

## 回路構造の冗長化と多様化によるニューラルネット 学習の高精度，高速化<sup>†</sup>

熊 沢 逸 夫<sup>††</sup>

近年，ニューラルネットの学習方式として，バックプロパゲーションが提案されているが，この手法は，勾配法に基づきパラメータを逐次修正していくため，誤差評価関数が極小点を持つ場合，最小点への収束が保証されない。また最小点が求まったとしても，そこで誤差が0になるとは限らない。本論文では，誤り訂正符号の理論に基づき，回路構造の冗長化と多様化の戦略をあらかじめ学習手続きの中に組み込んでおくことにより，学習結果に不完全さがあっても，それを補償し，少ないサンプルを使って，高精度に目的の機能を得る方式を提案する。冗長に設定された目標に対して，その目標を異なる手続き，異なる手順で実現する多様な構造の回路が得られれば，それらは独立に誤りを生じると考えられる。したがって，あらかじめ学習目標に設定されている冗長性を活用して，出力の誤りを訂正することが可能になる。このような立場を取ると，不完全でも良いから，構造の異なる多様な回路をできるだけ多数構成することが学習の課題となる。そのような回路を幾らでも多数構成できるならば，誤り訂正により，幾らでも精度良く目標を実現できるはずである。本論文では，回路構造に多様性を生じさせる工夫を幾つか提案し，この可能性を簡単なパターン認識の実験によって確認する。実験結果によれば，認識率，学習速度ともに改善されることが明らかになる。

### 1. はじめに

近年，バックプロパゲーション<sup>1)</sup>を初めとするニューラルネットの学習方式が注目を集め，パターン認識，人工知能，信号処理等の分野で盛んに応用されている。しかしながら，これらの学習手法は，その振舞いが理論的に不明確なまま，とりあえず適用してみたら良い結果が得られたというので評判になっているのが実状であり，実際には，従来から学習手法が抱えている基本的問題点は何も解決されていない。

最近のニューラルネットのブームは，実は，このように理論的保証のない，実験的成功例をよりどころとして成り立っていると言っても過言ではない。もともとバックプロパゲーションの原型は，甘利<sup>2)</sup>によって1967年に発表されているが，計算機の性能が極めて低く，実験的検証が不可能であった当時，この理論が注目を集めなかったのも，やむをえないことであったのであろう。

現在，学習手法が抱えている問題点を整理してみると次のようになる。

[ニューラルネット学習方式の問題点]

(i) 勾配法を基礎とした学習方式では，逐次修正手続きが誤差評価関数の極小点に捕まってしまう恐れがあり，必ずしも誤差評価関数を最小とするパラメー

タを求めることができない。

(ii) たとえ，誤差評価関数の最小点が見つかったとしても，そこで誤差が0になるとは限らない。

(iii) 初期構造をどのように与えたら良いのかわからない。回路の層数，素子数，結線構造は，ランダムに与えるか，実験結果によって試行錯誤的に決めるしかない。

(iv) 勾配法では，パラメータの値を少しずつ修正することしかできないため，誤差曲面の形状いかんによっては，学習に長時間を必要とする。

(v) 初めて与える入力に対して，適切な出力が得られるとの保証が何もない。つまり汎化の能力を理論的に説明できない。そのため得られたシステムの利用には常に不安が付きまとう。

最近では，以上の問題点を解決するために，回路にどのような初期構造を与えることが望ましいかという問題，いわゆる pre-wiring の問題が研究されている。もしも万能，完全な学習方式が存在すれば，それを用いることにより，pre-wiring で与えるべき初期構造でさえ，自動的に形成することができるはずである。しかしながら，最近の研究アプローチでは，そのような万能な学習方式を得ることはあきらめるべきであるとする立場が取られている。実際，人間の脳は，かなりの部分が先天的に定められ，特異的な構造を有している。それは進化の過程で形成されたものであり，そのような構造までも学習によって形成するのは困難であると考えられる。

ところで，pre-wiring によって学習の性能を上げよ

<sup>†</sup> Improvement of the Learning Capability of Neural Networks  
Based on a Redundant and Diverse Structuring by Itsuo  
KUMAZAWA (Department of Computer Science, Faculty of  
Engineering, Tokyo Institute of Technology).

<sup>††</sup> 東京工業大学工学部情報工学科

うとする場合、いかにして問題領域依存的にならないように初期構造を与えるか、ということが問題になってくる。システムが問題固有のものになってしまい、特定の問題にしか使えないのでは、学習方式の利点は薄れてしまうのである。

本論文では、次のような立場から、問題領域に依存せず、一般的に与えることができる pre-wiring の一手法を提案する。

誤り訂正符号の理論によれば、伝送誤りが生じる不完全な通信路も、送信する信号を冗長符号化することにより、目的上は完全な通信路として利用することができる。これは、通信路の性能を物理的に向上することをあきらめ、その使い方を工夫することにより確実な通信を実現したものであるといえる。

本論文でも、同様の観点から、学習方式の性能を向上することを試みる。すなわち、回路の初期構造に適当な冗長性を与えておき、学習の結果構成された回路に不完全な部分があっても、それを補償し目的上は完全な回路として利用できるようにするのである。

従来の学習方式の改善の試みは、学習方式自身の精度を直接高めようとしたものであり、通信路の物理的性能の向上に対応するものであった。それに対して、本論文では、外部から要求される目標を学習方式の直接の目標とするのではなく、内的な目標を設定し、それが不完全に満たされるだけで外部から要求される目標は完全に満たされるようにする。このようにしておけば、学習結果が不完全で、内的な目標が不完全にしか満たされていないとしても、外部から要求された目標は正しく実現できることになる。これは、通信路の利用法の工夫に対応したアプローチである。

ただし、この冗長化によるアプローチを成功させるためには、1つ、大きな問題を解決しなければならない。それは回路構造の多様化の問題である。

通信の誤り訂正では、コード中の多数のビットに誤りが同時に生じると、誤りを訂正することができなくなる。ニューラルネットの構造の冗長化でも、同じ問題が生じる。例えば、回路の単純な多重化によって構造を冗長化した場合を考えると、各回路の誤りの特性が同じで、各回路の出力に誤りが同時に生じてしまうと、それらの多数決をとっても誤りを減らすことはできない。冗長化が効果を発揮するためには、各回路の誤りの特性が異なり、誤りが独立に発生するようになっている必要がある。つまり、各回路に同一の目標出力が要求される場合でも、各回路は、それを、そ

れぞれ異なる方法、異なる内部構造によって実現しなければならないのである。

このように、冗長な目標出力を得るために、誤りの特性が異なる様々な実現手段（異なる構造の回路）を見つけ出すことを、本論文では、回路構造の多様化と呼ぶ。また、この考え方を一般化し、多重化に限らず、一般にハミング距離の離れた目標出力ベクトルが要求される回路で、出力ベクトルの各成分の誤りが独立に生じるとき、その回路の構造に多様性があると呼ぶことにする。

本論文では、まず、以上の観点から、回路構造の冗長化と多様化に基づく学習原理を定式化した後、具体的に、回路の構造に多様性を生じさせる工夫を幾つか述べる。最後に、それらの工夫が学習の高速化と精度改善にどのような効果を持つかということを実験的に確認する。

## 2. 回路構造の冗長化、多様化と学習性能の改善

### 2.1 外部から要請される目標と内的な目標

本論文では、フィードフォワード型の回路を対象として、回路構造の冗長化と多様化の問題を検討することにする。1章に述べたように、学習が抱える様々な問題を解決するためには、回路のすべての構造を学習によって形成しようとするのではなく、一部の構造を先天的に与えておく必要がある。ここでは、図1に示す構造をあらかじめ与えることにする。

図1で、学習の目標は、入出力関係  $y = F(x)$  を実現することである。図に示す回路では、これを内部に設けられた表現  $z$  を介して実現する。 $z$  は、写像  $y = T(z)$  によって、 $y$  に対応付けられるが、 $y$  に対応付く  $z$  が、互いにハミング距離の離れたものとなるように、写像  $T$  を定める。写像  $T$  を実現する回路は、初期構造として与え、学習の間固定しておく。

写像  $z = G(x)$  を学習によって形成するが、学習の

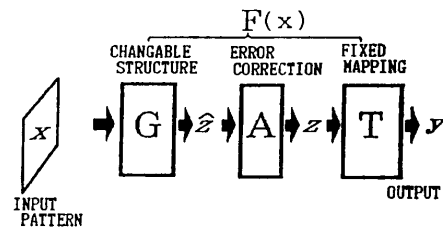


図1 冗長な内部表現と誤り訂正の機構を持つシステム  
Fig. 1 The network structure using redundant representation for error correction.

結果得られる  $G$  は正しく形成されるとは限らないので、誤りを含む可能性のあるその出力を  $\hat{z}$  によって表しておく。誤り訂正符号の理論によれば、 $z$  間の最小ハミング距離を  $L$  とした時、 $\hat{z}$  の成分が  $\lfloor (L-1)/2 \rfloor$  個まで誤っても、それを訂正することができる。ここで、 $\lfloor \ ]$  はその内部の数を越えない最大の整数を意味する。

図中の  $A$  は誤りの訂正を行う回路であるが、1つの方法として、平衡状態に収束しようとするリカレントな回路を用いて誤りを訂正することができる。互いにハミング距離の離れた平衡状態を持つリカレントな回路の構成方法は文献 3)、4) に示されている。

以上のように回路を構成すれば、内的な目標 ( $z = G(x)$ ) を不完全に満たすだけで、外部から要求される目標  $y = F(x)$  を完全に満たすことができるようになる。

2.2 冗長化の問題点

簡単のために、以上に述べた方針に従って構成した最も単純な回路 (図 2) を使って説明する。

この回路では、外部から要求される出力  $y$  は、0 または 1 の 2 つの値を取る。また写像  $G$  は、 $\hat{z}$  の成分  $\hat{z}_i (i=1, 2, \dots, M)$  ごとに独立な回路  $g_i (i=1, 2, \dots, M)$  を用いて実現する。出力  $y$  が 2 通りの値しか取らないので、それらに対応付く  $z$  は、ハミング距離を最も離そうとすると、例えば、ベクトル  $(0, 0, \dots, 0)^T$ ,  $(1, 1, \dots, 1)^T$  になる。この場合、冗長化は単純な多重化となり、先に述べた誤り訂正手続きは、 $g_i (i=1, 2, \dots, M)$  の出力の多数決を取って  $y$  を決めることに等しくなる。

通常、学習によって  $g_i$  内部のパラメータ  $\omega_i$  を調整しても、要求される目標出力  $z_i$  を完全に正しく実現することはできない。例えば、 $g_i$  内部のパラメータを  $\omega_i$  として、誤差評価関数：

$$\epsilon(\omega_i) = E[|\hat{z}_i - z_i|^2] \tag{1}$$

の形状の一例を図 3 に示す。上式で  $E[ \ ]$  は平均操作を意味している。 $z_i$  は要求される目標出力、 $\hat{z}_i$  は

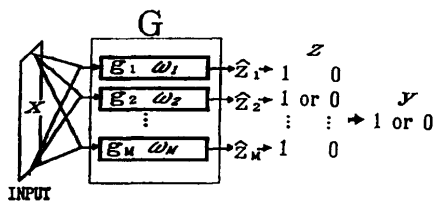


図 2 多重化によって冗長化した回路  
Fig. 2 A redundant structure by multiplication.

$g_i$  の実際の出力である。普通  $\epsilon(\omega_i)$  は多数の極小点を持ち、 $\omega_i$  は、それらの極小点の 1 つに捕まってしまう。また、たとえ最小点が求まったとしても、そこで誤差は 0 とならないのが普通である。

学習を行った結果、各  $g_i$  のパラメータ  $\omega_i$  がすべて同じになってしまったとする。この時、すべての  $g_i$  の出力は常に同一となり、それらの誤りも同時に起こる。これでは、 $\hat{z}_i (i=1, 2, \dots, M)$  の多数決をとっても誤りを訂正することはできない。

2.3 多様性の必要性

多数決によって誤り訂正を行えるようにするためには、各  $g_i$  の誤りは同時に起こってはならない。つまり、各  $g_i$  は同一の出力を出すことを目標として学習を進めるが、学習の結果形成される各  $g_i$  の内部構造は異なり、誤りを独立に生じるようになっていなければならない。

本論文では、各  $g_i$  の内部構造がそれぞれ異なる時、また、多重化に限らず、一般に図 1 の構成の回路で回路  $G$  の出力  $z$  が、各成分独立に誤りを生じる時、 $G$  に多様性があると呼ぶことにする。

従来、バックプロパゲーションの欠点として、学習の結果、パラメータが誤差評価関数の極小点に捕まってしまう、誤差を最小化できないことが指摘され、その改善方法が研究されてきた。しかしながら、構造を多様化するという観点からは、むしろ、各  $g_i$  のパラメータ  $\omega_i$  は、それぞれ異なる極小点に捕まったほうが好ましい。このような発想から、本論文では、従来と全く異なる次のような方針に基づき、高速、高精度な学習方式を導くことを試みることにする。

$g_i$  の出力の誤差は、でたらめに構成した時に比べてある程度小さくなっていけば、最小化されていなくとも構わない。このように内的に設定された目標をある程度正しく満たす  $g_i$  で、内部構造が異なり、したがって誤りを独立に生じるものを、幾らでも多数構成できるならば、内的な目標の冗長性を利用してそれらの出力の誤りを訂正することで、外部から要求される

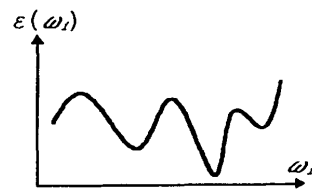


図 3 誤差評価関数の一例  
Fig. 3 An example of the error evaluating function.

目標を幾らでも高い精度で実現することができるはずである。

このような観点からは、これまでの学習研究の「いかにして極小点に捕まらないようにするか」という目標は、「冗長性を持つ目標出力に対して、異なる内部構造を持つ多様な回路を、いかにして多数構成するか」という目標に置き代わることになる。

回路構造を冗長化して、横方向に素子数を増やしても、縦方向の層数が限られているのでは、実現できる写像能力に限界があるのではないかと、この疑問が当然生じてこよう。しかしながら、最近、舟橋によって、4層の回路でも十分な数の素子を用いれば、任意の連続写像を、幾らでも高い精度で実現できることが示されている<sup>6)</sup>。したがって、たとえ回路の層数に制限があっても、横方向に十分な数の素子を用い、さらに学習結果が極小点に捕まる問題を解決すれば、任意の連続写像を実現することができるはずである。本論文で取るアプローチは、まさに、冗長化による補償の機構を通じて極小点問題を克服しながら、回路を横方向に拡張していく具体的方針を与えることになる。

また、層数に制限があると、得られる構造の多様性に限界があるのではないかと、という気がするが、舟橋の結果から、十分な素子を持つ4層の回路によって、幾らでも多数の多様な構造を持つ回路を構成できることがわかる。

### 3. 回路構造を多様化する工夫

前章では、単純な多重化を例として、学習における回路構造の冗長化と多様化の意義を示した。特に多様性を与える工夫をしなくとも、誤差評価関数が図3のような形状をしている時、各  $g_i$  の初期パラメータの値をランダムに与えておけば、学習の結果、 $g_i$  はそれぞれ異なる極小点に陥り、多様性が得られる可能性がある。しかしながら、本章では、多様性の出現をこのような偶然に任せるのではなく、積極的に多様性を生じさせるために、学習手法に導入すべき幾つかの工夫を示すことにする。

多様性を生じるためにまず考え付く方法は、各  $g_i$  の学習手続きの間に相互作用を設けて、多様な構造が得られるように、各  $g_i$  のパラメータの値を制御することである。しかしながら、このような相互作用を設けると、回路は単純なフィードフォワード型ではなくなり、学習アルゴリズムを見つけることが難しくなる。

そこで、本論文では、このような相互作用によらず、多様性を形成する方法として、次に述べる2つの方針を提案する。

#### 3.1 多様化のための工夫1—各 $g_i$ にできるだけ異なる入出力関係を要求する

簡単のために、図4に示す構成の回路を考えることにする。この回路では、内的な目標  $z$  を、その成分  $z_i$  ごとに独立な回路  $g_i$  を用いて実現する。ただし、図2とは異なり、 $z$  は4通りのベクトルを取り、その成分  $z_i$ 、すなわち、各回路  $g_i$  に要求される出力はそれぞれの回路に対し、異なる値となる。つまり単純な多重化ではない。また各  $g_i$  への入力も、 $x$  が、それぞれ異なる  $h_i$  によって変換されるため、異なったものになる。

このように、図4では、各  $g_i$  には、それぞれ異なる入出力関係が要求されて、それを満たすべく内部パラメータ  $\omega_i$  が学習により定められることになる。したがって、すべての  $g_i$  に同一の入出力関係が要求される図2とは異なり、図4の各  $g_i$  の内部構造 ( $\omega_i$ ) はそれぞれ異なったものとなり、誤りも各  $g_i$  独立に生じやすくなると考えられる。

ここで、異なる入出力関係という言葉を使ったが、この言葉を使う場合には注意が必要である。例えば、次のような例を考えてみよう。

2つの回路  $g_i$  と  $g_j$  への入力は同一として、それぞれに要求される出力は、常に反転の関係にあるとする(つまり、一方に0(1)が要求されるときは、他方には1(0)が要求される)。この時、両回路に要求される出力は、その絶対値を比較する限りにおいて大きく異なっている。しかしながら、常に反転の関係にある出力を持つ回路は、一方の回路の最終層の結線強度と閾値の±の符号を反転したものを、もう一方の回路とすることで、容易に構成することができる。この時、2つの回路の出力の間には、常に(誤りが起きている場合にも)反転の関係が生じるため、両回路の誤りは同時に生じるようになってしまう。

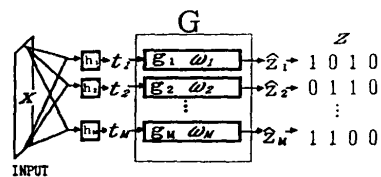


図4 構造多様化の工夫を導入した回路  
Fig. 4 A network structure which yields diversity in learning.

このような問題が生じることを防ぐためには、正確には、上記の議論の中で用いた「異なる入出力関係」を「独立な入出力関係」と言い直さなければならない。上に述べた2つの回路の出力は、絶対値を比べると大きく異なるが、両者の間には、常に「反転の関係」という従属関係があり、独立な入出力関係が要求されていないのである。要求される出力の間に、このような従属関係があると、常に（誤りが起きた場合にも）出力の間にそのような従属関係を持つ回路が形成されてしまい、誤りは独立でなくなってしまうのである。

要求される出力 $z$ の成分の間の独立性の程度は、実験計画法の理論で知られている「強さ」という概念を使って定式化することができる。

$z$ の2成分の間に反転の関係があるということは、その2成分、例えば $z_i, z_j$ が、本来取りうる4通りの組み合わせ：00, 01, 10, 11の中の01, 10の組み合わせしか取らなくなることを意味している。一般に、成分の間に関係があるということ、成分の取る値の組み合わせが特殊なものに限られることとして定義することができる。この定義によれば、2成分があらゆる4通りの値の組み合わせを取りうるならば、その2成分は（その2成分を局所的に見る限りにおいて）独立であると考えられることになる。

実験計画法では、 $z$ のどの $r$ 個の成分も、 $2^r$ 通りのあらゆる値の組み合わせを取るが、ある $r+1$ 個の成分を見ると取りうる値の組み合わせが限られ、従属関係が見つかる時、 $z$ の成分の間に「強さ $r$ 」の独立性があると呼んでいる。

ハミング距離を離すという目的からすると、 $z$ の全集合の中から互いにハミング距離の離れたベクトルのみを選ばなければならないが、 $z$ の独立性の強さを $z$ の長さ $M$ に一致させることはできない。したがって、全成分を同時に見た時に、それらの間に従属関係が生じることは避けられない。学習によって多様な構造を形成するためには、多くの成分を同時に見なければ従属関係が検出できないように、すなわち、できるだけ「強さ」が強くなるように、うまく $z$ を選び、目標出力ベクトルとすることが大切である。

幸いにして、実験計画法の分野で、有限体論を用いてハミング距離と強さの両方を同時に大きくする $z$ の集合の構成方法が確立されており、その理論を本論文の目的にそのまま利用することができる。紙数の都合でその詳細は省くが、この理論は文献5)に詳しく論

じられているので参照されたい。

### 3.2 多様化のための工夫2—入力冗長性の利用

入力パターン $x$ が、 $N$ 次元2値ベクトルであるとする、それは $2^N$ 通りの異なるベクトルを取りうるが、現実の世界で実際に生じるパターンは、その中のごく一部に限られているのが普通である（図5）。例えば、文字パターンを入力する場合、手書き文字は様々に変動し、多様なパターンを取りうるが、それでも意味を持つパターンとなるのは $2^N$ 通りのパターンの中のほんの一部である。この部分集合を $X$ で表すことにする。このように認識対象となる入力が特別なものに限られている時、入力に冗長性があると言う。

入力に冗長性がある場合には、通常出現しえない、したがって認識対象とならないパターンを利用して、構造を多様化することができる。

まず、学習のモードを2つに分け、知識獲得モードと構造多様化モードと呼ぶことにする。知識獲得モードでは、回路は要請される入出力関係の因果関係を学習し、それを反映した構造を形成する。構造多様化モードでは、そうして形成された回路の構造から、重複している部分を捜し出し、それを除いて構造を多様化する。具体的には、それぞれのモードで次のように学習を行えば良い。

知識獲得モードでは、入力パターンと、それに対する目標出力のサンプルを与えながら、例えば、バックプロパゲーションを用いて通常の学習を行う。

これに対して、構造多様化モードは、少し複雑である。まず、簡単のために、図2に示す、多重化によって冗長化した回路で、出力の多数決を取って結果を得る場合について説明する。

各回路 $g_i$ には、学習の結果、同一の出力を出すようになることと、各回路の誤りの特性が異なり、誤りを独立に生じるようになることが同時に要求される。後者の要求を満たすためには、各回路が入力から出力を得るのに用いる手続き、つまり各回路の内部構造ができるだけ異なっていなければならない。これは、一見矛盾する要求である。なぜならば、数学的には、関

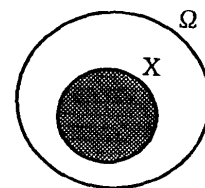


図5 入力パターンの持つ冗長性  
Fig. 5 Redundancy among input patterns.

数は1つのブラックボックスであり, その構造の同一性は, 同一の入力に対して同一の出力を出すかということで判断されるからである. この数学上の定義によれば, 入出力関係を同一にするということは, 構造を同一にすることを意味しているのであり, 上記の目標には, 根本的に両立し得ない矛盾があることになる. しかしながら, 入力に冗長性がある場合には, この矛盾を次のように解決することができる.

上に述べた2つの要求のうち, 回路に同一の入出力関係を要求する方は,  $X$  に属する入力に対してのみ言えば良いのである. したがって, 回路の内部構造の違いは,  $X$  に属さない入力に対して, 各回路の出力がどの程度異なったものになるか, ということによって評価することができる.

この内部構造の違いを計る評価尺度を用いれば, 構造多様化モードでの学習は次のように行えばよい.

まず, 回路を多重化した場合には,  $X$  に属さない入力をわざと与え, それに対して各回路  $g_i$  に異なる出力を要求して学習を行う. この要求を満たそうとすると, 各回路の構造は必然的に異なったものとなり, 多様性が得られる.

また, 単純な多重化でなく, 回路にハミング距離の離れた目標出力が要求される一般の場合には, 次のように学習を行う. 目標出力  $z$  として, ハミング距離の離れたベクトルのみを用いるわけであるから, 必然的に  $z$  の成分間に従属関係が生じてくる. 構造多様化モードでは,  $X$  に属さない入力をわざと与え, その時  $z$  の成分間に, この従属関係が生じないことを要求して学習を行うことで, 多様性が得られる.

以上に述べた方法で,  $X$  に属さない入力を与えて学習を行うモードは, 形式上, ボルツマンマシンで用いられる unlearning のモードと類似したものとなっている. 本論文の手法では, unlearning のモードは回路構造を多様化する働きをしている.

## 4. 実験

### 4.1 実験方法

以上までの議論によって得られた予想を, 簡単な実験によって確認する. 本章の実験を通じて, 回路構造の冗長化と多様化に基づく, 新しい学習アプローチの有効性を明らかにする.

回路への入力は,  $16 \times 16 = 256$  個の画素から成る濃淡画像である. ETL8 の手書き教育漢字データベースから75種(カテゴリ)の手書き平仮名を選び出す.

このデータベースでは, 各カテゴリは128個のサンプルを持つが, 1つのサンプルは,  $64 \times 63$  画素の2値画像となっているので, 画面を  $16 \times 16$  個の格子状の領域に分け, それぞれの領域の黒画素の密度を入力画像の一画素の値とする.

回路は, 図4に示す構成とし, その中の1つの  $g_i$  として図6に示す3層のフィードフォワード型回路を用いる. また, 図4の回路では最初の段階で入力  $x$  を変換し(図中の  $h_i$ ), 各  $g_i$  に異なる入力が与えられるようにしているが, 以下の実験では, この変換を次のように行う.

まず, 一様乱数で,  $[-1, 1]$  の範囲の数をランダムに発生して, 行列  $H_i$  の  $256 \times 256$  個の成分を定める. 次に, 入力  $x$  に行列  $H_i$  を作用させて,

$$t_i = H_i x, \quad i = 1, 2, \dots, M$$

により256次元のベクトル  $t_i$  を得て, 各  $g_i$  への入力とする.

目標出力  $z$  は, ハミング距離が互いにできるだけ離れ, なおかつ成分間の独立性の程度が強さ2以上となるように注意して定める. また, 従来使われてきた認識細胞的表現, すなわち,  $z$  のどの成分の値が最大となるかということでカテゴリ名を表す方式との比較を行う. 認識細胞的表現では, 目標出力  $z$  はカテゴリ名に対応する成分のみが1で他の成分は0となっているため, ハミング距離は互いに2しか離れていない. 認識細胞的表現に対しても, 回路の構成は図4とし, 個々の  $g_i$  には図6の回路を用いる. いずれも, 学習アルゴリズムとしてバックプロパゲーションを用いる. したがって, 両者の違いは, 出力層の素子数と目標出力  $z$  だけになる.

なお, 今回の実験では, 学習に, 3.2節に述べた構造多様化モードは導入しない.

バックプロパゲーションを用いると出力はアナログ値となる. 誤り訂正は, 各目標出力  $z$  (カテゴリ名を

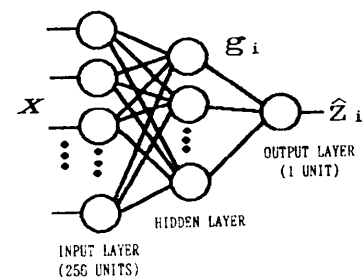


図6 回路  $g_i$  の構造  
Fig. 6 Structure of the network  $g_i$ .

表すベクトル) と回路の実際の出力  $\hat{z}$  との間で、

$$D(\hat{z}, z) = (\hat{z} - z)'(\hat{z} - z)$$

により距離を計算し、実際の回路の出力に対して、最も小さな距離を与える目標出力が表すカテゴリ名を認識結果とする。ただし2位の結果との距離の差が小さい時は、結果が曖昧であると考え、認識誤りとしてカウントする。これと比較する認識細胞的表現では、出力ベクトルの中で最大値を取る成分が表すカテゴリ名を認識結果とし、2位との差が小さい時は、やはり認識誤りとしてカウントする。

4.2 実験結果

以下に示す実験条件のもとで、認識細胞的表現で従来どおりの学習を行った場合と、今回提案した、回路構造を多様化、冗長化する工夫を導入した場合とで、学習の速度と精度を比較した。各グラフは、横軸を学習時のサンプル提示回数とし、その各時点で、過去500 (1,000) 回の試行中、認識誤りが何回生じているかということを示したものである。ただし、学習を始めたばかりで、過去の誤りの累積のない時点0では、過去500 (1,000) 回すべて誤っていると、誤り回数を計算している。

通信では、冗長化する単位となるコード長を長くするほど、冗長化の効果は顕著に現れ、また符号化の効率も良くなるから、回路構造の冗長化もカテゴリ数が増えるほど、効率と効果が高まると予想される。そこで、以下、文字カテゴリ数をいろいろ変えて実験を行ってみた。

以下の実験では、図6の  $q_i$  の hidden layer の素子数はすべて8としている。したがって、回路の構成で変化するのは、図4中の回路  $q_i$  の数だけである。

【実験1】

認識対象とするカテゴリ数を平仮名「あ」、「い」、「う」、「え」の4種とする。回路の構成(図4中の回路  $q_i$  の数  $M$ ) は次のように定める。

|                  | 冗長化した表現 | 認識細胞的表現 |
|------------------|---------|---------|
| $q_i$ の数 ( $M$ ) | 14      | 4       |

各カテゴリ128個のサンプルをすべて学習に用い、それを繰り返し与えて学習を進めながら、各時点で過去500回に生じた誤りの回数を図7に示す。

両者の結果に大きな違いはないが、5,000回学習した時点で、冗長化した表現では誤りが0となるのに対して、認識細胞的表現では0とならない。ただし、両表現で使われる回路の規模は異なり、前者では、14/4倍の素子が使われている点に注意する必要がある。

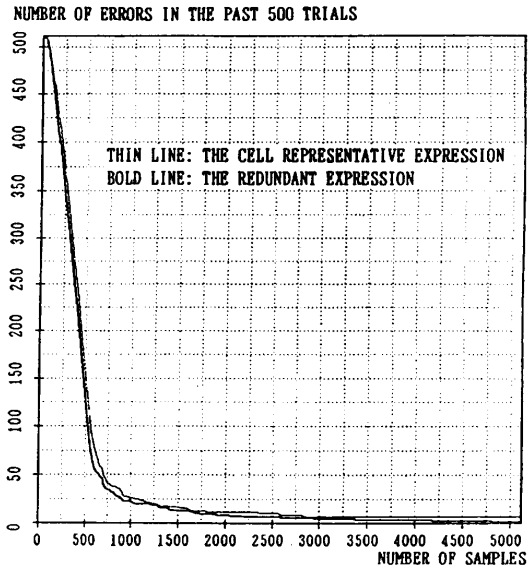


図7 実験1の結果

細線は、認識細胞的表現、太線は冗長化した表現に対する結果を示す。

Fig. 7 Result of the experiment No. 1.

The thin line shows the number of errors by the representative cell expression and the bold line shows the number of errors by the redundant expression.

【実験2】

認識対象とするカテゴリ数を15に増やし、回路の構成は次のように定める。

|                  | 冗長化した表現 | 認識細胞的表現 |
|------------------|---------|---------|
| $q_i$ の数 ( $M$ ) | 15      | 15      |

各カテゴリ128個のサンプルをすべて学習に用い、それを繰り返し与えて学習を進めながら、過去1,000回に生じた誤りの回数が減少する様子を図8に示す。

認識細胞的表現では誤りがいつまでも残ってしまうが、表現を冗長化した場合には、16,000回サンプルを与えたところから誤りはほとんど0となる。この場合、両結果を得るのに使われた回路の規模(素子数)は同じである。

【実験3】

認識対象とするカテゴリ数を31に増やし、回路の構成を次のように定める。

|                  | 冗長化した表現 | 認識細胞的表現 |
|------------------|---------|---------|
| $q_i$ の数 ( $M$ ) | 31      | 31      |

後で汎化の効果を調べるため、各カテゴリで用意されている128個のサンプル中の64個を学習に用いる。過去1,000回に生じた誤りの回数が減少する様子を図

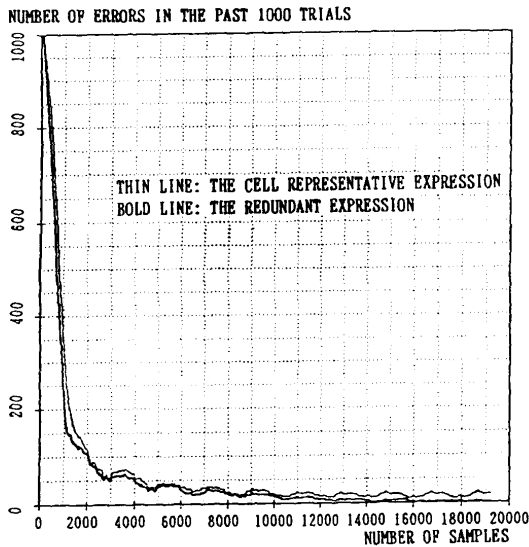


図 8 実験 2 の結果

細線は、認識細胞的表現、太線は冗長化した表現に対する結果を示す。

Fig. 8 Result of the experiment No. 2.

The thin line shows the number of errors by the representative cell expression and the bold line shows the number of errors by the redundant expression.

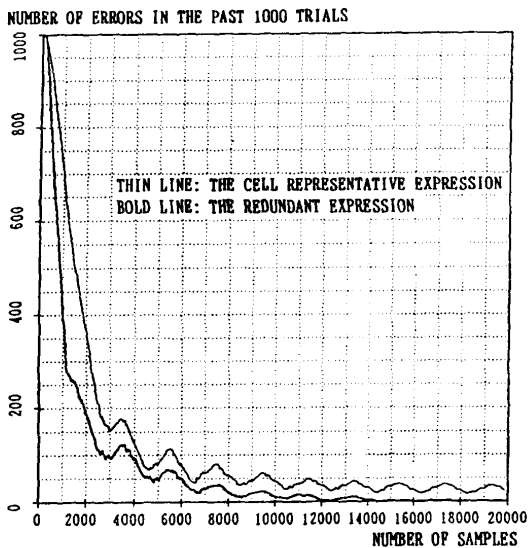


図 9 実験 3 の結果

細線は、認識細胞的表現、太線は冗長化した表現に対する結果を示す。

Fig. 9 Result of the experiment No. 3.

The thin line shows the number of errors by the representative cell expression and the bold line shows the number of errors by the redundant expression.

9 に示す。

結果を見ると、16,000 回学習した時点で、認識細胞的表現では多数の誤りが残るが、表現に冗長化した場合には、誤りはほとんど 0 となっている。両結果は、同じ規模の回路で得られたものである。

【実験 4】

全カテゴリ (75 カテゴリ) を認識対象とする。回路の構成は次のように定める。

|                  | 冗長化した表現 | 認識細胞的表現 |
|------------------|---------|---------|
| $g_i$ の数 ( $M$ ) | 63      | 75      |

後で汎化の効果を調べるため、やはり 128 個のサンプル中の 64 個を学習に用いる。過去 1,000 回に生じた誤りの回数が減少する様子を図 10 に示す。

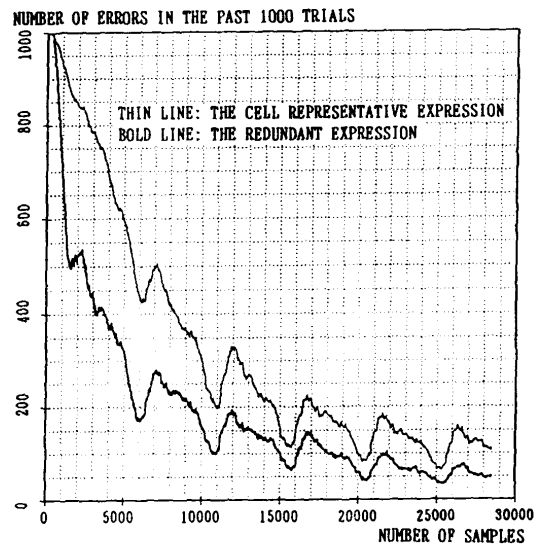


図 10 実験 4 の結果

細線は、認識細胞的表現、太線は冗長化した表現に対する結果を示す。

Fig. 10 Result of the experiment No. 4.

The thin line shows the number of errors by the representative cell expression and the bold line shows the number of errors by the redundant expression.

表 1 実験 5 の結果

未学習文字に対する認識率

Table 1 Result of the experiment No. 5.

Recognition error rates of the representative cell expression and the redundant expression for the unlearning pattern set.

|         | 学習文字  | 未学習文字 |
|---------|-------|-------|
| 認識細胞的表現 | 87.2% | 66.3% |
| 冗長化した表現 | 96.2% | 77.6% |



表現を冗長化した場合、認識細胞的表現に比べ、学習精度、速度とも改善されることがわかる。冗長化した表現では、回路の規模がむしろ小さくなっている点に注意して欲しい。

#### 【実験5】

実験4で形成した回路で、未学習文字（各カテゴリ64サンプル）の認識率を調べ、汎化の能力を確認した。結果を表1に示す。冗長化した表現では汎化の能力も高まっていることがわかる。

## 5. おわりに

本論文では、ネットワークの構造の冗長化と多様化という、従来と異なる観点から、高性能な学習方式を導くことを試みた。提案した手法は、次のような発想に基づいている。

冗長性を持つ学習の目標を内的に設定し、それが不完全に満たされるだけで、外部から要求された目標が完全に満たされるようにしておく。そうすれば、学習の結果が不完全であっても、誤り訂正を行うことにより、外部からの要求を正しく実現することができる。ただし、出力の各成分に誤りが同時に生じてしまったら、誤り訂正を行うことができない。したがって、冗長な結果を、異なる方法、異なる手続きで実現する多様な構造の回路を形成し、誤りが独立に生じるようにしなければならない。

本論文では、まず、冗長に設定された内的な目標と誤り訂正の機構を、一種の pre-wiring として回路に与え、多様な構造の回路を形成するために、次の2つの方法を提案した。

(1) 各回路に要求される入出力関係ができるだけ異なったものとなるように学習の目標を工夫して与える。

(2) Unlearning のモードを設け、認識対象外となるパターンを活用して回路の構造を多様化する。

最後に簡単な実験を行い、本論文に示した手法により、従来用いられていた認識細胞的表現に比べて、小さい規模の回路で、優れた学習速度、学習精度が得られることを確認した。1つの実験結果から一概に結論

を得るのは危険であるが、これにより、本論文で提案したアプローチの可能性を示すことができた。

今後の課題としては、多様性を生じるメカニズムをより明確にすること、汎化能力を理論的に検討すること、様々な課題に対して本アプローチの有効性を確認することなどがある。

**謝辞** 本研究に対して、貴重なご助言を頂いた本学小川英光教授、また実験に協力して頂いた本学大学院生松田哲君、C. ポーンシャイ君に感謝いたします。

## 参 考 文 献

- 1) Rumelhart, D. E., McClelland, J. L. and PDP Research Group: *Parallel Distributed Processing*, Vol. 1, Vol. 2, MIT Press, Cambridge, MA (1986).
- 2) Amari, S.: A Theory of Adaptive Pattern Classifiers, *IEEE Trans.*, Vol. EC-16, No. 3, pp. 299-307 (1967).
- 3) Bruck, J. and Blaum, M.: *Neural Networks, Error-Correcting Codes and Polynomials over the Binary N-Cube*, IBM Research Report (1987).
- 4) 熊沢逸夫, 小川英光: ハミング距離の離れたアトラクタを用いた拘束条件下でのラベル付けと表現の冗長性を利用した温度制御の自動化, 信学技報, MBE 87-121 (1988).
- 5) 高橋磐郎: 組合わせ理論とその応用, 岩波書店 (1979).
- 6) 舟橋賢一: ニューラルネットワークの capability について, 信学技報, MBE 88-52, pp. 124-134 (1988).

(平成元年8月31日受付)

(平成2年4月17日採録)



熊沢 逸夫 (正会員)

昭和56年東京工業大学電気電子卒業。昭和61年同大学大学院博士課程修了。同年同大工学部情報工学科助手。平成2年助教授。パターン認識、信号処理、ニューラルネットワークの研究に従事。工学博士。電子メールアドレス: kumazawa@cs. titech. ac. jp