

複合型ニューラルネットワークによる文字認識能力の検討
Recognition ability of multiple structured neural networks

藤木 なほみ[†]、大久保武尊[†]、藤木澄義[‡]

Nahomi M. Fujiki, Takeru Ohkubo and Sumiyoshi Fujiki

1. はじめに

ニューラルネットワークの中で、最も一般的に用いられている学習アルゴリズムは、1986年にRumelhartによって提唱された階層型ニューラルネットワークのエラーバックプロパゲーション学習則であろう。この方法は、出力層からの出力値と教師信号の差を二乗した自乗誤差を評価関数とし、理想的な出力が得られるようニューラルネットワークに学習をさせるが、学習が準安定状態に陥りやすいことや最適なネットワークサイズを事前に知っておかなくてはならないという問題点も挙げられる。この問題を解消するエラー測度として、ある確率を持った入力パターンに対して出力層に目的とする出力パターンが出力される条件付確率と、目的の出力パターンが教師信号として与えられる条件付確率の比からなる、カルバック測度が提唱された[1]。

我々は、このカルバック測度を評価関数として階層型ニューラルネットワークにおけるエラーバックプロパゲーション学習則を導出し、手書き文字データベース“ETL-3”を用いた文字認識問題を通して、ニューラルネットワークの学習や認識の能力について検討してきた。[2,3,4,5,6]。

このカルバック測度をエラー測度とする学習則の特徴は、ネットワークのサイズに依存することなく常に安定した学習傾向を示すスケラビリティに優れた点にあり、この特徴を活かしたスケルフリーの高効率な認識システムが構築されることが期待される。本研究では、この特長を活かし簡単に学習済みのネットワークを構築できることから、個々に学習を進めた複数のネットワークから構成される複合型ニューラルネットワークを考えることで、単体では十分引き出せなかったその能力を容易に高めることができることを示し、手書き数字の認識問題を通してその具体的な能力を検討した結果を報告する。

2. カルバック測度を用いた学習則

本研究で用いたネットワークの学習則は以下になる。図1に示すような3層の階層型ネットワークにおいて、各ニューロンは±1の離散値を取る確率的ネットワークを考える。各層間のニューロンは互いにシナプス結合し、信号は入力層から出力層に向かって1方向のみに送られる。

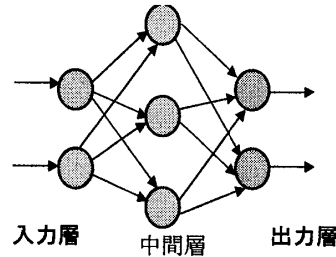


図1 3層構造ネットワーク

今、一般に l 層の i 番目のニューロンを $\sigma_i^{(l)}$ と表すとするとき、 $l+1$ 層の j 番目のニューロンがとる状態の確率は、 l 層の総てのニューロンからの入力によって式(1)で表わされる条件付確率で与えられる。

$$P(\{\sigma_j^{(l+1)}\}|\{\sigma_i^{(l)}\}) = \prod_j \frac{\exp[\beta \sigma_j^{(l+1)} h_j^{(l+1)}]}{2 \cos \beta h_j^{(l+1)}} \quad (1)$$

ここで、

$$h_j^{(l+1)} = \sum_i w_{ji}^{(l+1)} \sigma_i^{(l)} + \theta_j^{(l+1)} \quad (2)$$

は、ニューロンの内部状態を、 β 、 $w_{ji}^{(l+1)}$ 、 $\theta_j^{(l+1)}$ は、それぞれ温度係数、ニューロン間のシナプス結合の強さ、ニューロンの持つ閾値を示す。このようにして、入力層(第0層)のニューロンの状態に対応して出力層(第 L 層)のニューロンがとる状態確率は、各層毎の条件付確率の積として式(3)で与えられる。

(3)

このネットワークにおいて、最小化するべき量としてカルバック

$$P(\{\sigma_j^{(L)}\}|\{\sigma_i^{(0)}\}) = \sum_{\sigma^{(0)} \dots \sigma^{(L-1)}} \prod_k \frac{\exp[\beta \sigma_k^{(1)} h_k^{(1)}]}{2 \cos \beta h_k^{(1)}} \dots \prod_j \frac{\exp[\beta \sigma_j^{(L)} h_j^{(L)}]}{2 \cos \beta h_j^{(L)}}$$

測度は次式で与えられる。

$$S = - \sum_{\mu} p(\mu) \sum_{\nu} Q(\{\sigma_j^{(L)}\}_{\nu}|\{\sigma_i^{(0)}\}_{\mu}) \ln \frac{P(\{\sigma_j^{(L)}\}_{\nu}|\{\sigma_i^{(0)}\}_{\mu})}{Q(\{\sigma_j^{(L)}\}_{\nu}|\{\sigma_i^{(0)}\}_{\mu})} \quad (4)$$

ここで、 $p(\mu)$ は入力パターン μ が呈示される確率、 $P(\{\sigma_j^{(L)}\}_{\nu}|\{\sigma_i^{(0)}\}_{\mu})$ は入力パターン μ が与えられた時に、出力層に目的とする出力パターン ν が出力される条件付確率、 $Q(\{\sigma_j^{(L)}\}_{\nu}|\{\sigma_i^{(0)}\}_{\mu})$ は目的とする出力パターン ν が教師信号として与えられる条件付確率である。

[†]仙台電波工業高等専門学校専攻科

[‡]東北文化学園大学科学技術学部

この確率的ネットワークに平均場近似を施すことで、決定論的アナログニューロンからなる階層型ネットワークの学習則が得られる。平均場近似の結果、 $l+1$ 層のニューロンの平均値 $m_j^{(l+1)}$ は前層の各ニューロンの平均値 $m_i^{(l)}$ によって決定される。

$$m_j^{(l+1)} = \tanh(\beta h_j^{(l+1)})$$

$$h_j^{(l+1)} = \sum_i^{M^{(l)}} w_{ji}^{(l+1)} m_i^{(l)} + \theta_j^{(l+1)} \quad (5)$$

また、前述のカルバック測度は、以下の式に置き換わる。

$$S_{mfa} = - \sum_{\mu} p(\mu) \frac{1}{2} \sum_j \sum_{\sigma_j^{(L)} = \pm 1} (1 + \sigma_j^{(L)} \tau_{j\mu}^{(L)}) \ln \frac{(1 + \sigma_j^{(L)} m_{j\mu}^{(L)})}{(1 + \sigma_j^{(L)} \tau_{j\mu}^{(L)})} \quad (6)$$

ここで $\tau_j^{(L)}$ は教師信号、 $m_j^{(L)}$ は出力層の出力値を表す。式(6)で与えられた平均場近似されたカルバック測度を式(7)、(8)に示すように各シナプス結合と各ニューロンの閾値についての変分をとることで、カルバック測度を最小にするようシナプス結合、閾値の更新量を決定する修正式が導出される。この修正式は、従来の自乗誤差をエラー測度とした場合の学習則とよく似た修正式であるが、この修正式にしたがった学習は、ネットワークサイズに関わらず常に安定した学習傾向を持つことが、我々のこれまでの一連の研究において示されている[2, 3, 4]。

$$\Delta w_{ji}^{(l)} = -\alpha \frac{\delta S_{mfa}}{\delta w_{ji}^{(l)}} \quad (7)$$

$$\Delta \theta_j^{(l)} = -\alpha \frac{\delta S_{mfa}}{\delta \theta_j^{(l)}} \quad (8)$$

ここで、 α は学習係数とよばれる任意の定数である。

3. 実験概要

本研究では、ニューロンの数をそれぞれ入力層 1140 個、中間層 K 個 (512 個~8192 個)、出力層 10 個とした 3 層の階層型ニューラルネットワークにおいて、カルバック測度を評価関数とした学習則に基づいて学習を行った。結合荷重と閾値の初期値は、[-1.0~+1.0]区間の一様乱数で設定し、学習係数 α は 0.1、温度係数 β は 1.0 とした。各ニューラルネットワークの学習は、その出力層からの出力と教師信号との自乗誤差を学習終了条件として、様々な場合において、その認識能力と汎化能力について検討を行った。

本研究で用いた手書き文字は、産業技術総合研究所より提供されている文字データベースのうち、手書き文字 200 人分のデータを収録してある“ETL-3” [7]を用いた。

ETL-3 は一文字あたり 72×78[pixel]の 16 階調グレイスケールで収録されているが、この文字情報を閾値 12 で 2 値化変換を施し、さらに正規化しデータ情報を含まない周囲を削除することで 30×38[pixel]に縮小した。このように正規化処理を行った“ETL-3”に収録されている 200 人分の“0”から“9”の手書き数字データを、100 人分を学習用に、残りの 100 人分を汎化能力を見る認識用とし

た。

以上のような設定環境のもと、各ネットワークサイズについて初期状態を違えた複数のネットワークに同じ手順によって 100 人分の手書き数字の学習を行わせ、それらの学習効率及び未学習データに対する認識の様子を考察した。また、学習終了条件として目的とする出力値と実際の出力値との自乗誤差を用い、終了条件を変える事で学習の習熟度の違う様々なネットワークを用意し検討を行った。

ネットワークの初期状態をランダムに設定するため、同じデータ

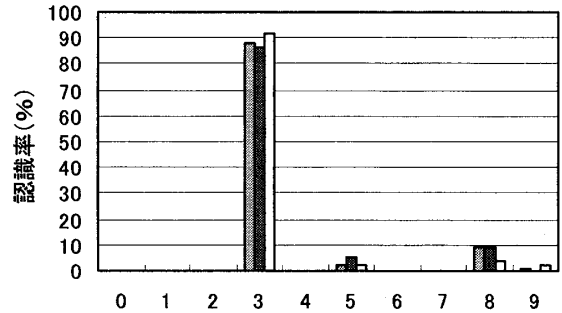


図2. 100 人分の手書き数字“3”の認識の様子

を学習させてもネットワーク毎に多少の認識の偏りが生じる。特に、この偏りは、学習終了条件を緩めた場合に顕著に現れる。複数のネットワークを用いてその能力を測る場合、結果を単純平均するか、多数決法によって認識率を判断するのが一般的である。しかし、図2に例として 100 人分の未学習文字“3”に対する認識の様子を示したように、学習済ネットワークが複数の類似パターンに同時に反応するのは自然なことである。図2は中間層ニューロン数を 8192、学習終了条件を 0.01 未満としたときの 3 つの学習済ネットワークが、未学習文字“3”を 90% 近くは“3”として、10% 程度は“8”、数%は“5”として認識したことを示している。我々は、このような他の文字との類似性に関する情報も重要な判断材料となると考え、本研究では図3の模式図に示すように初期状態を変えて学習させた複数の学習済ネットワークを並列に結合し、各文字に対応す

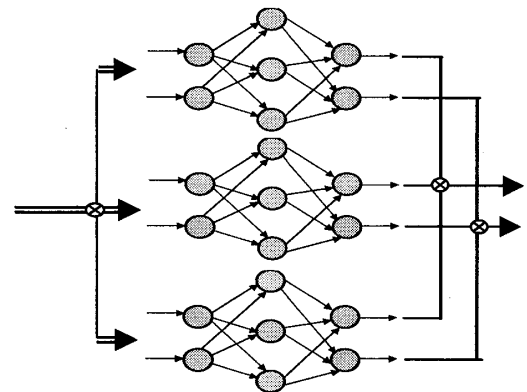


図3 複合型ネットワークの模式図

る出力ニューロンの値の総和の平均値 (以後、複合平均と呼ぶ) か

ら認識の判定を行う、複合型の認識用ネットワークシステムを考案した。複数のネットワークの出力の総和をとることでこのネットワークが、どのパターンに最も類似していると判断したかをより公平に評価できることが期待される。学習の習熟度の違いやネットワークサイズの違いによる認識能力の変化及びそれらのネットワークを複合する数によって未学習データ認識率にどのような違いが現れるかを詳細に検討した。

4. 実験結果

4.1 学習と識別能力

今回の研究で、ネットワークのサイズを中間層ニューロンを512個から8192個まで変化させた場合の手書き文字の学習曲線の様子を図3に示した。中間層ニューロン数が増えることにより学習回数が減少し、いずれの場合も学習曲線は滑らかに収束していく様子が分かる。また、表1には、中間層ニューロン数を1024個、学習終了条件を自乗誤差が0.1未満及び0.01未満になるまで手書き文字の学習を行わせた後に、学習データがどの程度正しく認識されているかの識別率を示した。自乗誤差0.01未満でほぼ100%の識別率であった。

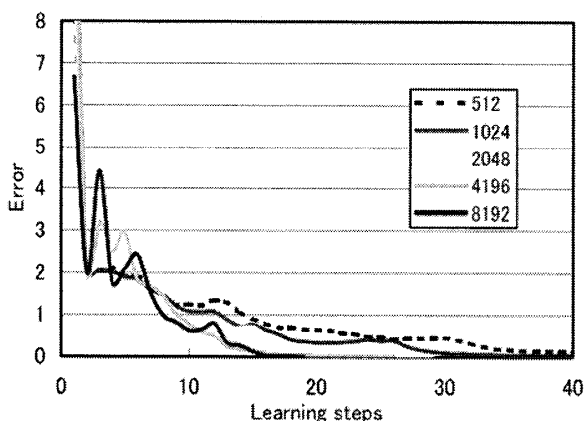


図4 ネットワークのサイズ(中間層のニューロン数)の違いによる学習の様子の違い

表1 正規化を行った学習済み手書き文字の平均識別率 (%)

中間層ニューロン (個)	512	1024	2048
終了条件 0.1 未満	94.4	95.8	96.9
終了条件 0.01 未満	99.2	99.4	99.5

4.2 未学習データに対する認識能力

4.1において学習を行わせた様々な学習済みのネットワークに未学習100人分の手書き文字データを読み込ませ、各文字がどの数字として認識されたかを検討した。図4に中間層ニューロン数1024個とし、自乗誤差が0.01未満になるまで学習を行わせること

を条件としたときの未学習手書き文字の文字別認識率を示した。学習済ネットワークの初期値に依存した差は多少認められるが、3つの単体の学習済ネットワークによる認識率は56~97%であり、他の文字と特徴が似通った文字の認識率は低くなる傾向が求められた。

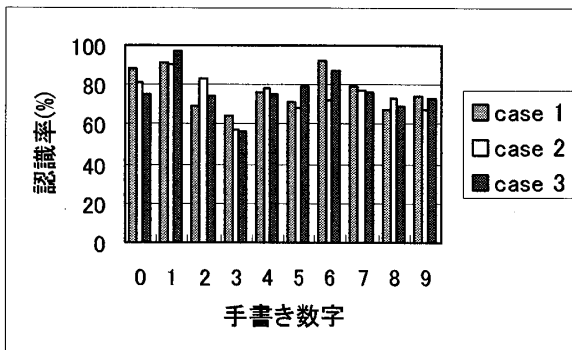


図5 中間層ニューロン1024個、終了条件0.01未満での未学習手書き文字認識率

また、表2には中間層のニューロン数を増加させネットワークサイズを変えて、認識率を検討した結果を示した。表よりサイズが大きくなるにしたがって認識率もかなり向上していくことが判る。また、表に示す3つの学習済ネットワークの認識率の単純平均と、3つのネットワークの各出力ニューロン毎の出力値の総和から認識判定を行う複合平均による認識率の算出結果を比較すると、複数の学習済ネットワークを複合して用いることで、格段に認識率が向上することが判る。また、この向上の度合いは、比較的小さなネットワークにおいてその効果が顕著であり、サイズ512では7%強向上したのに対して8192では4%弱に留まっているが、複合平均をとることで90%を超える認識率を得られることが判った。

表2 3つのネットワークの出力値の単純平均と複合平均による中間層ニューロン数別の未学習文字平均認識率(%)

サイズ	512	1024	2048	4096	8192
複合平均	85.5	87.3	89.9	91.9	92.8
単純平均	77.63	79.07	83.2	87.33	89.07

さらに、我々は学習終了条件を変化させることにより、学習の習熟度に依存して認識率がどのように変化するかを検討を行った。複合型ネットワークとした場合の効果がより顕著となることが期待される比較的小さいサイズである中間層ニューロン数512及び1024個について、終了条件を非常に緩めた場合(終了条件0.5未満)から少し厳しくした場合(終了条件0.01未満)へと変化させて学習を行わせた後の認識率を表3、4に示した。各表には、比較のために認識判定に単純平均を用いた場合と複合平均を用いた場合の認識率の違いを考察した結果も示した。複合平均をとった場合の認識率は、学習終了条件を緩めた場合においても、かなり高い認識率となっていることが判る。

表3. 中間層ニューロン512個のネットワークにおける平均認識率(%)の、複合するネットワーク数(セット数)及び学習の習熟度(終了条件の自乗誤差値)依存性

セット数	0.5未満まで学習		0.1未満まで学習		0.01未満まで学習	
	単純平均	複合平均	単純平均	複合平均	単純平均	複合平均
1	71.9	71.9	75.5	75.5	74.1	74.1
3	70.7	85.8	75.6	89.3	76.1	89.4
5	70.8	89.9	75.7	91.7	76.4	92.8
10	70.3	91.8	75.1	93.5	76.5	94.5
25	70.5	94.5	75.2	94.4	76.2	95.2

表4. 中間層ニューロン1024個のネットワークにおける平均認識率(%)の、複合するネットワーク数(セット数)及び学習の習熟度(終了条件の自乗誤差値)依存性

セット数	0.5未満まで学習		0.1未満まで学習		0.01未満まで学習	
	単純平均	複合平均	単純平均	複合平均	単純平均	複合平均
1	74.2	74.2	77.9	77.9	77.5	77.5
3	74.7	88.9	77.7	88.9	77.6	89.7
5	75.0	91.4	78.1	92.2	78.3	92.8
10	75.1	93.7	78.0	93.9	78.8	94.8
25	74.9	95.0	78.0	95.9	78.9	95.7

5. まとめ

カルバック測度を評価関数としたエラーバックプロパゲーション学習則を用いた場合、ネットワークサイズに左右されずに安定した学習傾向を示すというスケラビリティに優れた性質を持つことを活かし、優れた手書き文字の学習・認識能力を持つ複合型階層型ネットワークシステムを構築し、その能力の詳細な検討を行った。

カルバック測度をエラー測度とした場合の大きな特徴として、単体のネットワークでは中間層ニューロンの数が増加に伴って学習及び認識能力が向上したが、十分な認識率を得るためには、かなり大きなサイズが必要であった。しかし、今回の研究で、複数のネットワークからの出力値をより公平に考慮できるよう複合することで、小さなネットワークサイズであっても高い能力を持たせることができることが明らかにされた。特に、学習の習熟度が十分でない状態における複合型ネットワークの効果は顕著であり、適当な数のネットワークを複合することで、95%程度の認識率が容易に得られることが判明したことは意義が大きいと考える。

佐々木・平井の研究[8]ではETL-1のデータを用い、位置と大きさの正規化を行いエラーバックプロパゲーションを用いて文字の学習と認識を行った結果、未学習手書き文字の認識率が95.2%得ら

れ、さらに方位選択性を持った特徴抽出を行うことで98.6%の認識率を得られることが報告されている。直接比較することはできないが、本研究で用いたネットワークにおいても適当な特徴抽出を付加することで、未学習文字認識率をさらに向上できる可能性が示唆される。

参考文献

- [1] B. S. Wittner and J. S. Denker: Strategies for Teaching Layered Networks Classification Tasks, American Ins. of Physics 52, pp.850-859 (1988) 他
- [2] Fujiki and Fujiki: A learning process of a stochastic feed forward neural network, J. Phys. Soc. Japan 64, pp.757-765 (1995)
- [3] 藤木なほみ、及川康志: エラーバックプロパゲーション学習則におけるエラー測度の選択と学習効率向上に関する考察, 仙台電波工業高等専門学校研究紀要 第29号, pp.47-52 (1999)
- [4] 藤木なほみ、馬場庸二、原田祐介: カルバック測度を用いたエラーバックプロパゲーション学習則の学習及び認識能力の評価, 仙台電波工業高等専門学校研究紀要 第33号, pp.65-70 (2003)
- [5] 阿部, 藤木: カルバック測度をエラー測度とするニューラルネットワークの文字認識と汎化能力の検討, 仙台電波工業高等専門学校研究紀要 第34号, pp.55-60 (2004)
- [6] 藤木なほみ、阿部容子、藤木澄義: カルバック測度をエラー測度とするニューラルネットワークの文字認識を通じた能力の検討, H18年度電気関係学会東北支部連合大会予稿集, pp.254 (2006)
- [7] ETL CHARACTER DATABASE Electrotechnical Laboratory [ETL-3], (1993)
- [8] 佐々木信、平井有三: DOG フィルターによる線分特徴抽出を用いた手書き数字認識系の構成, 電子情報通信学会 信学技法 NC97-37, pp.17-23 (1997)