

独立成分分析を用いた数字パターンの特徴抽出

正員 小谷 学 (神戸大学)

正員 小澤 誠 一 (神戸大学)

Feature Extraction of Digit Patterns Utilizing Independent Component Analysis

Kotani Manabu, Member, Ozawa Seiichi, Member (Kobe University)

In this paper, an approach to feature extraction utilizing independent component analysis (ICA) is proposed. In our approach, input patterns are transformed into feature vectors using ICA-bases that are obtained through two-layer neural network learning. A k -NN classifier is applied to these ICA feature vectors when the recognition accuracy is evaluated. Hand-written digits in MNIST database are used as target characters. Fast ICA algorithm is applied to these images in order to learn ICA-bases. In recognition experiments, we demonstrate that the ICA approach realizes a potential feature extraction method for hand-written digits. Furthermore, we show the addition of noise patterns to training data is effective for elimination of redundant basis functions.

キーワード：独立成分分析, 手書き, 数字認識

1. はじめに

高次モーメント量を用いて入力信号から独立な信号を分離する独立成分分析 (Independent Component Analysis, ICA) が注目されている^{(1) (2) (3)}. ICA は, これまで主にブラインド信号分離問題に適用され, 混合信号から音声や画像などの原信号を分離するのに応用されてきた⁽⁴⁾. 一方, 最近ではパターンの特徴抽出も ICA の有望な応用として期待されている. 例えば, 自然画像に対して ICA を適用したいくつかの研究では, 基底関数が視覚野における単純細胞の受容野特性によく似ているという報告がなされている^{(5) (6) (7)}. つまり, ICA は入力画像から方向をもったエッジの検出を行っているということを示している.

また, ICA によって抽出される特徴を利用してパターン認識を行う試みも報告されている. Bartlett & Sejnowski は, ICA によって顔画像から抽出された特徴ベクトルが主成分分析 (Principal Component Analysis, PCA) に基づいて抽出された特徴ベクトルに比べて, 視点に対して不変な性質をもっていることをシミュレーション実験で示した⁽⁸⁾. PCA が 2 次以下のモーメント量を用いて無相関な成分に分離する手法であることを考えれば, 上記の結果は, ICA による高次特徴量の抽出が顔パターンにおける不変な特徴の抽出に貢献していることを示唆している. このような ICA 特徴ベクトルの性質は, 他のパターン認識問題においても有効であると考えられる. 我々もこれまでに, ETL-4 データベースなどを用いて手書き文字画像への適用を検討して

いる^{(9) (10)}. しかし, これらの研究では, 認識率をより向上させるための基底関数の選択方法についての検討が充分にされていない.

本論文では, 手書き数字認識を対象として, ICA によって得られる独立成分の特徴量としての有効性を評価することを目的としている. さらに, 基底関数の選択方法を提案する. なお, 有効性の評価としては k 近傍法による認識率を用いる.

第 2 節では, 独立成分を抽出するための目的関数を説明し, そこから導出される ICA アルゴリズムについて述べる. 第 3 節では ICA を用いた手書き数字パターンの特徴抽出法と認識実験の結果について述べ, 第 4 節でノイズ付加について考察する. 第 5 節では結論を示し, 今後の課題について述べる.

2. 独立成分分析

独立成分を得るための目的関数の違いから, これまでいくつかの ICA アルゴリズムが提案されており, 通常, これらの ICA アルゴリズムで得られる独立成分は互いに異なると考えられる. しかしながら, どの ICA アルゴリズムが数字パターンの特徴抽出に適しているかという議論は困難であるため, ここでは計算の高速性から Hyvärinen らが提案した Fast ICA アルゴリズム⁽¹¹⁾を採用することにする.

時刻 t で観測される m 次元入力信号 (時間平均ゼロ) を $\mathbf{v}(t) = \{v_1(t), \dots, v_m(t)\}'$ とおき, これを白色化 (無相関で分散の等しい信号への変換) した n 次元信号を

$\mathbf{x}(t) = \{x_1(t), \dots, x_n(t)\}'$ とおく。ここで l は行列およびベクトルの転置を表す。 \mathbf{v} と \mathbf{x} は $n \times m$ ($n \leq m$) の白色化行列 \mathbf{M} を用いて以下のように表される。

$$\mathbf{x}(t) = \mathbf{M}\mathbf{v}(t) = \mathbf{D}^{-1/2}\mathbf{E}'\mathbf{v}(t) \dots\dots\dots (1)$$

ここで、 \mathbf{D} は固有値行列、 \mathbf{E} は固有ベクトル行列である。いま、観測信号 $\mathbf{v}(t)$ が統計的に独立な n 個の成分 $\mathbf{s}(t) = \{s_1(t), \dots, s_n(t)\}'$ からなっていると仮定しよう。このとき、白色化信号 $\mathbf{x}(t)$ から独立成分 $\mathbf{s}(t)$ への式 (2) の線形変換が存在するはずである。

$$\mathbf{s}(t) = \mathbf{W}\mathbf{x}(t) \dots\dots\dots (2)$$

行列 $\mathbf{W} = \{\mathbf{w}'_1, \dots, \mathbf{w}'_n\}'$ は分離行列と呼ばれ、2 層ニューラルネットの学習を通して得られる⁽³⁾。このニューラルネットは n 個の出力をもち、これら出力を $\tilde{\mathbf{s}}(t) = \{\tilde{s}_1, \dots, \tilde{s}_n\}'$ で表すことにする。

成分 s_i が統計的に独立とは、式 (3) に示すように $s_1 \dots s_n$ の同時確率密度がそれぞれの周辺確率密度の積で表されることを意味する。

$$p[s_1, \dots, s_n] = \prod_{i=1}^n p_i[s_i] \dots\dots\dots (3)$$

ここで、 $p[\cdot]$ は確率密度関数を表す。上記の確率密度関数は一般に未知であるので、式 (3) をなるべく満たすニューラルネットの出力 $\tilde{\mathbf{s}}(t)$ が得られるように (つまり $\tilde{\mathbf{s}}(t) \simeq \mathbf{s}(t)$)、適切な目的関数を設定して分離行列を求める必要がある。Karhunen と Oja は、出力信号 $\tilde{\mathbf{s}}$ に関して以下の目的関数を最大化する方法を提案した⁽³⁾。

$$J(\tilde{\mathbf{s}}) = \sum_{i=1}^n |E\{\tilde{s}_i^4\} - 3[E\{\tilde{s}_i^2\}]^2|, \dots\dots\dots (4)$$

ここで $E\{\cdot\}$ は全サンプルに対する期待値を表す。式 (4) は 4 次キュムラントで kurtosis (尖度) と密接に関連した目的関数である。 s_i の分布がガウシアン⁽⁴⁾ のとき、この目的関数は 0 となることから、式 (4) の目的関数を最大化することは、 s_i が非ガウシアン分布となるような分離行列 \mathbf{W} を求めることに相当する。これは、独立信号の分布が非ガウシアンであれば、中心極限定理によって独立信号が混合された観測信号の分布はガウシアンに近づくことになる。すなわち、観測信号の目的関数値は、独立信号の目的関数値よりも小さくなる。そこで、目的関数を最大化すれば、元の非ガウシアン分布の成分 (独立信号) を求めることができることになる。なお、独立成分は非ガウシアンという前提が必要であるが、独立成分のうち一つだけはガウシアンであっても問題ないとされている。

分離行列の学習アルゴリズムは、式 (4) の \mathbf{w}_i に関する勾配ベクトルから導出される。Hyvärinen と Oja は、式 (4) に結合ベクトル \mathbf{w}_i の大きさを制限するためのペナルティー項を加えた目的関数から、以下のような学習アルゴリズムを導出している。

$$\begin{aligned} \mathbf{w}_i(t+1) &= \mathbf{w}_i(t) \pm \mu(t)[\mathbf{x}(t)(\mathbf{w}_i(t)^T \mathbf{x}(t))^3 \\ &\quad - 3 \|\mathbf{w}_i(t)\| \mathbf{w}_i(t) + f(\|\mathbf{w}_i(t)\|^2)\mathbf{w}_i(t)] \dots (5) \end{aligned}$$

ここで f は連続で微分可能な関数、 $\mu(t)$ は学習率である。なお、 \mathbf{W} の初期値は乱数により設定される。Hyvärinen らは式 (5) の固定点を求める方程式を逐次的に (または一括に) 解きながら独立成分を高速に求める Fast ICA アルゴリズム⁽¹¹⁾を提案しており、以下ではこれを採用することにする。

3. ICA を用いた数字パターンの特徴抽出

第 2 節で述べたように、ICA アルゴリズムは式 (3) をできるだけ満足するように入力信号を分離する。このような性質は数字パターンの特徴抽出に応用できる。

式 (1), (2) から、ICA の入出力関係は次式で与えられる。

$$\tilde{\mathbf{s}}(t) = \tilde{\mathbf{W}}\mathbf{M}\mathbf{v}(t) = \mathbf{B}\mathbf{v}(t) \dots\dots\dots (6)$$

$\tilde{\mathbf{W}}$ は ICA アルゴリズムによって求められた分離行列で直交行列となり、 $\mathbf{B} = \tilde{\mathbf{W}}\mathbf{M}$ は $n \times m$ 行列である。いま、ICA の入力 $\mathbf{v}(t)$ を t 番目に提示された数字画像に対応付ける。このときの ICA 出力 $\tilde{\mathbf{s}}(t)$ はその数字の独立成分を出力することになり、これを特徴量として数字認識に利用することにする。ここで、 \mathbf{B} の i 番目の行ベクトル \mathbf{b}'_i ($i = 1, \dots, n$) は n 次元特徴空間を張る基底ベクトルに相当すると考えられる。以後、この基底ベクトルを ICA 基底と呼び、ICA の出力として得られる特徴ベクトルを ICA 特徴ベクトルと呼ぶことにする。

前節において、入力画像が n 個の独立な成分から構成されていると仮定した。しかし、この独立成分の個数を前もって知ることは一般に難しい。また、ICA アルゴリズムは厳密な意味で独立成分に分離できるのではなく、なるべく統計的に独立な性質を満たすよう分離するだけである。よって、実際に特徴抽出を行う際に、適切な独立成分の個数 (つまり特徴ベクトルの次元) をなんらかの方法で決定する必要がある。本論文では独立成分の個数を変化させる簡単な方法として、白色化する際に固有値の大きさに従って、入力信号の次元を削減することを検討する。

ICA 特徴ベクトルを得るには、いくつかの数字画像サンプルを用いて ICA 基底の学習を前もって行っておく必要がある。よって、認識率を評価するためのテストサンプルだけでなく、ICA 基底の学習に用いるための訓練サンプルも必要になる。これらテストサンプルおよび訓練サンプルは、主成分分析により固有値の大きさに従って次元削減されて、ICA 基底の学習および認識率の評価に利用される。なお、学習時に用いられる訓練サンプルは、Fast ICA アルゴリズムが収束するまで繰り返し提示される。

$L \times L$ ピクセルの T 個のオリジナル画像が n 次元に削減されたとすると、ICA 基底は i ($i = 1, \dots, n$) 番目の特徴ベクトルの要素 (ICA 出力) $\{\tilde{s}_i(1), \dots, \tilde{s}_i(T)\}$ に対する 4 次のモーメント量 (kurtosis) が最大または最小となるよう

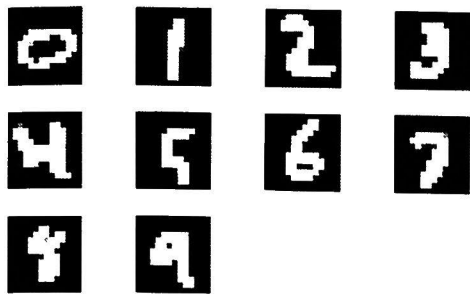


図 1 平均化処理された数字画像の例
Fig. 1. Examples of shrunk digit patterns.

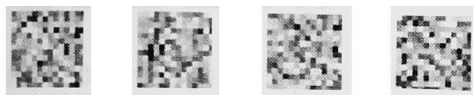


図 2 ノイズ・パターンの例
Fig. 2. Examples of noise patterns.

に学習される．このことと、 \tilde{s}_i の分布がガウス分布となる
ときに kurtosis がゼロとなることを考え合わせると、ICA
基底は \tilde{s}_i の分布が非ガウス分布となるような特徴抽出を
行っているといえる⁽³⁾．

4. 認識実験

特徴抽出法として ICA を用いる有効性を議論するため、
ICA 特徴ベクトルに対する認識率を調べた．認識実験に用
いた手書き数字データベースは、MNIST⁽¹²⁾ に収録されて
いる 28 × 28 ピクセルで 0 ～ 255 のグレイスケール画像で
ある．また、訓練パターン数は 60,000 個で、テストパター
ン数は 10,000 個収録されている．このデータベースに対
して計算時間の低減のため、平均化処理によってオリジナル
数字画像を、14 × 14 ピクセルで 0 ～ 255 グレイスケール
画像に縮小化した．このような平均化処理された数字画像
の例を図 1 に示す．この図では黒が 0、白が 255 のピク
セル値を表している．これらの数字画像は、前節で述べたよ
うな固有値の大きさに従って、 n 次元に次元削減される．

訓練パターンはプロトタイプの生成に用いられるだけで
なく、ICA 基底の学習にも用いられる．ICA 基底の学習時
には、オリジナル画像だけでなくノイズ・パターンもニュー
ラルネットの入力として用いた．なお、ノイズは正規乱数
で 0 ～ 255 の整数値である．図 2 にノイズ・パターンの例
を示す．

クラス c に属する i 番目の訓練パターンに対する ICA
特徴ベクトルを \mathbf{p}_i^c 、 j 番目のテストパターンを $\tilde{\mathbf{s}}_j(t)$ とす
る．認識においては、 \mathbf{p}_i^c と $\tilde{\mathbf{s}}_j$ の間に以下の類似度を採用
した．

$$S_j(c, i) = \frac{\tilde{\mathbf{s}}_j^T \mathbf{p}_i^c}{\|\tilde{\mathbf{s}}_j\| \|\mathbf{p}_i^c\|} \dots\dots\dots (7)$$

表 1 交差確認法による ICA 特徴ベクトルを用い
た認識率 [%]

Table 1. Recognition accuracy [%] for ICA Using
7 folds cross validation

n	Accuracy	n	Accuracy
20	90.8	55	92.2
25	92.5	60	92.2
30	93.2	75	91.3
35	93.7	100	89.7
40	93.6	125	86.8
45	93.1	150	84.5
50	92.9	177	84.1

表 2 ノイズなしで学習した種々の特徴ベクトル
の次元に対する ICA 特徴ベクトルと PCA 特徴ベ
クトルを用いた認識率 [%]

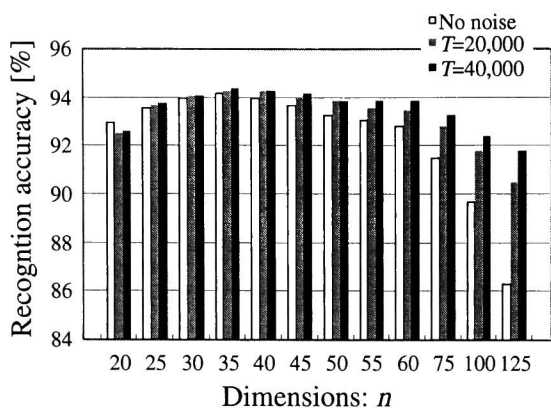
Table 2. Recognition accuracy [%] for ICA and
PCA feature vectors trained with only digit
patterns.

n	ICA	PCA	n	ICA	PCA
20	93.0	92.8	55	93.1	93.9
25	93.6	93.3	60	92.8	93.9
30	94.0	93.7	75	91.5	93.9
35	94.2	93.8	100	89.7	93.7
40	94.0	93.9	125	86.3	93.6
45	93.7	93.9	150	84.2	93.9
50	93.3	93.9	177	83.8	93.6

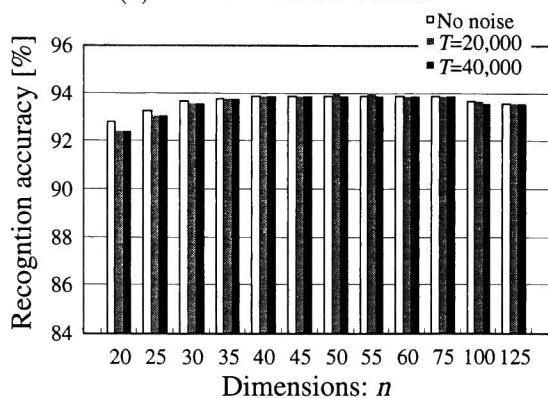
この類似度に対して k 近傍 (k -NN) 法を適用し、認識を行
った．

まず、今回のデータに対して最適な特徴ベクトルの次元
を選択するために、主成分分析によって得られた固有値の大
きさに基づいて次元を削減し、その様々な次元で ICA の特
徴ベクトルを学習した．なお、訓練と評価に用いたパターン
は、全訓練パターン 60,000 個からランダムに選んだ 7,000
個である．これらのパターンを 7 組に分けて、交差確認法
により識別率を算出した．表 1 に特徴ベクトルの次元 n
を 20 ～ 177 とした時の k 近傍法による認識結果を示す．な
お、 k は 3, 5, 7, 10 の各場合で認識率を求め、その中で
の最高の認識率を示している．また、訓練パターンにはノイ
ズを含めていない．表 1 の結果のように、 $n = 35$ の時に最
も高い認識率が得られたので、今回のデータに対しては特
徴ベクトルの次元は 35 が最適と考えられる．

次に、MNIST に含まれている全評価パターン 10,000 個
に対して、 $n = 35$ だけでなく $n = 20 \sim 177$ の場合の識別
率を求めた結果を表 2 に示す．比較のため、PCA によっ
て求められた特徴ベクトルに対する認識率も調べた．なお、そ
他の実験条件は表 1 と同じである．表 2 より、ICA 特徴ベ
クトルと PCA 特徴ベクトル共に特徴ベクトルの次元が低
くても高くても認識率が低下している．その傾向は ICA 特
徴ベクトルの方が顕著である．また、ICA 特徴ベクトル
の方が特徴ベクトルの次元による認識率の変化が大きい．一
方、PCA 特徴ベクトルの方は特徴ベクトルの次元 30 から
100 まではほとんど認識率は変っていない．さらに、ICA 特



(a) For ICA feature vectors



(b) For PCA feature vectors

図3 種々のノイズ・パターン数で学習した場合の認識率

Fig. 3. Recognition accuracy in the case of noise patterns, $T = 20,000$ and $T = 40,000$, and no noise patterns, $T = 6,000$.

特徴ベクトルでは $n = 35$ の時に 94.2% の認識率が得られているが、これは PCA 特徴ベクトルの最高である 93.9% より、0.3% 程度高い結果となっている。これは、パターン数にして 10,000 個中 30 個である。以上より、適切な次元を選択すれば、ICA 特徴ベクトルは PCA 特徴ベクトルよりも優れた特徴抽出法であると考えられる。

次に、ノイズ・パターンを含めて学習した場合の認識率を図 3 に示す。図 3 には訓練パターンとノイズパターンの和を T として、 $T = 20,000$ と $40,000$ の場合とノイズ・パターンのない場合 ($T = 6,000$) の認識率を併せて示している。

このように、PCA 特徴ベクトルを用いた場合ではノイズ・パターンを含めても認識率はほとんど差がないことがわかる。一方、ICA 特徴ベクトルの場合には、特徴ベクトルの次元数の違いによる認識率の差が小さくなっている。この傾向はノイズ・パターンを多くした方がより明確になっている。特に、 $n = 75, 100, 125$ などの場合でノイズの付加による認識率の改善が顕著である。さらに、 $T = 40,000$

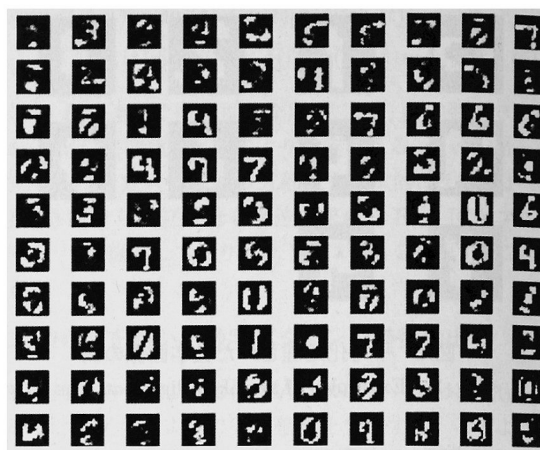


図4 150 次元の特徴ベクトルでノイズなしで学習した場合の基底関数例

Fig. 4. Examples of basis functions trained with only digit patterns when dimensions are 150.

の場合では最高の識別率はノイズなしに比べて約 0.2% 向上した 94.4% となっている。このように、ICA 特徴ベクトルを学習するときには、ノイズ・パターンを含める方が有効であることわかる。

5. ノイズ・パターンの効果に関する考察

前節では、ノイズ・パターンを訓練パターンに含めることによって、特に特徴ベクトルの次元が高い場合にテスト・パターンに対する認識率が大きく向上することを示した。ここでは、ノイズ・パターンを訓練パターンに含める効果について考察する。まず、図 4 に $n = 150$ でノイズなしで学習した基底関数の内 100 個を示す。基底関数とは A の列ベクトルであり、 $A = ED^{1/2}\tilde{W}'$ となる。なお、以下の各基底関数の図では黒が最小のピクセル値で、白が最大値となっている。

図 5 に $n = 150$ でノイズ・パターンを含めて ($T = 40,000$) 学習した基底関数のうち 100 個を示す。白黒が反転している基底関数があるが、これは ICA で求める分離行列の任意性によるものである。このように反転しても、独立成分の係数の符号が逆になる程度で認識への影響はない。これらの結果より、学習時にノイズ・パターンを付加することによって、基底関数がかかなり変わることがわかる。図 4 では、数字パターンに類似した基底が多くあるが、どの数字に一致するかとなると不明確である。さらに、類似した基底関数もいくつか見受けられる。それに対して、ノイズを付加して学習した図 5 では、数字パターンに類似した基底関数は少ないが、類似した基底関数はかなり明瞭に数字と一致していることがわかる。また、ノイズ・パターンに類似した基底関数が増えているが、図 4 ほど相互に類似した基底関数は多くない。

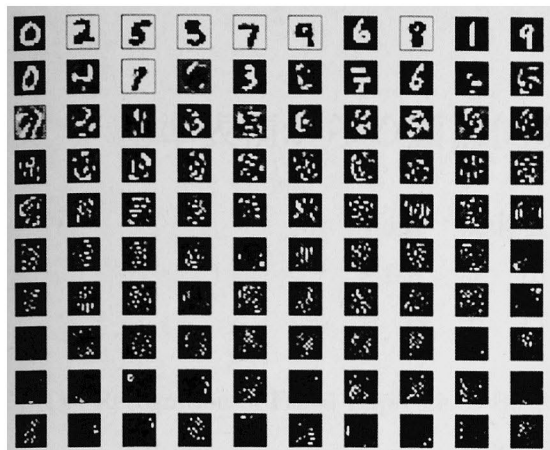


図5 150次元の特徴ベクトルでノイズ・パターンを含めた $T = 40,000$ で学習した場合の基底関数例

Fig.5. Examples of basis functions trained with digit patterns and noise patterns, $T = 40,000$, when dimensions are 150.

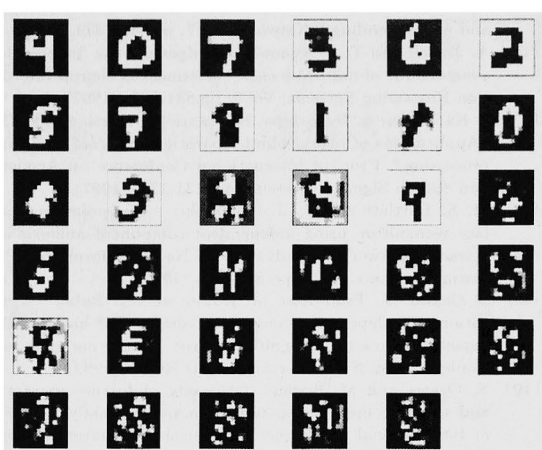


図7 35次元の特徴ベクトルでノイズ・パターンを含めた $T = 40,000$ で学習した場合の基底関数例

Fig.7. Examples of basis functions trained with digit patterns and noise patterns, $T = 40,000$, when dimensions are 35.

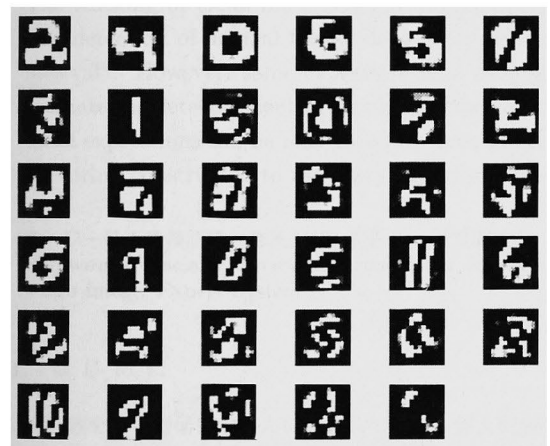


図6 35次元の特徴ベクトルでノイズなしで学習した場合の基底関数例

Fig.6. Examples of basis functions trained with only digit patterns when dimensions are 35.

図6と図7にノイズを付加しない場合と付加した場合の $n = 35$ の基底関数を示す。これら図より、ノイズの有無による基底関数の違いが、 $n = 150$ の場合ほど大きくないことがわかる。さらに、ノイズなしで学習した場合でも類似した基底関数はほとんどない。

これらの結果は、PCAによって得られる固有値の大きい順に特徴ベクトルの次元を適切に設定すれば、ノイズの有無に関係なく相互に類似した基底関数が減少するので、認識率という点では大きな改善はないことを意味している。しかし、特徴ベクトルの次元が高すぎて類似した基底関数

が多く存在するような場合には、ノイズを付加することによって、冗長でない基底関数が得られる。それによって、認識率が向上していると考えられる。

6. おわりに

独立成分分析を用いた特徴抽出へのアプローチを提案し、手書きひらがな数字に対する認識率を調べることで、その有効性を検討した。その結果、高い認識性能を得るためには独立成分分析による特徴ベクトルの次元を主成分分析によって適切に設定すべきであることを示した。また、独立成分分析による基底の学習で与える訓練サンプルに、ノイズを加えることも有効であることを示した。

今回は独立成分分析によって得られた特徴の有効性を把握するために、識別方法として比較的単純な k -近傍法を適用したが、今後の課題として抽出された独立成分に適した識別方法について検討していきたいと考えている。

(平成13年03月26日受付、同13年11月15日再受付)

文 献

- (1) A. J. Bell and T. J. Sejnowski: "An information maximization approach to blind separation and blind deconvolution," Neural Computation, Vol.7, pp.1129-1159, (1995).
- (2) S. Amari, A. Chichocki, and H. Yang: "A new learning algorithm for blind signal separation," Advances in Neural Information Processing Systems, Vol.8, pp.757-763, (1996).
- (3) J. Karhunen and E. Oja: "A class of neural networks for independent component analysis," IEEE Trans. on Neural Networks, Vol.8, No.3, pp.486-503, (1997).
- (4) A. Hyvärinen: "Survey on independent component analysis," Neural Computing Surveys, Vol.2, pp.94-128, (1999).
- (5) B. A. Olshausen and D. J. Field: "Natural image statistics

- and efficient coding," Network, Vol.7, pp. 333-339, (1996).
- (6) A. J. Bell and T. J. Sejnowski: "Edges are the 'independent components' of natural scenes," Advances in Neural Information Processing Systems, Vol.9, pp.831-837, (1997).
 - (7) J. Karhunen, A. Hyvärinen, R. Vigario, J. Hurri, and E. Oja: "Applications of neural blind separation to signal and image processing," Proc. of International Conference on Acoustics and Speech Signal Processing, pp.131-134, (1997).
 - (8) M. S. Bartlett and T. J. Sejnowski: "Viewpoint invariant face recognition using independent component analysis and attractor networks," Advances in Neural Information Processing Systems, Vol.9, pp.817-823, (1997).
 - (9) S. Ozawa, T. Tsujimoto, M. Kotani and N. Baba: "Application of independent component analysis to hand-written Japanese character recognition", Proc. of International Joint Conference on Neural Networks, CD-ROM, (1999).
 - (10) S. Ozawa and M. Kotani: "A study of feature extraction and selection using independent component analysis", Proc. of International Conference on Neural Information Processing, CD-ROM, (2000).
 - (11) A. Hyvärinen and E. Oja: "A fast fixed-point algorithm for independent component analysis," Neural Computation, Vol.9, pp.1483-1492, (1997).
 - (12) <http://www.research.att.com/~yann/exdb/mnist/index.html>

小 谷 学 (正員) 1983 年, 神戸大学大学院修士課程計測工学専攻修了。同年 (株) 神戸製鋼所入社。89 年神戸大学工学部計測工学科助手を経て現在情報知能工学科助教授。電子的計測, 知能システムなどの研究に従事。計測自動制御学会, 電子情報通信学会, システム制御情報学会, 日本神経回路学会などの会員 (博士 (工学))。



小 澤 誠 一 (正員) 1989 年, 神戸大学大学院修士課程計測工学専攻修了。同年奈良工業高専情報工学科助手。1993 年大阪教育大学教育学部教養学科助手, 講師。2000 年神戸大学大学院自然科学研究科助教授。ニューラルネットの学習理論に興味をもち, パターン認識や連想記憶への応用にも従事。計測自動制御学会, 電子情報通信学会, システム制御情報学会, International Neural Network Society などの会員 (博士 (工学))。

