⁾で学習させたニューラルネットを手書き数字、手書き漢字に適用し、従来手法に劣らない認識率が得られている⁽²⁾⁽³⁾。一方、学習により形成される隠れユニットの動作を解析することによって、入力データの構造を反映した内部表現についての洞察を得ようとする試みも多い。実際、手書き漢字認識において、中間層に文字の空間情報を識別するユニットが形成されるという報告がある⁽⁴⁾。

しかしながら、このようなニューラルネットの文字認識への適用は、入出力ベクトルの次元数が増えると、急速に計算量が増加するという規模依存性の悪さから、数字、類似漢字の識別など、問題の一部分に試験的に用いているにとどまっている。一方、音声認識の分野では、上位層のユニットに下位層の近傍ユニットとのみ接続関係をもたせることによって、局所的な特

徴の組合せから全体的な特徴を抽出するようなニューラルネットを構成し、連続音声の中のホルマント抽出に適用した結果、その有効性が確認されている⁽⁴⁾。

本稿では、中間層の1つのユニットが接続される入力層のユニットの範囲を限定したニューラルネットを手書き英数字認識に適用し、このモデルを適用することによる、計算量の削減能力及び文字の部分空間情報の有効利用の可能性について検討した結果を報告する。

また、ニューラルネットを手書き文字認識に適用する際、当然、手書き文字認識を行うネットワークにはできるだけ多くのデータを学習させる方がよい。しかし、余り多過ぎても性能は飽和状態となり学習効率が悪い。このような観点から、何人程度のデータをネットワークに学習させるのが適当であるかについても評価を行った。

(2) ニューラルネットの構成と逆伝搬法

2. 1 モデルの構成

ニューラルネットは、入力層、中間層、出力層からなり各層内での結合はなく、信号は上層に向かって一方向に伝達されるフィードフォワード型のネットワークである。各ユニットは接続されている下層のユニットの出力とその接続の度合である重みを掛け合わせたものの総和を入力とし、その入力をしきい値関数で変換した後、その値を上層への出力として伝達する働きをしている。今回用いたしきい値関数は(1)式に示すsigmoid関数で、この関数は微分可能な非線形単調増加関数である。

$$y_j = 1 / (1 + e^{-x_j}) \quad (1)$$

$$x_j = \sum y_i w_{ji} \quad (2)$$

y_j : ユニットjの出力

x_j : ユニットjの入力総和

y_i : ユニットiの出力

w_{ji} : ユニットiからjへの結合重み

2. 2 逆伝搬法

図1に示した階層ネットワークを学習させるには、入力層の各ユニットに入力データを与え、出力層から出力された値を望ましい値(教師データ)と比較し、その差を減らすように結合の強さを変えればよい。ここでは、学習アルゴリズムとしてRumelhartら⁽¹⁾の提案した逆伝搬法を用いた。この学習アルゴリズムについて簡単に述べる。

評価関数

階層ネットワークにある入力データを与えたときに出力層から出力される値 y_j と、教師データとして外部から与えられる望ましい出力の値 t_j から(3)式のような関数を定義し、この関数の値で現在の結合状態(重み係数の大きさ)でのネットワークの性能を評価する。

$$E = \frac{1}{2} \sum (y_j - t_j)^2 \quad (3)$$

逆伝搬法

ネットワークを学習させるには、この2乗誤差総和Eを減少させるように(4)式にしたがって結合の強さを変えればよい。

$$\Delta w_{ji} = -\epsilon \partial E / \partial w_{ji} \quad (4)$$

ϵ : 学習レート

この(4)式を変形すれば Δw_{ji} は(5)式の様に定式化できる。

$$\Delta w_{ji} = -\epsilon \delta_j y_i \quad (5)$$

y_i : ユニットiからユニットjへの入力値

δ_j は、ユニットjが出力ユニットか中間ユニットかで異なる。

出力ユニットの場合は

$$\delta_j = (y_j - t_j) \partial y_j / \partial x_j \quad (7)$$

中間ユニットの場合は

$$\delta_j = \partial y_j / \partial x_j \sum \delta_k w_{kj} \quad (8)$$

である。 k は、ユニットjが接続している上層ユニットを表しているので、 Δw はその前段の Δw が決まらないと計算できない。このように、学習は入力データの処理とは逆方向、つまり後向きに進む。

逆伝搬法による学習では、これを収束するまで繰り返すが、収束を速める目的で(9)式に示すように、結合強度を変更する際に前回の変更量が影響するように変更するのが一般的である。

$$\Delta w_{ji}(n) = -\epsilon \partial E / \partial w_{ji} + \alpha \Delta w_{ji}(n-1) \quad (9)$$

n : 学習回数

α : 加速度レート

(3) 英数字認識実験

3. 1 実験概要

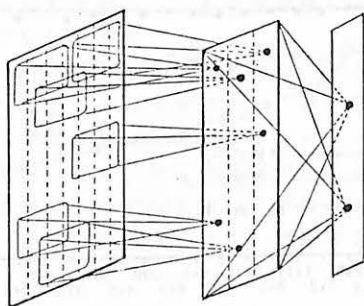
手書き漢字認識において、中間層に文字の空間情報を識別するユニットが形成されることを考慮すると、中間層の各ユニットと全ての入力層を結合するよりも、2次元的に結合範囲を限定する方が、文字の空間情報を中間ユニットの内部表現に効率よく反映でき、しかも結合数が減少することから計算量の削減にもなると考え

られる。そこで、今回の実験では、中間層の各ユニットが接続される入力層の範囲（以下、ウィンドの大きさ）を2次元的に変えることが可能なニューラルネットを手書き英数字認識に適用し、結合範囲の大きさが認識率に与える影響及び、計算量削減の可能性について検討した。

また、ニューラルネットを手書き文字認識に適用する際、学習に必要なデータ数（筆者数）の検討を行った。当然、手書き文字認識を行うネットワークにはできるだけ多くのデータを学習させる方がよい。しかし、余り多過ぎても性能は飽和状態となり学習効率が悪い。そこで、ネットワークの学習に用いるデータ数と認識性能との関係についても検討を行った。

3. 1. 1 モデルの構成

今回認識実験に用いたモデルは、図1に示すように入力層、中間層、出力層からなり、出力層の各ユニットは全ての中間層のユニットと接続されているが、中間層の各ユニットと入力層との接続については結合範囲を2次元的に変えることが可能な階層的ネットワークである。



入力層 中間層 出力層
(8×11個) (4×5～7×8個) (36個)

図1 階層的神経回路網

各層のユニット数は、以下のように設定した。

入力層: 88 (8×11) 個

中間層: 20 ～ 56 個

出力層: 36 個 (数 字 10 文字、
アルファベット大文字 26 文字)

中間層の各ユニットと入力層のユニットの結合

範囲は、3×3～8×11の16通りについて検討した。

3. 1. 2 入力パターン

今回用いた手書き英数字データは、図2に示すように、見本文字を参考に所定の枠内に数字10文字(1～10)、大文字のアルファベット26文字(A～Z)合わせて36文字を記入してもらったもので、80人分のデータを用意した。



図2 手書き文字サンプル

手書き文字データは64×63の二値マトリックスで構成されている。このマトリックスデータに対して、位置変動の吸収と入力次元の削減を目的として、文字部分の切り出しを行い、正規化した後メッシュ特徴を求め、それらをネットワークの入力層に与えた。

メッシュ特徴は、正規化後のマトリックスを、88 (8×11) 個の粗いメッシュ領域に分割し、各メッシュ領域に含まれる文字部分をつまみ黒点部分の全体に対する割合を計数することにより求めた。

3. 1. 3 学習パラメータと学習方法

図1に示した階層ネットワークを学習させるには、(4)式に従って結合の強さを変えればよい。学習レートεの値は学習データに依存す

るので簡単には決められないが、ここでは試行錯誤的に決定し0.2とした。今回の実験では、重み変更時に前回の変更量が影響するように(9)式を用い、加速度レート α は0.95とした。

また、重み変更はパターン提示時に行わずにそのときモデルに与えたサンプル全てを提示した後で一度に変更を行う一括重み変更方式も提案されているが⁽³⁾、ここではパターン提示毎に重みを変更する方法を採用した。

3. 1. 4 中間層の大きさと認識率

図3に中間層のユニット数を20、30、42、56個と変えた場合の認識結果を示す。ここでは、ウインドの大きさを 8×11 とし、100回学習後の学習データ(40人分)及び未学習データ(40人分)に対する認識結果を示す。

学習データ、未学習データ共に中間ユニットの数が多いほど認識率が高くなり、特に未学習データの方が改善効果は大きい。ただどちらの場合も中間ユニット数が30個ぐらいから飽和傾向にあり、認識率からみて中間ユニット数は56個程度で十分であると考えられる。そこで、以下の実験では中間ユニット数が56個の場合について検討する。

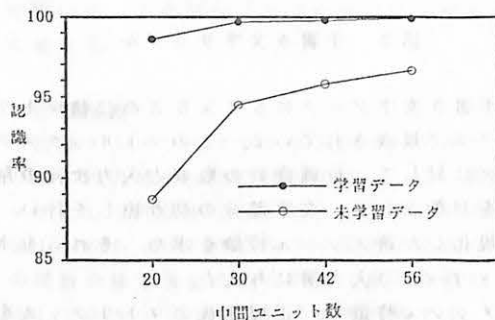


図3 中間ユニット数と認識率

3. 2 モデル規模の検討

3. 2. 1 ウインドの大きさと認識率

図4-1、図4-2にウインドの大きさを $3 \times 3 \sim 8 \times 11$ の16通りに変化させたときのネットワークの結合数の変化と学習データ及び

未学習データに対する認識率の変化を示す。ここでは、40人分のデータを学習データとし、残りの40人を未学習データとして使用した。

結合数については、ウインドの大きさが 8×11 の場合(中間層の各ユニットと全ての入力ユニットを結合した場合)を1としてそれとの比(結合数比)で表している。図4-1は100回学習後、図4-2は200回学習後のネットワークについて実験した結果である。

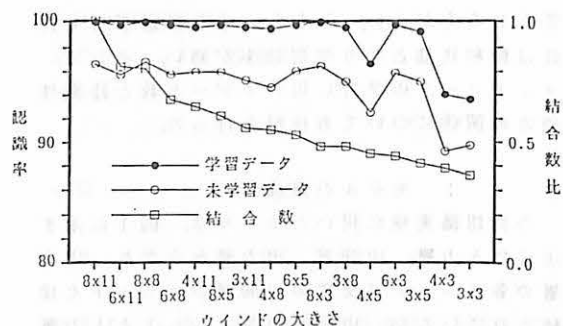


図4-1 結合数と認識率(100回学習)

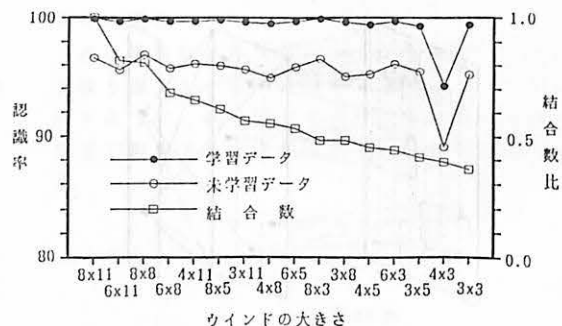


図4-2 結合数と認識率(200回学習)

学習データ、未学習データともに、200回学習のネットワークに比べると100回学習の方は、ウインドの大きさによる認識率のばらつきが目立つものの、どちらの場合もウインドの大きさのある程度小さくしても、それほど認識率の劣化はみられない。特に、未学習データに対するウインドの大きさが 8×3 、 6×3 の場合などウインドの取り方によっては、結合数が半分以下のモデル規模でもかなり高い認識率を

示している。

3. 2. 2 ウィンドの大きさと学習速度

ウィンドの大きさが学習速度に与える影響を調べるために、ネットワークの学習性能を表す誤差総和 E ((3) 式) が一定値以下になるまでの学習速度とウィンドの大きさとの関係を求めた。図 5-1、図 5-2 にその結果を示す。

図 5-1 は誤差総和 E が 0.1 以下、図 5-2 は誤差総和が 0.2 以下の場合である。

ただし、ウィンドの大きさによって 1 回の学習に要する時間が違うので、学習速度がネットワークの結合数に比例すると仮定して、誤差総和 E が一定値以下になるまでの学習回数に前述した結合数比を掛けることにより補正し学習速度を求めた。ここでは、ウィンドの大きさが 8×11 の場合 (中間層の各ユニットと全ての入力ユニットを結合した場合) の学習速度を 1 としてそれとの比 (学習速度比) で表している。

ウィンドの大きさが 4×8 、 4×5 の場合など一部学習速度に劣化が生じているが、これは誤差総和 E が局所極小値に陥込みそこから抜け出すのに時間を要しているためで、全体的にみると、結合数を減らしても学習速度にはほとんど影響がないと思われる。

ここでは評価の都合上誤差総和 E がしきい値以下になるまで学習を続けたが、最適学習回数の決定は非常に困難であることから、ネットワークを学習させるときは誤差総和 E の収束が確認できるまで十分余裕を持って学習させる場合が多い。この点を考慮すると、認識率に劣化が

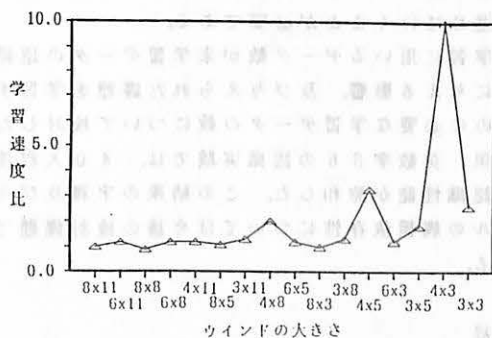


図 5-1 ウィンドの大きさと学習速度
(誤差総和 $E \leq 0.1$)

ないのであれば、むしろ中間層と入力層の間の結合数を削減し、一回の学習に要する時間を少なくした方が、学習効率の点、またハードウェア実現の点においても有利であると思われる。

3. 3 学習データ数の検討

学習に用いるデータ数が未学習データの認識率に与える影響、及び与えられた課題を学習するのに必要な学習データの数について検討するために学習データに関してクローズの実験 (学習データと評価データが同じ) とオープンの実験 (学習データと評価データが異なる) を行った。

クローズ実験とオープン実験を比較すると、前者の方が認識率が高いのが普通であり、両者の差は、学習後のネットワークのデータ依存性が高いほど開き、学習データ数を非常に多く用いれば両者は一致すると考えられる。そこで、学習データ数を順次増やして、オープンとクローズの評価を行い、認識率の傾向を調べ、学習データ数と、ネットワークの学習データ依存性の関係について検討した。

オープンの評価実験法⁽⁵⁾

80人分の学習データを機械的に10人ずつ8クラスに分割し、そのうち1クラスを評価用とし、他の7クラスを用いて10人~70人用の学習データをそれぞれ作成し、評価する。そしてデータによる結果の片寄りを除くため、評価用クラスと学習データ作成用クラスを4通りに入れて評価し、4回の平均値を求めた。

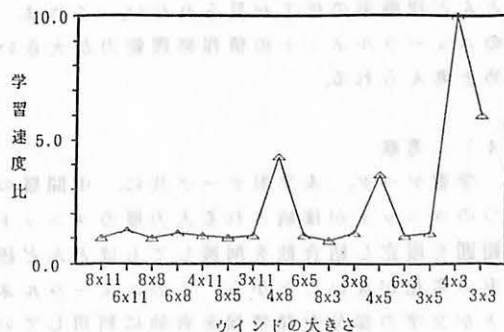


図 5-2 ウィンドの大きさと学習速度
(誤差総和 $E \leq 0.2$)

クローズの評価実験法

オープンの評価実験と同様に、データを8クラスに分割し、学習に用いたデータをそのまま評価用データとした。データによる結果の片寄りを除くため、同一データ数の評価に対するデータを4通りに替えて評価し、4回の平均値を求めた。

図6に、オープン及びクローズの評価実験結果を示す。この場合もウインドの大きさは 8×11 、学習回数は100回とした。

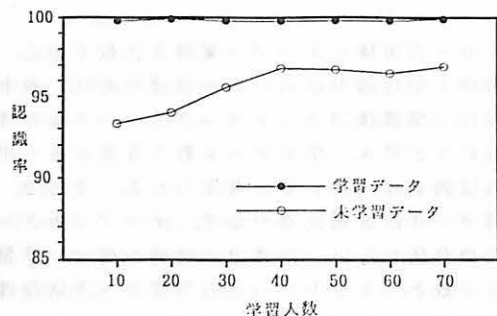


図6 学習データ数と認識率

オープン評価では、学習に用いるデータ数を増加させるにつれて認識率が向上し、40人を境に飽和している。これはデータの増加と共にネットワークの汎用性が増すためと思われる。

一方、クローズ評価は一般的にいて学習データ数が増加するとネットワーク内に入力データ構造を反映しきれなくなることから、認識率が低下するはずであるが今回の実験においてはほとんど認識率の低下が見られない。これは、このニューラルネットの情報処理能力が大きいためと考えられる。

〔4〕 考察

学習データ、未学習データ共に、中間層の1つのユニットが接続される入力層のユニットの範囲を限定し結合数を削減してもほとんど認識率の劣化がないことから、このニューラルネットが文字の部分空間情報を有効に利用していることを示唆していると考えられる。

また、認識率を劣化させずにネットワークの結合数が削減できることから計算量の削減が可

能となり、認識時間の短縮につながるだけでなく、一回の学習に要する時間が短くて済み学習効率が向上するといった点についてもある程度確認できた。これは、これまでニューラルネットを実用的な文字認識システムに適用する際問題となっていた規模依存性の悪さという点に対してこのモデルが有効である可能性を示している。

学習データ数の実験については、40人程度で認識率が飽和状態になるという実験結果を得たが、認識対象文字が36英数字であることと今回のモデル規模とに依存すると考えるのが妥当である。いずれにせよ、不特定筆者用の文字認識システムを設計する上での設計指針が与えられたことには違いない。

〔5〕 まとめ

本稿では、中間層の各ユニットが結合する入力層の範囲を限定したニューラルネットを手書き英数字認識に適用することによる、計算量の削減能力と部分空間情報の有効利用の可能性について検討した。その結果、中間層と入力層の接続範囲を限定し結合数を削減してもほとんど認識率の劣化はなく学習データ数で99%以上、未学習データで96%前後の高認識率を得た。

文字の部分空間情報の有効利用という点からすると、このモデルは英数字よりもいろいろな部首からなる漢字に適用する方がより有効と思われる。今後手書き漢字に適用し、文字の空間情報が中間層の内部表現にどの様に反映されているか、またどのぐらいのウインドの大きさが適当であるかなどについて、さらに詳細な検討を進めて行くことが必要である。

学習に用いるデータ数が未学習データの認識率に与える影響、及び与えられた課題を学習するのに必要な学習データの数について検討した結果、英数字36の認識実験では、40人程度で認識性能が飽和した。この結果の字種及びモデルの規模依存性については今後の検討課題である。

謝辞

日頃御指導頂く、新田所長ならびに松島室長に感謝致します。特に貴重な助言を頂いた坪香主任研究員をはじめ、日頃御討論頂く情報科学

研究室の諸氏に感謝致します。また、この認識実験を行うにあたり、貴重な文字データを御提供頂いた情報システム研究所に感謝致します。

参考文献

- 1) D.E.Rumelhart et al., "Learning Representation by Back-Propagation Errors," Nature, Vol.323, pp533-536, Oct.1986
- 2) D.J.Burr, "A Neural Network Digit Recognizer," IEEE-SMC, pp1621-1625, 1986
- 3) 森、横澤、梅田、"PDPモデルによる手書き漢字認識", 信学技報, MBE87-156, Vol.87, No.428, pp407-414, 1988年3月
- 4) 高木、坪香、"ニューラルネットを用いた疑似ホルマント抽出フィルタリング", 音講論 3-P-11, pp249-250, 1988年3月
- 5) 二矢田、藤井、森井、"不特定話者を対象とした音声認識のための特徴パラメータと距離尺度に関する考察", 信学論(A), Vol.J69-A, No.5, pp628-638, 1986