

パターン認識のためのクラスタ分析に基づく3層ニューラルネットワークの重み初期値の一設定法

正員 市村 直幸<sup>†</sup>      非会員 竹内 俱佳<sup>†</sup>      正員 永井 豊<sup>†</sup>

Initial Weight Values Setting of Three Layered Neural Network Based on Cluster Analysis for Pattern Recognition

Naoyuki ICHIMURA<sup>†</sup>, *Member*, Tomoyoshi TAKEUCHI<sup>†</sup>, *Nonmember* and Yutaka NAGAI<sup>†</sup>, *Member*

あらまし 本論文では、パターン認識を行う3層ニューラルネットワークの中間層の働きを、学習パターンに含まれるクラスタごとの判別分析とみなし、与えられた学習パターンのクラスタ分析の結果に基づいて、重みの初期値を設定する手法を提案する。提案する手法は、(1)学習パターンの直和解によるクラスタ分析、および、(2)各クラスタに属するパターンの、そのクラスタに対応する内部表現(代表ベクトル)への集塊化の2段階の処理からなる。クラスタ分析の結果の評価基準である平均ひずみ測度と、各カテゴリーのクラスタ数との関連を調べ、各カテゴリーのクラスタ数(中間層のユニット数)の設定法についても考察する。分布の単峰性を仮定できないパターンに対する数値実験を行った結果、提案手法は、乱数で重みの初期値を設定した場合に比べ、学習の進行を促進し、同一の学習回数で高い認識率を得られること、また、学習係数の変化に対し頑健であるという性質をもつことを確認した。

キーワード 3層ニューラルネットワーク、クラスタ分析、内部表現、判別分析、パターン認識

1. まえがき

非線形ユニットによって構成されるフィードフォワード型ニューラルネットワークは、画像や音声の認識を始め、広くパターン認識に応用されている。特に、3層ニューラルネットワーク(以下、3層 NN)は、単独でも、また、大規模ニューラルネットワークの構成要素としても、最もよく利用されている。

3層 NN によってパターン認識機構を構成する際には、学習の前にユニット間の重みの初期値を設定する必要がある。通常、この重みは乱数によって設定されるが、より適切な重みの初期値の設定については、シグモイド関数の活性領域という観点<sup>(1)</sup>や3層 NN 以外のパターン認識法との関連<sup>(2)</sup>から議論がなされている。しかし、それら以外に、3層 NN の構造と密接な関連があると思われる、学習後の中間層におけるデータ表

現(以下、内部表現)からの観点も考えられる。すなわち、どのような内部表現がパターン認識において有効であるかを考察し、その内部表現を実現するのに適した重みを設定すると、学習の進行の促進効果が期待できる。

パターン認識を行う3層 NN の内部表現は、線形ユニットを用いた場合には線形判別分析を、非線形ユニットを用いた場合には非線形判別分析を、学習パターンに対して行った結果と本質的に同一であることが示されている<sup>(3),(4)</sup>。しかし、パターンの分布が単峰性を仮定できない複雑な分布の場合の内部表現については明らかではない<sup>(4)</sup>。

本論文では、パターン認識を行う3層 NN を対象とし、中間層の働きを学習パターンに含まれるクラスタごとの判別分析、つまり、(1)学習パターンの直和解によるクラスタ分析、および、(2)各クラスタに属するパターンの、そのクラスタに対応する内部表現(代表ベクトル)への集塊化、とみなし、与えられた学習パターンのクラスタ分析の結果に基づいて、重みの初期値を

<sup>†</sup> 電気通信大学電子工学科, 調布市  
Faculty of Electro-Communications, University of Electro-Communications, Chofu-shi, 182 Japan

設定する手法を提案する。更に、クラスタ分析の結果の評価基準である平均ひずみ測度と各カテゴリーのクラスタ数との関連を調べ、クラスタ数の設定についても考察する。この際、パターンの分布の単峰性を仮定しない。

以下、2.において3層 NN の定式化、および、それによってパターン認識機構を構成した際の中間層の働きについて述べる。次に、3.において、学習パターンのクラスタ分析に基づく重みの初期値の設定について述べる。そして、4.において、提案した手法の有効性を数値シミュレーションによって確認すると共に、各カテゴリーのクラスタ数の設定についても議論する。

## 2. パターン認識を行う3層 NN

### 2.1 3層 NN の定式化

以下の議論に必要な記号を定義する。

$C_i$ : カテゴリー ( $i=1, \dots, m$ )

$N_i$ :  $i$  番目のカテゴリーの学習パターン数

$N = \sum_{i=1}^m N_i$ : 全学習パターン数

$X = [x_1^1, \dots, x_{N_1}^1, \dots, x_1^m, \dots, x_{N_m}^m]$ : パターンベクトル  $x_j \in R^n$  を要素とする  $n \times N$  行列

$B = [b_1^1, \dots, b_{N_1}^1, \dots, b_1^m, \dots, b_{N_m}^m]$ :  $x_j^i$  に対応する内部表現  $b_j^i \in R^p$  を要素とする  $p \times N$  行列

$Y = [y_1^1, \dots, y_{N_1}^1, \dots, y_1^m, \dots, y_{N_m}^m]$ :  $x_j^i$  に対応する出力ベクトル  $y_j^i \in R^m$  を要素とする  $m \times N$  行列

$T = [t_1^1, \dots, t_{N_1}^1, \dots, t_1^m, \dots, t_{N_m}^m]$ :  $x_j^i$  に対応する教師信号  $t_j^i \in R^m$  を要素とする  $m \times N$  行列

但し、 $i=1, \dots, m, j=1, \dots, N_i$

$u = [1, 1, \dots, 1]^t \in R^n$

$X^* = \begin{bmatrix} X \\ u^t \end{bmatrix}$ :  $(n+1) \times N$  行列

$B^* = \begin{bmatrix} B \\ u^t \end{bmatrix}$ :  $(p+1) \times N$  行列

$W_1$ : 入力層から中間層への結合重み ( $p \times (n+1)$  行列)

$W_2$ : 中間層から出力層への結合重み ( $m \times (p+1)$  行列)

$f(x) = 1/(1 + \exp(-x))$ : ユニットの出力関数

各ユニットのしきい値は、 $X^*$ ,  $B^*$  を用いることによって、常に1を出力するユニットとの間の重みとして表現する。また、教師信号は各カテゴリーに対応した単位ベクトル (例えば、2 カテゴリーの場合は、 $(0, 1)$  と  $(1, 0)$ ) とする。

以上の記号を用いると、入力層から中間層における処理は、次式で表される。

$$B = f(W_1 X^*) \quad (1)$$

ここで、 $f$  は行列の各要素ごとに作用するものとする。また、中間層から出力層における処理は、次式で表される。

$$Y = f(W_2 B^*) \quad (2)$$

### 2.2 双峰性をもつパターンに対する数値実験

3層 NN の内部表現を観察するために、次に示す平均と共分散をもつ正規乱数を用いて実験を行った。

[カテゴリー1]: 200 サンプル, unimodal Gaussian

$$m = [0.3, 0.5]^t \quad (3)$$

$$\Sigma = \begin{bmatrix} 0.005 & 0.001 \\ 0.001 & 0.005 \end{bmatrix} \quad (4)$$

[カテゴリー2]: 200 サンプル, bimodal Gaussian

モード1: 100 サンプル

$$m_1 = [0.5, 0.7]^t \quad (5)$$

$$\Sigma_1 = \begin{bmatrix} 0.005 & -0.004 \\ -0.004 & 0.005 \end{bmatrix} \quad (6)$$

モード2: 100 サンプル

$$m_2 = [0.5, 0.3]^t \quad (7)$$

$$\Sigma_2 = \begin{bmatrix} 0.005 & 0.004 \\ 0.004 & 0.005 \end{bmatrix} \quad (8)$$

実験に用いた学習パターンの散布図を図1に示す。この図より、明らかにパターンの分布の単峰性は仮定できない。3層 NN の各層のユニット数を2-1-2, 2-2-2として実験を行った結果の学習曲線と内部表現を図2~4に示す。判別分析を行うと、内部表現(特徴)空間

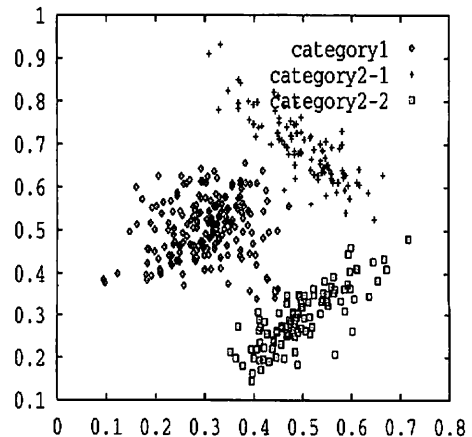


図1 双峰性をもつ学習パターンの散布図  
Fig. 1 A scatter diagram of learning patterns that have bimodal distribution.

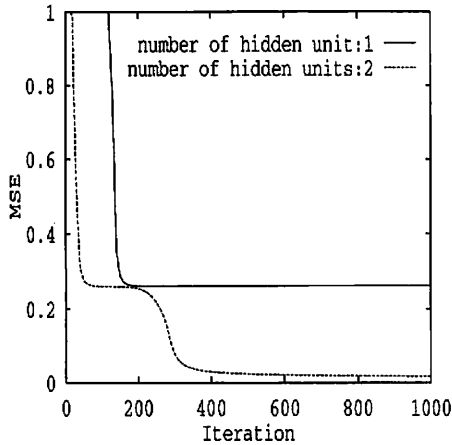


図2 図1のパターンに対する学習曲線  
Fig. 2 Learning curves for patterns of Fig. 1.

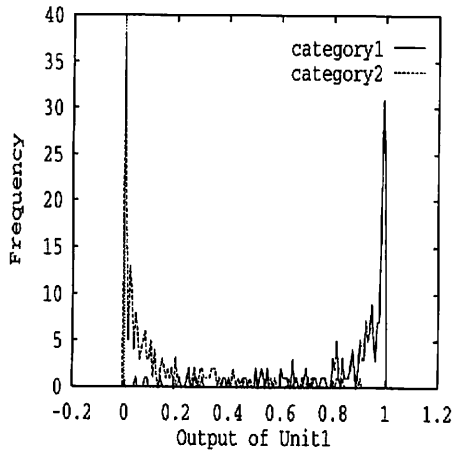


図3 図1のパターンに対する内部表現(中間層のユニット数: 1)  
Fig. 3 Internal representation for patterns of Fig. 1 (Number of hidden unit : 1).

の次元はカテゴリ数から1を引いた1となる。しかし、中間層のユニット数が一つの場合は、図3よりわかるように、二つのカテゴリを分離するには不十分な内部表現しか生成されず、学習時の平均2乗誤差も約0.26より減少しないことが図2よりわかる。

中間層のユニットが二つの場合には、図4よりわかるように、カテゴリ2のパターンに対する内部表現は、ほぼ(0,0)、(1,1)の2点に集中しており、カテゴリ2のもつ二つのモードに対応した内部表現が生成されていることがわかる。このように、パターンの分布のモード一つ一つに対し内部表現が生成され、それらの内部表現に、それぞれのモードに属するパターン

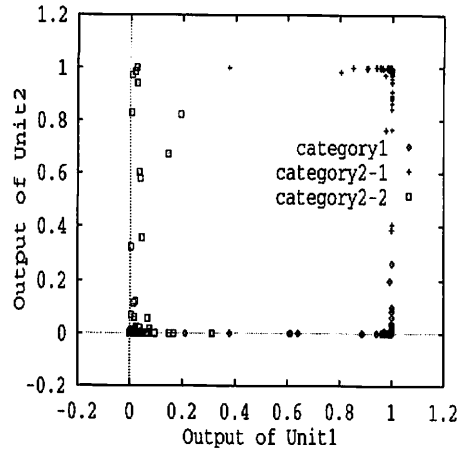


図4 図1のパターンに対する内部表現(中間層のユニット数: 2)  
Fig. 4 Internal representation for patterns of Fig. 1 (Number of hidden units : 2).

の集塊化が行われている。そのことによって、二つのカテゴリの分離が中間層のユニットが一つの場合と比較して明確になり、学習時の平均2乗誤差もより小さい値となる(図2)。非線形判別分析においても、各カテゴリに割り当てられた代表ベクトルへのパターンの集塊化が基本とされる。しかし、3層NNでは、二つのモードを含むカテゴリに対しても、一つの教師信号しか与えていないのに、中間層では全体としてあたかも三つのカテゴリがあるかのように振る舞っている点が異なっている。

また、図3, 4どちらの場合にも共通していることは、シグモイド関数の有界性によって、内部表現は2値表現に近くなることである。

### 3. クラスタ分析による3層NNの構成

2.の数値実験の結果から、パターン認識を行う3層NNの中間層の働きを、学習パターンに含まれるクラスタごとの判別分析、つまり、

- (1) 学習パターンの直和解によるクラスタ分析
- (2) 各クラスタに含まれるパターンの、そのクラスタに対応する内部表現(代表ベクトル)への集塊化とみなし、3層NNの重みの初期値を設定する手法を提案する。

#### 3.1 学習パターンの直和解

カテゴリ  $C_i$  の学習パターンの集合を  $V_i$  とする。

$$V_i = \{\mathbf{x}_j\} \quad (9)$$

$$i=1, \dots, m, \quad j=1, \dots, N_i$$

$V_i$  中の二つの要素の類似度をユークリッド距離  $d(x_j^i, x_k^i) = \|x_j^i - x_k^i\|$  で計る。このユークリッド距離  $d$  に基づいて、 $V_i$  を  $M_i$  個のクラスタ  $\{S_i^1, S_i^2, \dots, S_i^{M_i}\}$  ( $i = 1, \dots, m$ ) に直和分解する。直和分解を見出すクラスタリングの手法は、いくつか存在するが、ここではベクトル量子化の手法である LBG アルゴリズム<sup>(5),(6)</sup>を用いる。LBG アルゴリズムによって得られるカテゴリー  $C_i$  のコードブックを  $\{c_j^i\}$  ( $j=1, \dots, M_i$ ) とし、 $V_i$  の部分集合

$$V_i(c_j^i, c_k^i) = \{y \in V_i \mid d(y, c_j^i) \leq d(y, c_k^i)\} \quad (10)$$

を用いると、ほかのどのコードよりも、コード  $c_j^i$  に類似した  $V_i$  の要素は、

$$V_i(c_j^i) = V_i(c_j^i, c_i) \cap \dots \cap V_i(c_j^i, c_{M_i}^i) \quad (11)$$

と表すことができる。求めるクラスタ  $S_j^i$  を

$$S_j^i = V_i(c_j^i) \quad (12)$$

とする。

### 3.2 各クラスタへの内部表現の割当て

得られたクラスタ  $S_j^i$  に対する代表ベクトルとして、一つの  $p$  次元ベクトル  $z_j^i$  ( $i=1, \dots, m, j=1, \dots, M_i$ ) を割り当てる。 $z_j^i$  は、異なるクラスタに同一のものを割り当てない限り、基本的にはどのようなものでもよい。ここでは、 $z_j^i$  は 2 値符号とし、2.2 の数値実験に見られた内部表現のように、ユニットの出力関数であるシグモイド関数の飽和領域を用いて表現する。また、更に、次の関係を満たすものとする。

$$z_j^i < z_k^i \Leftrightarrow i \leq k \text{ かつ } j < l \quad (13)$$

上式の不等号は、 $z_j^i$  を  $p$  [bit] の 2 進数とした場合の大小関係を表すものとする。よって、この条件は、全クラスタに対して、 $p$  [bit] の 2 進数で表現された 0 から  $M$  ( $= \sum_{i=1}^m M_i$ ) までの数を小さい順番に割り当て、それを  $z_j^i$  とすることを意味する。このような条件を満たすように、すべてのクラスタに異なる代表ベクトルを割り当てることによって、内部表現は線形分離可能なものとなる ( $p \leq 8$  まで確認している。付録参照)。この  $z_j^i$  を用いて、内部表現  $b_j^i$  を次のように設定する。

$$b_j^i = z_k^i \text{ if } x_j^i \in S_i^k \quad (14)$$

$$i=1, \dots, m, j=1, \dots, N_i, k=1, \dots, M_i$$

内部表現を 2 値符号とすることから、全学習パターンが  $M$  個のクラスタに分割されている場合の中間層のユニット数  $p$  は、

$$2^p \geq M \quad (15)$$

を満たすものでなければならない。

### 3.3 重みの初期値の設定

各クラスタに含まれるパターンが、そのクラスタに

対応する内部表現へ集塊するように 3 層 NN の初期値を設定する。3.2 までの過程で与えられる、各学習パターンに対応する内部表現および教師信号を用いると、式 (1) および式 (2) より、

$$f^{-1}(B) = W_1 X^* \quad (16)$$

$$f^{-1}(T) = W_2 B^* \quad (17)$$

となる。ここで、重みの初期値として

$$J_1 = \|f^{-1}(B) - W_1 X^*\|^2 \quad (18)$$

$$J_2 = \|f^{-1}(T) - W_2 B^*\|^2 \quad (19)$$

を最小にするものを用いる。但し、 $f^{-1}$  はユニットの出力関数  $f$  の逆関数であり、 $\|\cdot\|$  はユークリッドノルムを表す。 $J_1, J_2$  を最小とする重みを  $W_{1i}, W_{2i}$  とすると、

$$W_{1i} = f^{-1}(B)(X^*)^+ \quad (20)$$

$$W_{2i} = f^{-1}(T)(B^*)^+ \quad (21)$$

で与えられる。ここで、 $(X^*)^+, (B^*)^+$  はそれぞれ、 $X^*, B^*$  のムーア・ペンローズの一般逆行列<sup>(7)</sup> である。

$f^{-1}(B)$  および  $f^{-1}(T)$  は  $\epsilon$  を小さな正の定数として、 $b_j^i$  および  $t_j^i$  の要素が 0, 1 それぞれの場合に対して、

$$f^{-1}(\epsilon) = \ln \frac{\epsilon}{1-\epsilon} \quad (22)$$

$$f^{-1}(1-\epsilon) = \ln \frac{1-\epsilon}{\epsilon} \quad (23)$$

を対応させることによって構成する。

以上の手続きで、学習パターンのクラスタ分析の結果から、重みの初期値を決定することができる。また、中間層のユニット数は、各カテゴリーの学習パターンをいくつのクラスタに分割するかによって決まる。4. において、実際に本手法を適用してパターン認識を行う 3 層 NN を構成し、各カテゴリーのクラスタ数の決定問題も含め、検討を行う。

## 4. 数値実験

3. で提案した手法を、2.2 で用いたパターンと野外シーンを撮像した画像から得られるパターンに対して適用した結果を示す。また、クラスタ分析の結果の評価基準としての平均ひずみ測度と各カテゴリーのクラスタ数との関連を調べ、各カテゴリーのクラスタ数の設定についても考察を行う。

### 4.1 人工パターンの認識

#### 4.1.1 学習結果

図 1 に示す二つのカテゴリーのパターン認識を行う 3 層 NN を、提案手法で構成した。ここでは、まず、各カテゴリーにおけるクラスタ数を、カテゴリー 1 は 1 個、カテゴリー 2 は 2 個とした結果を示す。クラス

タ数は合計で3個であるので、 $2^p \geq 3$ を満たす  $p$  を選ぶ。実験では、この関係を満たす最小の値をとり、 $p=2$  とした。よって、3層 NN の構造は、2-2-2 となる。重みの初期値は、式(20)および式(21)に従って設定した。また、予備実験より、誤差逆伝搬学習則<sup>(9)</sup>における学習係数は0.3、慣性係数は0.7とし、式(22)および式(23)の  $\epsilon$  は  $10^{-3}$  とした。

学習パターンを1,000回提示した結果の提案手法による学習曲線、および、重みの初期値を $\pm 0.3$ の一様乱数で与え、独立に20回学習を行った結果の学習曲線を図5に示す。

乱数で重みの初期値を設定した場合には、ほとんどの場合、平均2乗誤差が0.26付近で一時的に減少しなくなっている。これは、ローカルミニマムの影響であると考えられる。一方、提案手法では学習開始時の誤差が非常に小さく、また、ローカルミニマムの影響を

受けることなしに誤差が減少している。このことから、乱数で重みの初期値を設定した場合に比べ、提案手法は誤差の収束性に優れていることがわかる。

次に、重みの初期値を設定した時点、および、1,000回学習を行った後の3層 NN による認識率を調べた。評価パターンの認識率は、学習パターンと同一の分布に従うパターンを20個発生させ、平均をとった。乱数で重みの初期値を設定した場合は、初期値に依存性があるため、20回学習して、それぞれの場合の認識率を平均した。提案手法では、20個の評価パターンに対する平均のみをとった。結果を表1に示す。表1より、提案手法で重みの初期値を設定した方が、初期値を設定した時点で高い認識率を示し、また、誤差の収束性に優れていることから、同一の学習回数で高い認識率が得られることがわかる。

ここで用いた人工パターンのように、クラスタ化が有効であることがわかっている問題では、重みの初期値を設定した時点で高い認識率を示すが、学習による更なる向上が見られる。これは重みの初期値の設定のためのクラスタリングが、各カテゴリーごとに行われ、ほかのカテゴリーの影響を考慮していないという点が、学習を通じて改善されているからであると考えられる。特に、ここで用いた人工パターンのように、カテゴリー間に一部オーバーラップしている部分があるような場合には、ほかのカテゴリーのパターンの影響を考慮することは重要である。

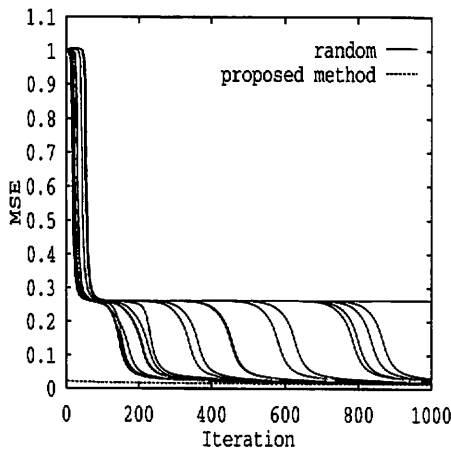


図5 提案手法による図1のパターンに対する学習曲線  
Fig. 5 Learning curves of three layered neural network that initial weights were set by proposed method for patterns of Fig. 1.

表1 図1のパターンに対する認識率 [%]

Iteration	Initial setting	Learning set	Test set
0	Proposed method	98.00	98.68
	Random	50.00	50.00
1000	Proposed method	99.50	99.34
	Random	98.55	98.53

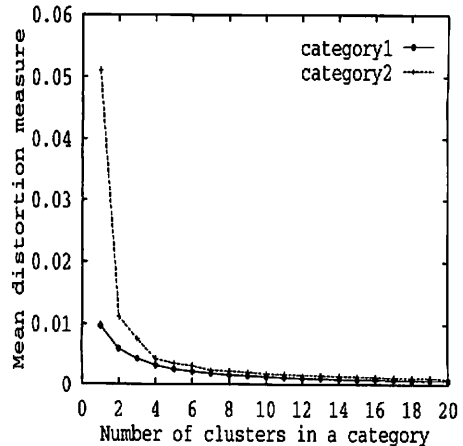


図6 図1のパターンに対するクラスタ数による平均ひずみ測度の変化

Fig. 6 Changes of mean distortion measure as a function of the number of clusters in a category for patterns of Fig. 1.

#### 4.1.2 クラスタ数と平均ひずみ測度

クラスタ分析に用いた LBG アルゴリズムによって得られるクラスタの評価基準として、次の平均ひずみ測度  $D_i$  を用いる。

$$D_i = \frac{1}{N_i} \sum_{j=1}^{M_i} \sum_{k=1}^{|S_j^i|} \|c_j - x_k\|^2 \quad x_k \in S_j^i \quad (24)$$

$$i=1, \dots, m$$

但し、 $|S_j^i|$  は、クラスタ  $S_j^i$  に含まれるパターン数を表すものとする。 $D_i$  は、与えられたクラスタ数  $M_i$  で、どの程度カテゴリ  $C_i$  のパターンの分布を近似しているかの目安となる。

図 6 に、図 1 に示すパターンの、各カテゴリのクラスタ数に対する平均ひずみ測度の変化を示す。この図より、平均ひずみ測度は、クラスタを増やすほど小さくなっていくが、ある程度で飽和し、それ以上クラスタ数を増やしても、平均ひずみ測度の減少には大きく貢献しないことがわかる。

中間層のユニット数が 2 個の場合、二つのカテゴリのクラスタ数の組合せとして、(1,2), (2,1), (2,2), (1,3), (3,1) が考えられる。これらすべての場合について学習を行い、クラスタ数の変化の学習への影響を調べた結果を図 7 に示す。この図より、パターンの分布が複雑なカテゴリ 2 の平均ひずみ測度が小さくなる (1,2), (2,2), (1,3) の組合せでは、学習初期の誤差の減少量が大きく、特に、(1,2), (1,3) では、1,000 回の学習で誤差がかなり小さくなっていることがわかる。このことから、平均ひずみ測度を飽和させるために多

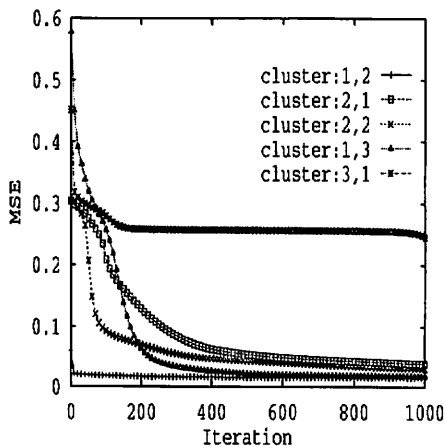


図 7 図 1 のパターンに対するクラスタ数による学習曲線の変化

Fig. 7 Changes of learning curve as a function of the number of clusters in a category for patterns of Fig. 1.

くのクラスタを必要とするカテゴリに、多くのクラスタを割り当てることは、学習の進行を促進すると考える。しかし、(1,3) と (1,2) を比較すると、カテゴリ 2 の平均ひずみ測度が大きくなる (1,2) の方が初期の誤差が非常に小さく、また、非常に少ない学習回数で誤差が小さくなる。このことから、不要なクラスタは、学習の進行を妨げていると考えられる。実際に、クラスタ数をどのように与えても、学習後の内部表現は、ほぼ図 4 のようになることが確認できた。

#### 4.1.3 クラスタ数の決定

この実験では、パターンの分布が既知であるので、各カテゴリの分布の各モードに一つのクラスタを割り当て、中間層のユニットを 2 個とした。通常の場合、パターンの分布は未知であるが、上記の平均ひずみ測度と学習の進行の関係は同様であると考えられる。

そこで、パターンの分布が複雑と思われるカテゴリに対し、多くのクラスタを割り当て、かつ、できるだけ不必要なクラスタを生成させないようにするために、各カテゴリの平均ひずみ測度のある値にそろえ、その値を実現するのに最低限必要なクラスタ数を決定する。例えば、図 6 において、各カテゴリの平均ひずみ測度を 0.01 にしようとする場合には、カテゴリ 1 に 1 個の、カテゴリ 2 に 2 個のクラスタを割り当てることとなる。次節で、分布が未知である画像から得られたパターンに対し、このような考えで各カテゴリのクラスタ数を選び、更に、その数から中間層のユニット数を決定して学習を行った結果を示す。

#### 4.2 画像パターンの認識

ここでは、図 8 に示すような、野外のシーンを撮像したカラー画像 (512×400 画素、RGB 各 8 bit) を認識対象とした。領域分割は  $R$ ,  $G$ ,  $B$  の三つの特徴を用い、統計的仮説検定法<sup>(10)</sup> により行った。その後、分



図 8 実験に用いた画像の一例

Fig. 8 An example image using the experiment.

割結果の各領域において、次の特徴量を計算した。

- ・色および輝度に関する特徴：R/B, R/G, 色相, 彩度, Y, I, Q, および、画像全体の Y の平均値と、特徴量を計算している領域の Y の平均値の比
- ・テクスチャに関する特徴：濃度共起行列<sup>(10)</sup>から求められるテクスチャのコントラスト、一様性、および、濃度の線形依存性を表す量
- ・領域の位置に関する特徴：領域の重心座標 ( $c_x, c_y$ ) ( $c_x$ は画像の何列目の画素かを表し,  $c_y$ は、何行目の画素かを表す)の  $c_y$  (シーンで言えば、高さを表す)

以上の特徴量より、12次元のパターンを構成した。認識するカテゴリは{道路, 木, 芝生, 空}の四つとした。また、パターンは41枚の画像から1,311個用意し、一様乱数を用いて学習用654個および評価用657個に分割した。

使用した画像は、撮像環境として、暗れ、曇り、雨の三つの天候を含んでおり、天候に依存したパターンの変化のため、分布の単峰性は仮定できない。一例として、カテゴリ[道路]の学習パターン、一つの要素(輝度の比)のヒストグラムを図9に示す。図9より、一つの要素のみでも二つ以上のピークを有していることがわかる。

#### 4.2.1 学習結果

図10に、各カテゴリのクラスタ数に対する平均ひずみ測度の変化を示す。各カテゴリのクラスタ数の決定は、前節で述べたように行う。実験では、平均ひずみ測度を飽和させるのに、多くのクラスタを必要とするカテゴリ1、および、カテゴリ2の平均ひず

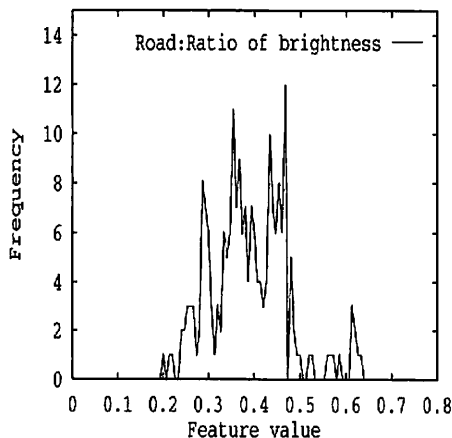


図9 特徴ベクトルの要素の分布

Fig. 9 A distribution of component of patterns obtained from color images.

み測度がほぼ飽和する値である0.01に、各カテゴリ平均ひずみ測度をそろえた。その結果、各カテゴリのクラスタ数は39, 39, 14, 8となり、合計で100個となった。中間層のユニット数は $2^p \geq 100$ を満たす最小の値の7とした。また、学習係数は0.3, 慣性係数は0.7とした。

学習パターンを10,000回提示した結果の提案手法による学習曲線、および、重みの初期値を $\pm 0.3$ の一様乱数で与え、独立に20回学習を行った結果の学習曲線を

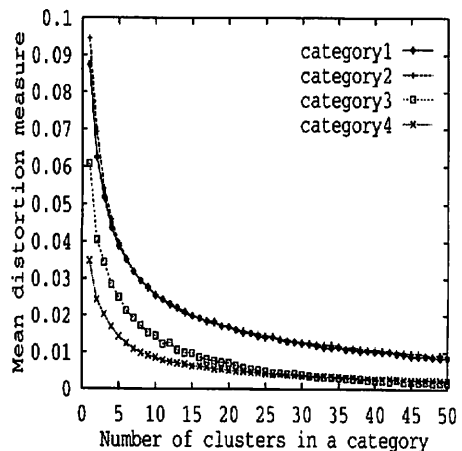


図10 画像から得られたパターンに対するクラスタ数による平均ひずみ測度の変化

Fig. 10 Changes of mean distortion measure as a function of the number of clusters in a category for patterns obtained from color images.

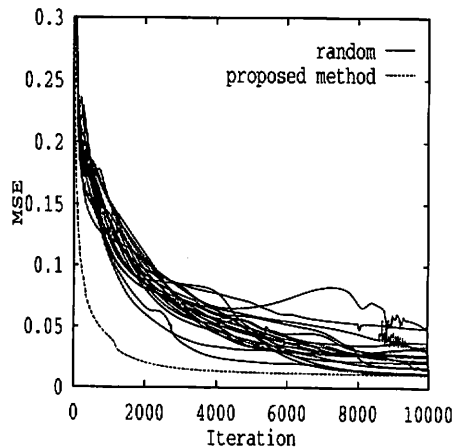


図11 提案手法で構成した3層NNの画像から得られたパターンに対する学習曲線

Fig. 11 Learning curves of three layered neural network that initial weights were setted by proposed method for patterns obtained from color images.

表2 画像パターンに対する認識率 [%]

Iteration	Initial setting	Learning set	Test set
0	Proposed method	75.99	72.30
	Random	26.96	26.94
4000 (learning coefficient :0.3)	Proposed method	99.85	97.72
	Random	99.04	97.60
10000 (learning coefficient :0.3)	Proposed method	99.85	97.72
	Random	99.59	97.60
4000 (learning coefficient :0.9)	Proposed method	99.85	97.56
	Random	97.71	97.41
10000 (learning coefficient :0.9)	Proposed method	99.85	97.87
	Random	98.62	96.80

図 11 に示す。

乱数で重みの初期値を設定した場合には、誤差が一時的に増加したり、振動するなどの現象が見られる。一方、上記のようにクラスタ数を決定し、ネットワークを構成した提案手法では、そのような現象を生じることなしに、学習が進行し、誤差が減少していることがわかる。このことから、乱数で重みの初期値を設定した場合に比べ、提案手法は誤差の収束性に優れていることがわかる。

重みの初期値を設定した時点と、4,000 回および 10,000 回の学習を行った後の 3 層 NN を用いて、学習パターンおよび評価パターンの認識率を調べた。乱数の場合は、20 回の学習結果の平均である。結果を表 2 に示す。表 2 より、提案手法で重みの初期値を設定した方が、初期値を設定した時点で高い認識率を示し、また、誤差の収束性に優れていることから、同一の学習回数で高い認識率を得られることがわかる。特に、4,000 回学習した時点で、提案手法では、10,000 回学習した後の認識率と同じ値に達しているのに対して、乱数の場合には、10,000 回学習した後の認識率より、低い値にしか達しておらず、提案手法に比べ、学習パターンで約 0.8 % の差が見られる。

クラスタ数を選択する平均ひずみ測度を 0.015 にする

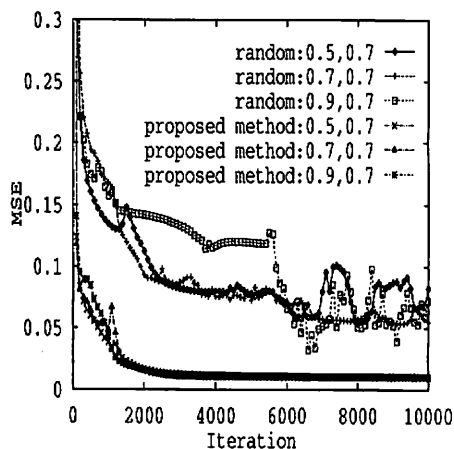


図 12 学習係数に対する学習曲線の変化  
Fig. 12 Changes of learning curve as a function of the learning coefficient.

と、中間層のユニットの数は 6 個となり、その場合も同様の結果を得た。しかし、平均ひずみ測度を 0.02 とすると、学習の進行は促進されなかった。この実験では、パターンの分布が複雑と考えられるカテゴリー 1 とカテゴリー 2 の総ひずみがほぼ飽和する点として、平均ひずみ測度を 0.01 としたが、最適な値の選択は今後の課題の一つである。

#### 4.2.2 学習係数の変化による影響

誤差逆伝搬学習則を用いる場合、学習係数および慣性係数を設定する必要がある。ここでは、各カテゴリーの平均ひずみ測度を 0.01 にそろえて構成した上記の 3 層 NN に対し、慣性係数を 0.7 に固定し、学習係数を 0.5, 0.7 および 0.9 と変化させた場合の学習曲線の変化を調べた。その結果を図 12 に示す。また、乱数で重みの初期値を設定した場合の結果も図 12 に示す。

乱数で重みの初期値を設定した場合には、学習係数を大きくしたために誤差が振動し、学習がうまく進行していないことがわかる。一方、提案手法では、ほとんど誤差が振動することなしに、学習が進行している。

以上のように提案手法は、学習係数の変化に対して頑健である。通常の場合、これらの係数は試行錯誤で決定しなければならないため、本手法のこの特徴は望ましいものである。

学習係数を大きくとることによって、更に学習は進行しやすくなる。提案手法において、認識率に変化が見られた、学習係数が 0.9 の場合の認識率を表 2 に示す (学習係数が、0.5, 0.7 の場合には認識率の変化は見られなかった)。表 2 より、10,000 回学習した後の評価パ



ターンに対する認識率が97.87%となり、学習係数が0.3の場合の認識率97.72%に比べ向上している。また、乱数で重みの初期値を設定した場合には、学習が不安定になる傾向を示すため、提案手法との認識率の差が大きくなっている。特に、学習パターンに対して、約1~2%の差が見られる。

## 5. むすび

パターン認識を行う3層ニューラルネットワークの中間層の働きを、学習パターンに含まれるクラスタごとの判別分析とみなし、与えられた学習パターンのクラスタ分析の結果に基づいて、重みの初期値を設定する手法を提案した。提案した手法は学習パターンの直和分解によるクラスタ分析、および、各クラスタに属するパターンの、そのクラスタに対応する内部表現(代表ベクトル)への集塊化の2段階の処理からなる。また、平均ひずみ測度と各カテゴリーのクラスタ数との関連を調べ、クラスタ数(中間層のユニット数)の設定について考察し、各カテゴリーの平均ひずみ測度にある値にそろえ、その値を実現するのに最低限必要なクラスタ数を設定する手法について考察した。分布の単峰性を仮定できないパターンに対する数値実験を行った結果、乱数で重みの初期値を設定した場合に比べ、提案手法は学習の進行を促進し、同一の学習回数で高い認識率を得られることを確認した。また、提案手法が学習係数の変化に対して頑健であるという性質をもつことも確認した。

提案手法では、出力層および中間層のユニットの出力を2値符号とし、シグモイド関数の飽和領域を用いるため、誤差の伝搬量が小さくなるようであるが、これは学習開始時の誤差が既にある程度小さい値となっているためである。しかし、提案手法と誤差の伝搬量を意図的に増加させるような学習高速化手法<sup>(9)</sup>を併用すれば、より学習の進行が促進されるものと期待される。

今後の課題としては、上記のような高速化手法との併用についての検討や、クラスタ数(中間層のユニット数)の決定法、および、その汎化能力との関連の解析が挙げられる。

## 文 献

- (1) 賈 棋, 戸田尚宏, 白井支朗: "ニューラルネットワークにおける逆伝搬学習アルゴリズムの初期値設定に関する一考察". 信学論(D-II), J73-D-II, 8, pp. 1179-1185 (1990-08).

- (2) Smyth S. G.: "Designing multilayer perceptrons from nearest-neighbor systems", IEEE Trans. Neural Network, 3, 2, pp. 329-333 (1992).
- (3) Gallinari P., Thiria S. and Soulie F. F.: "Multilayer perceptrons and data analysis", Proc. of ICNN-88, pp. I-391-399 (1988).
- (4) Asoh H. and Otsu N.: "Nonlinear data analysis and multilayer perceptrons", Proc. of IJCNN-89, pp. II-411-415 (1989).
- (5) 上坂吉則, 尾関和彦: "パターン認識と学習のアルゴリズム", 文一総合出版(1990).
- (6) Linde Y., Buzo A. and Gray R. M.: "An algorithm for vector quantizer Design", IEEE Trans. Communications, COM-28, 1, pp. 84-95 (1980).
- (7) 柳井晴夫, 竹内 啓: "射影行列, 一般逆行列, 特異値分解", 東京大学出版会(1983).
- (8) Rumelhart D. E., Hilton G. E. and Williams R. J.: "Learning representations by back-propagating errors", Nature, 323, 9, pp. 533-536 (1986).
- (9) 田中啓夫, 古村光夫: "誤差逆伝量の特異点解消による学習の高速化", 信学論(D-II), J75-D-II, 5, pp. 1000-1008 (1992-05).
- (10) 田村秀行: "コンピュータ画像処理入門", 総研出版(1985).

## 付 録

### 内部表現の線形分離性について

$p$ 次元ベクトル  $x_i (i=1, \dots, M; M=2^p)$  を,  $0, 1$  を要素とする2値符号としたとき, この  $x_i$  が

$$x_i < x_j \Leftrightarrow i < j$$

を満たすならば,

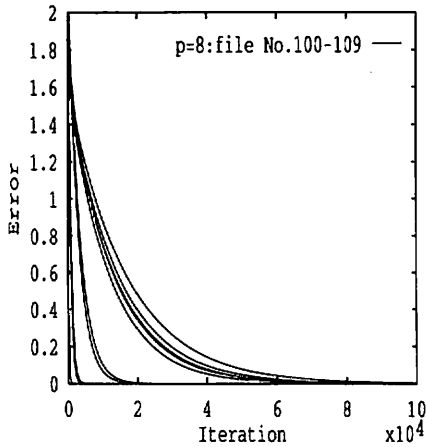
$$x_{i_1}, \dots, x_{M_1} \in C_1$$

$$x_{M_1+1}, \dots, x_M \in C_2 \quad (1 \leq M_1 < M)$$

の二つのカテゴリーは線形分離可能であるということを確認する。より複数のカテゴリーがあっても、ある一つのカテゴリー以外のすべてのカテゴリーを、まとめて一つのカテゴリーとみなし、順次に二つのカテゴリーとして分けていくことができるので、二つのカテゴリーの場合を考えれば十分である。一般的証明を与えることはできなかったが、ここでは、 $p \leq 8$  の場合について、線形分離可能か否かを判断するアルゴリズム<sup>(5)</sup>を適用することによって、線形分離性を確認した。このアルゴリズムは、カテゴリー1のパターンの要素に-1を掛け、更に、すべてのパターンの第1要素として1を加えたパターンを行とする行列を  $A(M \times (n+1))$ 、すべての要素を正とするベクトルを  $b(M \times 1)$  としたとき、

$$e = \|Aw - b\|$$

を0とする重みベクトル  $w((n+1) \times 1)$  が存在するか否



図A・1 線形分離性を判断するアルゴリズムの収束の様子 ( $p=8$ )

Fig. A.1 Convergence process of algorithm that checks up linear separability.

かを調べる。  $e$  が 0 となれば、二つのカテゴリーは線形分離可能である。  $x_i$  は 2 値符号であるので、超立方体  $[0, 1]^p$  の対称性から、  $M_1 = 2^{p-1}$  までの場合を議論すればよい。 よって、  $2^{p-1}$  個のパターンの組合せをすべて計算機を用いて発生させ、それらが線形分離可能か否かを調べた。

その結果、  $p \leq 8$  のすべての場合について線形分離可能であることが確認された。 処理の結果の一部として、  $p=8$  の場合の  $e$  が収束していく過程を、図A・1に示す。

(平成5年3月3日受付, 8月19日再受付)

### 永井 豊



昭57電通大・電子卒。昭59電気通信大学電気通信学研究科修士課程(電子工学専攻)了。同年電気通信大学電子工学科助手、現在に至る。工修。画像処理、ロボット、ファジィ制御の研究に従事。テレビジョン学会、日本ファジィ学会各会員。

### 市村 直幸



平1電通大・電気通信・通信工卒。平3同大大学院電気通信学研究科博士前期課程(電子工学専攻)了。現在、同大大学院博士後期課程在学中。画像処理、画像認識、知識の表現とその利用についての研究に従事。日本ファジィ学会会員。

### 竹内 倶佳



昭39電通大・電子卒。同年同大学助手、昭51電気通信大学短期大学部助教授、昭55電気通信大学助教授、平2同教授、現在に至る。海洋の計測・制御などの研究に従事。工博。電気学会、計測自動制御学会、日本ファジィ学会、日本音響学会、日本造船学会等の各会員。