階層型神経回路網のデイジタル構成におけるビット数低減方法

メタデータ	言語: Japanese
	出版者:
	公開日: 2017-10-03
	キーワード (Ja):
	キーワード (En):
	作成者: 中山, 謙二, 猪股, 悟, 竹内, 由幸
	メールアドレス:
	所属:
URL	http://hdl.handle.net/2297/5633

論　文

階層型神経回路網のディジタル構成におけるビット数低減方法

正　員　中山　謙二[†]　　准　員　猪股　悟[†]　　学生員　竹内　由幸[†]

Reductions in Number of Bits for Digital Realization of Multilayer Neural Networks

Kenji NAKAYAMA [†], *Member*, Satoru INOMATA [†], *Associate Member* *and* Yukou TAKEUCHI [†], *Student Member*

あらまし　神経回路網をディジタル回路で実現する方法は結合係数の可変性や学習アルゴリズムに対する柔軟性の点から有望な手段の一つである．しかし，神経回路網は非常に多くの非線形素子とその間の結合を必要とするため，LSIの集積度を上げるためには，効率の良い回路構成法が重要となる．本論文では，誤差逆伝搬アルゴリズムによる階層型神経回路網を用いたパターン認識を対象とし，ユニット数，結合係数とユニットの出力を表現するビット数，および結合密度とパターン認識能力の関係について検討する．以下に研究成果を要約する．結合係数を2ビットまで，また隠れ層と出力層のユニット出力を0，1の2値まで低減しても学習が収束することが確認できた．また，雑音に対して安定な認識を行うための学習方法としては，ビット数低減を学習中に段階的に行うこと，およびユニット出力の2値化のためにシグモイド関数の1次微分を微小な正定数で置換する方法が有効である．更に，ユニット数および結合密度を減らすことは雑音に対する認識能力の点から有効ではない．0～9の数字とアルファベット52文字を用いたシミュレーションにより以上のことを実験的に確認した．

1. まえがき

近年，非線形ユニットを超並列に結合した人工的な神経回路網の研究が活発になっている．パターン認識などの分野では階層型モデルを用い，学習アルゴリズムとして，誤差逆伝搬アルゴリズムが効果を上げている[(1)~(4)]．

神経回路網を実現する手段とし，ディジタル回路を用いる方法は，結合係数の可変性や学習アルゴリズムに対する柔軟性の点から有望な手段の一つである．しかし，神経回路網では，多くのユニットとユニット間結合を必要とし，またユニットの入出力変換が非線形であるため，現在のLSI技術ではごく小規模なものしか実現できない[(5)~(7)]．ディジタルLSIにおける神経回路網の集積度を上げるためには，機能を劣化させることなくハードウェアを削減する必要がある．ハードウェアを減らすための方策として，ユニット数の削減，結合係数とユニットの入出力を表現するビット数の削減，結合密度の低減，およびユニットの入出力変換関数の簡単化などが考えられる．

これまでにも，画像圧縮などの符号化への応用として階層型神経回路網の隠れ層のユニットの低減，およびその出力の量子化とビット数低減の可能性が検討されてきた[(1),(8)]．しかし，ディジタル型神経回路網全体のハードウェアを低減するための検討はまだ十分ではない．

本論文では誤差逆伝搬アルゴリズムによる階層型神経回路網を用いたパターン認識を対象とし，上記のパラメータと神経回路網の機能，すなわち学習の収束性および雑音が混入したパターンに対する認識能力との関係について検討する．更に，結合係数とユニットの入出力を2進符号表現するビット数の低減および非線形関数の簡単化に有効な学習方法を提案する．これらについて，「0」～「9」の数字パターンとアルファベットの大文字，小文字の52個のパターンを用いてシミュレーションを行い有効性を確認する．

2. 階層型神経回路網

本研究で対象とする2層の階層型神経回路網を図1

† 金沢大学工学部電気・情報工学科，金沢市
Faculty of Technology, Kanazawa University, Kanazawa-shi, 920 Japan

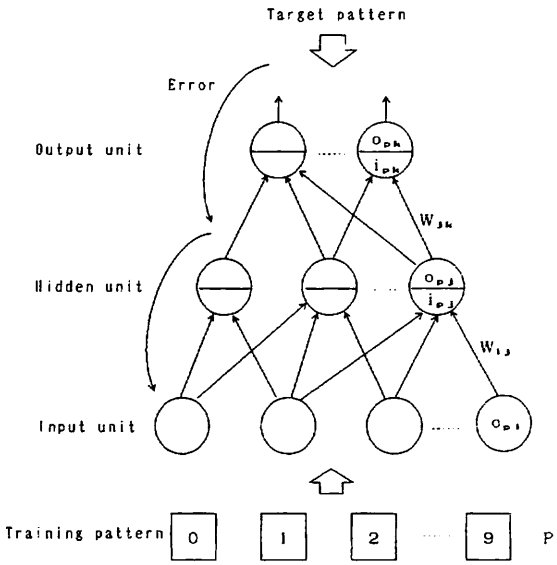


図1　誤差逆伝搬アルゴリズムによる階層型神経回路網
Fig. 1　A multilayer neural network, spervised by an error back propagation algorithm.

に示し，その学習方法を簡単に説明する．各層間の結合係数の学習は誤差逆伝搬アルゴリズム[1]により行う．入力層から隠れ層および隠れ層から出力層への結合係数をそれぞれ W_{ij} および W_{jk} とする．パターン p が提示されたときの入力層における第 i ユニットの出力を o_{pi}，隠れ層における第 j ユニットの入出力をそれぞれ i_{pj}，o_{pj} で表すと，これらは次式で関係づけられる．

$$i_{pj}=\sum_i W_{ij}o_{pi} \tag{1}$$

$$o_{pj}=f(i_{pj})=\frac{1}{1+\exp[-(\sum_i W_{ij}o_{pi}+\theta_j)]} \tag{2}$$

式(2)はシグモイド関数と呼ばれる非線形関数である．θ_j は隠れ層の各ユニットのバイアスを示す．出力層の各ユニットについても同様である．

各結合係数の修正量は次式である．

$$\Delta W_{ij}(n+1)=\mu\delta_{pj}o_{pi}+\alpha\Delta W_{ij}(n) \tag{3}$$

$$\Delta W_{jk}(n+1)=\mu\delta_{pk}o_{pj}+\alpha\Delta W_{jk}(n) \tag{4}$$

ここで，μ は1回の修正の大きさを決めるパラメータでステップサイズと呼ぶ．バイアスの項 θ_j は，常に1を出力するユニットに結合している結合係数と考えることができるので，これも同時に学習させる．また，結合の修正において，学習時の振動を減らし学習の収束を速めるために慣性項 $\alpha\Delta W$ が加えられている．ここで，α は小さな正の定数で，n は結合係数の修正回数を示す．

式(3)，(4)における δ_{pj} および δ_{pk} は次式で与えられる．

$$\delta_{pk}=(t_{pk}-o_{pk})\cdot f'(i_{pk})$$
$$=(t_{pk}-o_{pk})\cdot(1-o_{pk})\cdot o_{pk} \tag{5}$$

$$\delta_{pj}=(\sum_k W_{jk}\cdot\delta_{pk})\cdot f'(i_{pj})$$
$$=(\sum_k W_{jk}\cdot\delta_{pk})\cdot(1-o_{pj})\cdot o_{pj} \tag{6}$$

ここで，t_{pk} は教師信号である．また，$f'(\cdot)$ は $f(\cdot)$ の1次微分を示す．

学習手順は以下のようになる．

(1)　結合係数に乱数を用いて初期値を与える．

(2)　あるパターン p を提示し，出力を計算する．

(3)　最終層での誤差を計算し，結合係数を修正する．

(4)　(2)，(3)をすべての P 個のパターンについて繰り返す．

(5)　学習が収束したか否かをある誤差評価関係を用いてすべてのパターンについて判定する．

以上を1回の学習とする．

(6)　(2)～(5)を学習が収束するまで繰り返す．

3.　ビット数低減に対する学習方法

本研究では，ビット数低減に対する学習方法として，次の2通りを検討する．以下では，学習後に結合係数またはユニットの出力を表現する有効ビット数を m ビットとする．

(学習方法1)　学習時において常にビット数を m ビットに固定して学習する．

(学習方法2)　十分に多いビット数から学習を始め，学習の途中で段階的にビット数を減らし，最終的に m ビットにする．

学習方法1は，学習時においても回路を簡単化することを目的としている．しかし，結合係数やユニットの出力の初期値に対する依存性が高く，学習の収束性および認識率の劣化が予想される．

これに対して，学習方法2ははじめはビット数を大きく設定し，最適な結合係数や隠れユニットの出力に収束した後で量子化を行うため，最適解に近い良好な特性が得られることが期待できる．しかし，学習時に回路を簡単化できないという問題がある．現実的な解決策としては，学習は汎用計算機を用いてソフト処理で行い，パターン認識の運用時には専用ハードウェアを用いる方法が考えられる．

4. 結合係数におけるビット数の低減

4.1 量子化の方法

符号を除く有効ビット数を m とし，4捨5入で次のように量子化を行うことにする．

（1） 各層間における結合係数の値を仮に W とする．その最大値 W_{max} を求める．

（2） すべての結合係数の値を W_{max} で正規化する．W/W_{max}.

（3） 次式により W の量子化された値 W_q を求める．

$$W_q=[\mathrm{INT}(2^m W/W_{max}+0.5)]W_{max}/2^m \qquad (7)$$

ここで，INT(・)は整数化を行う演算子である．

例えば，$m=1$ の場合の量子化ステップは符号ビットまで含めて+1，+0.5，0，−0.5，−1の5ステップになる．これは通常の2進符号表示の量子化ステップとは若干異なる．

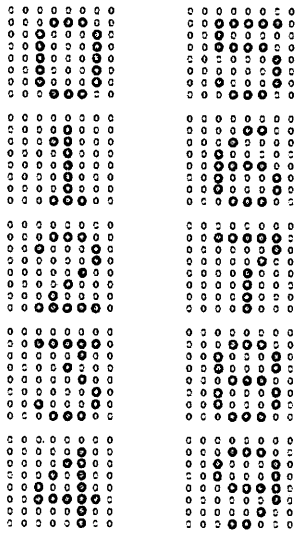


図2 数字パターン
Fig. 2 Number patterns.

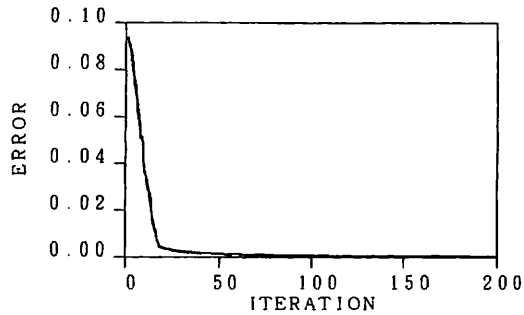
表1 教師信号

数字パターン	A (10ユニット)	B (4ユニット)
0	0000000001	0 0 0 0
1	0000000010	0 0 0 1
2	0000000100	0 0 1 0
3	0000001000	0 0 1 1
4	0000010000	0 1 0 0
5	0000100000	0 1 0 1
6	0001000000	0 1 1 0
7	0010000000	0 1 1 1
8	0100000000	1 0 0 0
9	1000000000	1 0 0 1

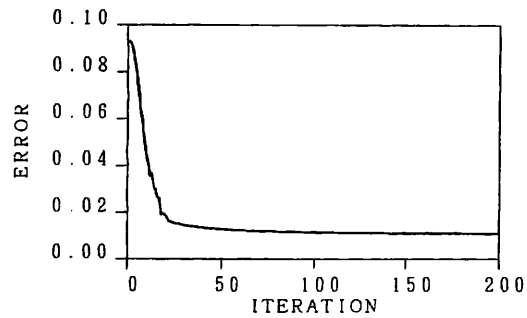
4.2 シミュレーション

認識対象のパターンは，図2に示す0，1の2値で表現される8×8行列の形の「0」〜「9」の10種類の数字パターンとする．また，それらの入力パターンに対応する教師信号としては表1に示すA，Bの二つのパターンについて検討する．但し，本章ではAを用いてシミュレーションを行う．結合係数の初期値は+0.5〜−0.5の範囲に分布する一様乱数とする．式(3)，(4)で結合係数を更新する際のステップサイズは $\mu=2.5$，慣性係数は $\alpha=0.6$ に設定する．隠れ層のユニット数は15個とする．

（1） 学習方法1　学習が収束するために必要な結合係数の最少ビット数を調べた．その結果，有効ビット数 m が5ビットまでは学習が収束することが確認できた．学習が収束した場合と収束しない場合の学習の様子をそれぞれ図3(a)，(b)に示す．横軸は学習回数を示している．2.で述べたように，「0」〜「9」のすべての入力パターンについて結合係数の修正を行うことを1回の学習回数としている．縦軸は出力層での各



(a)収束した場合（結合係数:5ビット）
(a) Convergence (Connections: 5bits).



(b)収束しない場合（結合係数:4ビット）
(b) Divergence (Connections: 4bits).

図3 学習の収束特性（学習方法1）
Fig. 3 Learning curves, using learning method 1.

ユニットに対する教師信号と実際の出力との誤差の2乗平均を更に入力パターンに対して平均化したものであり，次式で表される．

$$\mathrm{ERROR}=\frac{1}{P}\sum_{p=1}^{P}\left[\frac{1}{K}\sum_{k=1}^{K}(t_{pk}-o_{pk})^2\right] \qquad (8)$$

（2） 学習方法2　　結合係数のビット数を3段階に分けて低減した．学習回数が0～150回までは有効ビット数 m を20ビット，150～300回では8ビット，それ以降は目標とするビット数とした．目標ビット数は5～1ビットについて調べた．その結果，最終的な有効ビット数 m を1ビット（＋符号1ビット）まで減らしても学習が収束することが確認できた．ここで，1ビット＋符号1ビットとは，**4.1**でも述べたように，＋1，＋0.5，0，－0.5，－1の5ステップの表現を意味する．学習の収束の様子を図4に示す．誤差が3回目の量子化により一時的に増加するが，学習によりうまく収束している．

以上の結果から，結合係数のビット数はかなり減らすことができ，また，学習方法2がビット数低減には有効であることが確認された．

5. ユニットの出力を表現するビット数の低減

隠れ層と出力層の各ユニットの出力を表現するビット数の低減について検討する．各ユニットの出力を数ビットで表すことも考えられるが[(1),(8)]，$W_{jk}o_{pj}$ の乗算をなくし，式(2)のシグモイド関数を簡単化することにより，ディジタル回路の規模を大幅に低減できることから，本研究では各ユニットの出力を0,1の2値で表現する場合に限定して学習アルゴリズムおよび収束性について検討する．

5.1 学習アルゴリズムの変形

ユニットの出力を0，1で表現することは，しきい値を0とすると，ユニットの入出力関係は次式で与えられる．

$$i_{pj}\geqq 0 \quad ならば \quad o_{pj}=1 \qquad (9)$$
$$i_{pj}<0 \quad ならば \quad o_{pj}=0$$
$$i_{pk}\geqq 0 \quad ならば \quad o_{pk}=1 \qquad (10)$$
$$i_{pk}<0 \quad ならば \quad o_{pk}=0$$

すなわち，「しきい関数」を用いることに相当する．この場合は，式(5)～(6)において1次微分 $f'(i_{pk})$，$f'(i_{pj})$ の値が常に0になるため，結合係数の修正ができない．そこで，式(5)，(6)における1次微分を次のように微小な正定数で置換する．

$$\delta_{pk}=(t_{pk}-o_{pk})\cdot f'(i_{pk}) \rightarrow (t_{pk}-o_{pk})\cdot f_o \qquad (11)$$
$$\delta_{pj}=(\sum_k W_{jk}\cdot\delta_{pk})\cdot f'(i_{pj}) \rightarrow (\sum_k W_{jk}\cdot\delta_{pk})\cdot f_h \qquad (12)$$

f_o，f_h はそれぞれ出力層および隠れ層における1次微分項に対応する．

5.2 シミュレーション

隠れ層および出力層の出力を**4.1**の方法で量子化し，さらにその値を0，1の2値で表現して学習させた場合の学習の収束性について調べた．**4.**と同様に学習方法1および2の場合について検討した．式(9)，(10)の係数修正式における1次微分項，f_o および f_h は0.05とした．また，教師信号はAを用い，隠れユニットの数は15個とし，結合係数は量子化しないものとする．

（1） 学習方法1　　各ユニットの出力は，学習の最初から0，1の2値のみをとるという条件下で学習させた．学習の収束の様子を図5に示す．少し振動が見られるが良好に収束している．

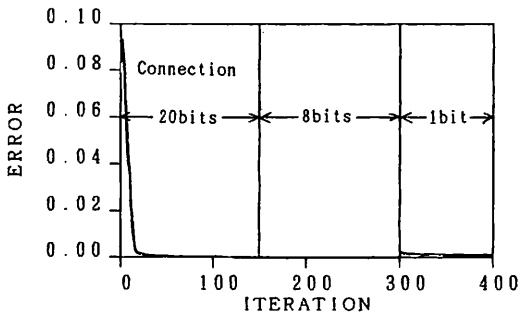


図4　学習の収束特性（学習方法2）
Fig. 4　Learning curve, using learning method 2.

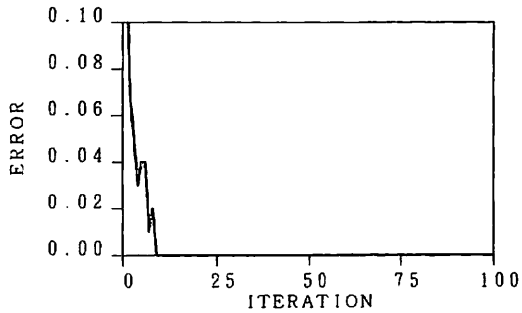


図5　学習の収束特性（学習方法1，ユニットの出力：2値）
Fig. 5　Learning curve, using learning method 2 (Outputs of units : 2-level (1, 0).

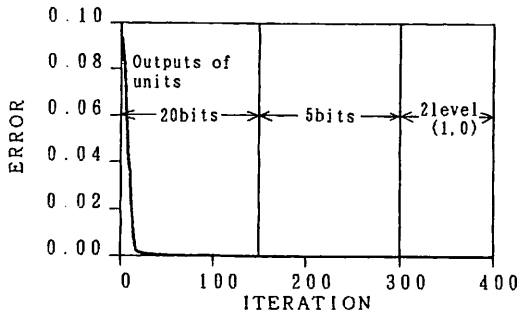


図6 学習の収束特性(学習方法2)
Fig. 6 Learning curve, using learning method 2.

(2) 学習方法2　出力の有効ビット数は，学習回数が0～150回までは20ビット，150～300回では5ビット，それ以降は0，1のみを用いて学習させた．学習の様子を図6に示す．この場合は途中の量子化においても誤差が振動することなく収束している．これは隠れユニットの出力が0，1に近づく傾向があるために量子化による変動が小さいことによるものである．

以上の結果から，各ユニットの出力を0，1の2値で表現し，非線形関数として微分不可能なしきい関数を用いても，学習アルゴリズムを変形することにより学習が収束することが確認できた．これにより，$W_{jk}O_{pj}$の乗算と，シグモイド関数による変換が不要となり，ハードウェアが大幅に簡単化できる見通しが得られた．

6. 結合係数およびユニット出力におけるビット数の低減

6.1 隠れユニット数を固定

本章では更に，各ユニットの出力および結合係数のビット数を同時に低減した場合の学習方法1および2による学習の収束性について調べる．隠れユニットの数は15個に固定し，その他の条件は**4.**，**5.**と同じとした．

(1)学習方法1　各ユニットの出力は0，1の2値とし，結合係数のビット数を低減して学習を行った．その結果，結合係数の有効ビット数が4ビットまでは学習が収束することが確認できた．学習の様子を図7に示す．

(2) 学習方法2　学習時における結合係数およびユニット出力のビット数を**4.2**および**5.2**と同様に段階的に低減して学習を行った．その結果，結合係数の有効ビット数を1ビットまで減らしても学習が収束可能であることが確認できた．その学習の様子を図8に示す．従って，この場合も学習方法2の方がビット数低減には有効である．

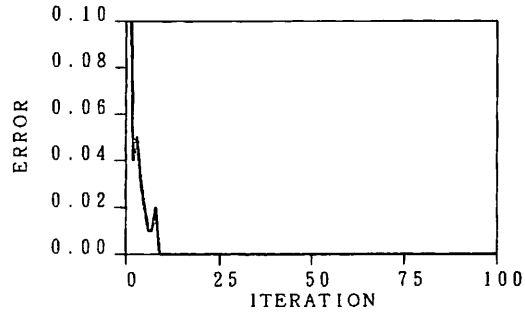


図7 学習の収束特性
Fig. 7 Learning curve, using learning method 2 (Connections : 4bits, Outputs of units : 2-level (1, 0)).

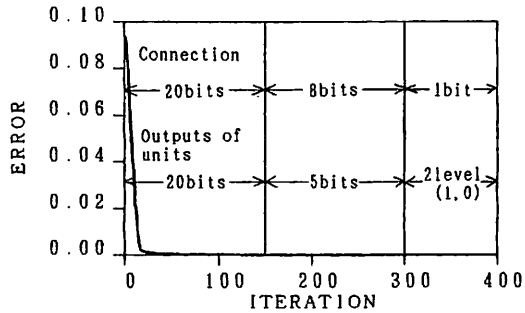


図8 学習の収束特性(学習方法2)
Fig. 8 Learning curve, using learning method 2.

6.2 隠れユニットの数と学習の収束性

6.1では隠れユニットは15個に固定したが，ここでは，隠れユニットの数と学習が収束可能な結合係数のビット数との関係について調べる．学習方法は1の方法を用い，教師信号は表1のAとBについて行い比較する．学習時の結合係数の初期値として5種類の乱数を用いて学習を行った．

シミュレーション結果を表2に示す．表中の数字は，結合係数の初期値として与えた5種類の乱数のうちで学習が収束した乱数の数を示す．上段が教師信号Aの場合で，下段がBの場合である．隠れユニットが3個以下および結合係数の有効ビット数が2ビット以下の場合は学習は全く収束しなかった．また，網かけの部分は5種類の乱数に対して常に学習が収束する領域である．

この結果より，いずれの教師信号についても，学習が収束するためには必要な隠れユニットの数と結合係

表2　隠れユニットの数と結合係数のビット数

隠れユニットの数［個］	3	4	5	6	7	8	9	10
12	4 1	3 4	3 3	5 6				
11	3 2	3 2	4 2	5 5				
10	3 0	2 0	3 3	4 4	5 4	5 5		
9		1 0	1 2	4 3	5 3			
8			2 0	5 4	5 4	5 5		
7				3 2	2 2	3 5	3 4	4 3
6					2 2	2 2	3 1	4 2
5	教師信号			A B			1 2	3 2
4								0 0

結合係数のビット数［bit］

数のビット数はほぼ反比例の関係にあることがわかる．すなわち，隠れユニットの数を減らしてユニット間結合を低減しようとしても，結合係数のビット数が増加するため，ハードウェアの規模としてはある下限が存在することになる．

同様のシミュレーションを学習方法2についても行った結果，結合係数のビット数を更に減らしても同様な傾向になることが確認できた．

6.3　ユニットの入力信号を表現するビット数

隠れユニットの入力信号は W_{ij} の加算となり，出力ユニットの入力も隠れユニットの出力を0，1で表現すれば W_{jk} の加算となる．従って，これらを表現するためのビット数はいずれも結合係数のビット数よりも増加する．この入力信号を表現するために必要なビット数を調べるために，結合係数および各ユニットへの入力の大きさをシミュレーションにより調べた．学習方法は2を用い，最終的な結合係数の有効ビット数は5ビット，各ユニットの出力は0，1の2値として学習させた．その結果，入力層から隠れ層への結合係数 W_{ij} と隠れユニットへの入力 i_{pj} の最大値は各々1.74，7.74，であり，隠れ層から出力層への結合係数 W_{jk} と出力ユニットへの入力 i_{pk} の最大値は各々3.42，16.1であった．ユニットの入力を表現するためのビット数は，これらの比

$7.74/1.74 \fallingdotseq 4.45$

$16.1/3.42 \fallingdotseq 4.71$

が $2^2 < 4.45$，$4.71 < 2^3$ を満たすことから，結合係数のビット数に比べて3ビット多く必要とする．

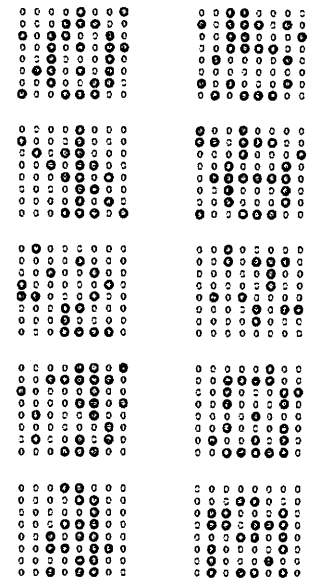


図9　雑音が混入した数字パターン
Fig. 9　Number patterns corrupted by random noise.

7.　雑音混入パターンの認識能力

4.～6.では，学習の収束性の観点からビット数低減の検討を行った．その結果，学習時に最初からビット数を低減する学習方法1でも，ユニットの出力を0，1の2値で表し，結合係数も4ビットまで減らすことができた．これは学習時においてもハードウェア規模を縮小できることを示唆するものである．

では，学習後の神経回路網を用いてパターン認識を行う場合，入力層に混入する雑音に対する安定性はどうであろうか．本章では，ランダムな雑音が混入したパターンの認識能力の点からビット数低減および学習方法の有効性を検討する．

7.1　ランダム雑音の発生

雑音は入力層の8×8＝64ユニットの(1，0)を乱数により反転させる形で発生させた．(1，0)が反転したユニットの数を雑音量とする．雑音量を1個ずつ増やしてシミュレーションを行い認識誤りを生じる雑音量を「0」～「9」のパターンについて求めた．雑音の混入したパターンの例を図9に示す．

7.2　学習方法による認識能力の比較

教師信号はAとし，隠れユニットの数を15個に固定して，各学習方法で得られた回路網による認識能力の比較を行う．

(1)　標準学習方法　　結合係数および隠れ層と出

力層の各ユニットの出力を量子化せずに学習する方法である．実際のシミュレーションプログラムでは単精度の浮動小数点で表示されている．

（2） 学習方法1　学習時の結合係数のビット数は6ビットに，また各ユニットの出力の値は0，1の2値に固定して学習させた．

（3） 学習方法2　結合係数のビット数は学習回数が0～150回までは20ビット，150～300回では8ビット，それ以降の最終的なビット数を5，3，1ビットの3通りの場合について学習させた．また，各ユニットの出力のビット数は学習回数が0～150回までは20ビット，150～300回では5ビットにし，それ以降は0，1の2値とした．

上の3通りの学習方法で得られた回路網を用いて，**7.1**で述べたランダム雑音が混入したパターンの認識を行った結果を表3にまとめる．表中の数字は，ランダム雑音が加わった各数字パターンに対して認識誤りを生じる最少の雑音量である．これは，30種類の雑音パターンについての平均値である．小数点以下は4捨5入されている．この表より，学習方法1によって得られた回路網では雑音混入により認識能力が大きく劣化していることがわかる．一方，学習方法2の場合は結合係数のビット数を3ビットぐらいまで減らしても，標準学習方法の場合に比べて認識能力はあまり劣化していない．これらの比較検討はさらに**7.3**で行うことにする．

更に，アルファベットの大文字，小文字の52個のパターンについて同様のシミュレーションを行った結果を表4に示す．この場合，入力層は16×16＝256個であり，教師信号はA，隠れユニットは52個とした．また，ステップサイズ$\mu=0.6$，$\alpha=0.5$とした．結合係数の最終ビット数は学習方法1が6ビット，学習方法2が3ビットである．表4から，標準学習方法と学習方法2はほぼ同じ結果が得られているが，学習方法1の雑音特性は大きく劣化していることがわかる．従って，雑音の混入まで考慮した場合は，学習方法2が一般のパターン認識に対して有効であることが期待できる．

10個の数字パターンについてのシミュレーションを教師信号Bについても行った．学習方法としては，標準学習と学習方法2の場合について行った．結果を表5に示す．条件は教師信号Aの場合と同じであり，結合係数の最終ビット数は5である．表3と比較するといずれの学習方法においても認識能力が劣化していることがわかる．従って，雑音に対する安定性からは出力ユニットを減らすことは必ずしも得策ではない．

7.3 隠れユニットの出力による学習方法の比較

7.2の数字パターンに対する結果について，隠れユニットの出力パターンの点から各学習方法による認識能力を比較する．教師信号はAとする．標準学習，学習方法1および2で学習させた回路網において，パターン認識時の各数字パターンに対する隠れユニットの出力を調べる．結合係数のビット数は学習方法1では6ビット，学習方法2では5ビットとした．また，各ユニットの出力の値は0，1の2値になっている．

表3 数字パターンに対して認識誤りを生じる最少雑音量

数字ﾊﾟﾀｰﾝ		0	1	2	3	4	5	6	7	8	9
標準学習		16	17	18	18	22	17	16	14	14	17
学習方法1		1	5	2	3	2	2	3	3	2	2
学習方法2	5ﾋﾞｯﾄ	13	15	17	14	18	15	15	14	14	16
	3ﾋﾞｯﾄ	13	15	14	15	19	15	15	13	14	15
	1ﾋﾞｯﾄ	7	15	17	8	17	10	6	8	8	16

表5 数字パターンに対して認識誤りを生じる最少雑音量（教師信号Bの場合）

数字ﾊﾟﾀｰﾝ	0	1	2	3	4	5	6	7	8	9
標準学習	13	12	13	14	16	15	15	18	12	18
学習方法2	8	7	11	10	14	16	12	16	10	17

表4 アルファベットパターンに対して認識誤りを生じる最少雑音量

ﾊﾟﾀｰﾝ	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
標準学習	29	28	20	28	24	27	37	23	34	32	36	17	37	23	22	33	33	31	34	31	23	24	34	30	30	30
学習方法1	4	3	1	1	2	1	1	1	1	2	1	1	1	1	2	2	2	1	2	1	2	2	1	1	2	2
学習方法2	34	32	18	32	28	29	45	21	27	32	39	13	33	22	22	36	33	30	33	33	28	26	38	34	30	33
ﾊﾟﾀｰﾝ	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
標準学習	40	25	28	32	26	31	31	27	15	31	33	28	46	39	27	21	36	31	28	23	28	40	42	19	23	30
学習方法1	1	2	1	1	3	2	1	2	1	1	2	1	1	2	1	1	3	2	1	2	1	2	1	1	2	1
学習方法2	48	26	31	41	27	35	29	30	14	37	32	34	58	42	22	19	30	31	31	23	36	42	45	30	28	38

表6　各学習方法における隠れユニットの出力

		隠れ層の出力							
パターン1	標準	0.89	0.15	0.97	0.80	0.97	0.06	0.01	1.00
		0.00	0.95	0.98	0.96	0.96	0.08	0.03	
	方法1	1.00	0.00	0.00	0.00	1.00	0.00	1.00	1.00
		0.00	1.00	1.00	1.00	0.00	0.00	1.00	
	方法2	1.00	0.00	1.00	1.00	1.00	0.00	0.00	1.00
		0.00	1.00	1.00	1.00	1.00	0.00	0.00	
パターン5	標準	0.83	0.05	0.03	0.92	0.98	0.11	0.44	0.02
		0.80	0.01	0.18	0.02	0.00	0.96	1.00	
	方法1	0.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00
		0.00	0.00	0.00	0.00	0.00	1.00	0.00	
	方法2	1.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00
		1.00	0.00	0.00	0.00	0.00	1.00	1.00	

数字パターン「1」と「5」に対する隠れユニットの出力を表6に示す．標準学習と学習方法2ではそのパターンが非常に類似している．これは，学習方法2では6.1の図8でも示したように，20ビットの精度で学習がほぼ収束した後で量子化を行っていること，および隠れユニットの出力が1または0に収束する傾向があることによるものである．後者は式(3)〜(6)の結合係数の修正量を求める際にシグモイド関数の1次微分が用いられていることに起因している．すなわち，ユニットの出力が1または0に近づくにつれて1次微分が小さくなり，結合係数の修正量が小さくなるためである．

これに対して，学習方法1の隠れユニットの出力パターンは前2者とは全く異なっている．これは，学習方法1ではじめからユニットの出力を0，1としているために，初期値に依存したある特殊解に一気に収束してしまうためである．

文献[9]でも検討したように標準学習，または十分大きいビット数を用いた学習による回路網では，隠れユニットの出力からなるベクトルは学習後においては各パターンに対して互いに線形独立に近い形になる．そのため，各入力パターンに対する隠れユニットの出力ベクトル間の距離が大きくなり，結合係数の大きさが平均化され，雑音に対する伝達係数が小さくできる．

これに対して，学習方法1による解は結合係数とユニット出力の初期値に大きく依存し，沢山ある解のうちのある特殊な解に収束してしまう．そのため，隠れユニットの出力ベクトルは各パターンに対して互いに線形独立な形に近づくことができない．すなわち，各パターンに対する出力ベクトル間の従属性が高くなる．

この点について，10個の数字パターンを用いたシミュレーションにより更に詳しく調べてみる．各パターンに対する隠れユニットの出力ベクトル間の距離を表7に示す．パターン「i」と「j」に対する隠れユニットの出力ベクトルを

表7　数字パターンに対する隠れユニットの出力ベクトル間の距離

パターン	0	1	2	3	4	5	6	7	8	9	
0		8	4	4	4	4	5	3	4	5	学習方法1
1	14		10	4	8	6	7	5	8	7	
2	7	7		6	4	6	7	5	6	5	
3	7	7	6		4	2	7	3	8	5	
4	8	8	7	7		4	7	5	6	3	
5	7	9	10	4	7		7	3	8	7	
6	7	7	4	6	7	8		6	3	6	
7	7	9	8	10	9	8	8		5	6	
8	5	9	6	6	9	8	6	6		5	
9	6	8	7	7	6	7	11	7	5		
	学習方法2										

$$\boldsymbol{o}_i=[o_{i1}, o_{i2}, \cdots, o_{iK}] \tag{13a}$$

$$\boldsymbol{o}_j=[o_{j1}, o_{j2}, \cdots, o_{jK}] \tag{13b}$$

とすると，表中の i 行 j 列の数値 $n(i, j)$ は次式で与えられる．

$$n(i,j)=\sum_{k=1}^{K} |o_{ik}-o_{jk}| \tag{14}$$

$n(i, j)$ はベクトル間の距離を表しており，この値が大きいほど独立性が高いことになる．表の右上半分が学習方法1の場合，左下半分が学習方法2の場合の $n(i, j)$ を表している．表7の結果から学習方法1の方がベクトル間の距離が小さく，従属性が高いことがわかる．

従属性の高い出力ベクトルを用いて，あらかじめ定められた教師信号を得るために結合係数の大きさにアンバランスが生じる．すなわち，結合係数は各パターンに対する隠れユニットの出力ベクトルの違いを強調するように決められるため，その大きさが平均化されない．大きな結合係数によって雑音は増幅されるため，容易に認識誤りを生じることになる．このように，学習方法1では初期値に依存する形で沢山ある解のうちの一つに収束することができても，雑音に対して安定な解に収束することは難しいと言える．

7.4　隠れユニットと結合密度の低減

回路規模の低減について考える場合，隠れユニットの数[1],[8]および結合の密度を減らすことも考えられる．そこで，それらを減らして得られる回路網の雑音に対する安定性について検討を行った．雑音のパターンは7.2と同じ30種類のパターンを用いた．

まず，隠れユニットの数を5個にして，学習後の回路網の雑音混入パターンに対する認識能力を調べた．学習方法としては標準学習と学習方法2について調べ

表8 隠れユニット5個の場合の認識結果

数字ﾊﾟﾀｰﾝ	0	1	2	3	4	5	6	7	8	9
標準学習	18	17	17	16	17	14	15	15	14	11
学習方法2	18	16	17	18	18	15	12	9	14	5

表9 結合密度を減らした場合の認識結果

数字ﾊﾟﾀｰﾝ	0	1	2	3	4	5	6	7	8	9
結合密度 20%										
標準学習	11	10	13	11	17	12	11	12	12	12
学習方法2	10	10	12	10	16	10	10	9	9	9
結合密度 8%										
標準学習	17	10	10	9	13	16	13	14	10	14
学習方法2	14	12	8	6	11	14	15	11	8	13

比較した．学習方法2における結合係数のビット数の低減は20ビット(0～150回)→8ビット(151～300回)→5ビット(301回～)とした．また，各ユニットの出力のビット数の低減は20ビット(0～150回)→5ビット(151～300回)→2値(0，1)(301回～)とした．結果を表8に示す．学習方法2の認識能力が少し劣化している．また，表3の隠れユニットを15個用いた場合に比べると，いずれの学習方法においても認識能力は悪くなっていることがわかる．

次に，結合密度を減らした回路網で学習を行い，雑音混入パターンに対する認識能力を調べた．結合密度の減らし方は，各結合係数に対して乱数を発生し，その値があるしきい値以下の場合に，その結合係数を零に固定する方法とした．隠れ層から出力層への結合密度は約70％に固定し，入力層から隠れ層への結合密度が約20％および8％の場合についてシミュレーションを行った．隠れユニットは15個に固定した．結果を表9に示す．表3と比較すると，いずれの学習方法でも認識能力は結合密度を減らすことによって悪くなっている．

以上のシミュレーションから，雑音に対する認識能力を考慮すれば，隠れユニットおよび結合密度を減らすことは得策ではないと言える．

8. 回路規模低減のための指針

これまでの検討より，雑音まで考慮したパターン認識を対象とした場合に階層形神経回路網のディジタル構成による回路規模を低減するための指針は次のようにまとめられる．

(1) 結合係数のビット数は数ビットまで減らすことができる．

(2) 各ユニットの出力は0，1の2値で表現可能である．従って，実際の回路において乗算器が不要となる．

(3) (2)は非線形関数としてシグモイド関数の代わりにしきい関数を用いることに相当するから，複雑な非線形変換が不要となる．

(4) (3)に対して誤差逆伝搬アルゴリズムを変形する．すなわち，1次微分の代わりに微小な正定数を用いる．

(5) 結合係数およびユニット出力のビット数は学習の初期には大きく設定し，学習の途中で段階的に低減する学習方法が有効である．そのため，学習時は回路規模を低減できない．これに対しては，学習は汎用計算機を用いたソフト処理で行い，実際のパターン認識としての運用時にはビット数を低減した専用ハードウェアで行う方法が実用的なアプローチである．

(6) 隠れユニットおよび結合密度を減らして回路規模を減らすことは雑音耐力の点から得策ではない．

(7) 教師信号については，表1のBよりもAを用いた方が雑音に対して安定であり，出力ユニットを減らすことは得策ではない．

9. むすび

本論文では，誤差逆伝搬はアルゴリズムによる階層型神経回路網を用いたパターン認識を対象とし，ディジタル回路で実現した場合のハードウェア低減の可能性について，ユニット数，結合係数とユニット出力を表現するビット数，結合密度などの低減および，ユニットの入出力変換関数の簡単化などの点から検討を行った．

その結果，結合係数のビット数は数ビットまで減らすことが可能であり，隠れ層と出力層のユニット出力は0，1の2値で表現できることがわかった．更に，このようなビット数低減に有効な学習方法の改良を行った．これは，ビット数の低減を学習中に段階的に行うこと，および誤差逆伝搬アルゴリズムにおける1次微分係数を微小な正定数で置換することの2点である．以上の検討内容については，10個の数字パターンと52個のアルファベットパターンを用いて計算機シミュレーションによりその有効性を確認した．

必要とされるビット数やユニット数は認識対象のパターンの数にも依存すると考えられるので，更にパターンの数を増やして検討する予定である．

謝辞 日ごろ御指導御討論頂く山本外史教授，なら

びに山本・中山研究室の学生諸氏に深謝致します．

文　献

(1) D. E. Rumelhart and J. T. McClelland : "Parallel Distributed Processing", MIT Press (1987).

(2) 麻生英樹："ニューラルネットワーク情報処理"，産業図書 (昭 63).

(3) T. J. Sejnowski and C. R. Rosenberg : "Parallel networks that learn to pronounce English text", Complex Systems, 1, pp. 145-168 (1987).

(4) 迫江博昭，磯　健一："ダイナミックニューラルネットワークの検討"，信学技報，SP87-101 (1987-12).

(5) J. Tomberg, et al : "Fully digital neural network implementation based on pulse density modulation", Proc. IEEE CICC, pp. 12.7.1-12.1.4 (1989).

(6) Y. Hirai,et al : "A digital neuro-chip with unlimited connectability for large scale neural networks", Proc. IJCNN89, 2, pp. 163-169 (1989).

(7) 平井，岩田："ニューラルコンピュータと LSI"，スペクトラム，2, 12, pp. 52-62 (1-12).

(8) G. W. Cottrell, P. Munro and D. Zipser : "Image compression by back propagation : an example of extentional programming", Tech. Rep. 8702, Univ. of California, San Diego, Institute for Cognitive Science (1987).

(9) 猪股　悟，中山謙二："誤差逆伝搬モデルによる雑音の加わった数字パターンの認識に関する検討"，信学技報，PRU89-23 (1989-06).

(平成元年 12 月 22 日受付，2 年 3 月 23 日再受付)

中山　謙二

昭 46 東工大・工・電子卒．昭 46〜47 同大学研究生．昭 58 同大学より工学博士の学位取得．昭 47 日本電気(株)入社．伝送通信事業部および C&C システム研究所にて通信用各種フィルタ，ディジタル信号処理，信号理論，ニューラルネットに関する研究開発に従事．昭 63 年 8 月金沢大学・工・電気情報助教授．平 2 年 4 月同教授．昭 62.9 IEEE Circuits & Devices Mag.論文賞受賞．著書「SC 回路網の設計と応用」(東海大学出版会) 他．IEEE シニア会員．

猪股　悟

平 1 金沢大・工・電気情報卒．現在，同大大学院修士課程在学中．階層型神経回路網によるパターン認識に関する研究に従事．

竹内　由幸

平 2 金沢大・工・電気情報卒．同年三谷産業入社．在学中，階層型神経回路網によるパターン認識に関する研究に従事．