クラス間距離を最大化する教師あり 独立成分分析による特徴抽出

非会員 坂口 善規* 正員 小澤 誠一*
正員 小谷 学*

Feature Extraction Using Supervised Independent Component Analysis
by Maximizing Class Distance

Yoshinori Sakaguchi*, Non-member, Seiichi Ozawa*, Member, Manabu Kotani*, Member

Recently, Independent Component Analysis (ICA) has been applied to not only problems of blind signal separation, but also feature extraction of patterns. However, the effectiveness of pattern features extracted by conventional ICA algorithms depends on pattern sets; that is, how patterns are distributed in the feature space. As one of the reasons, we have pointed out that ICA features are obtained by increasing only their independence even if the class information is available. In this context, we can expect that more high-performance features can be obtained by introducing the class information into conventional ICA algorithms.

In this paper, we propose a supervised ICA (SICA) that maximizes Mahalanobis distance between features of different classes as well as maximize their independence. In the first experiment, two-dimensional artificial data are applied to the proposed SICA algorithm to see how maximizing Mahalanobis distance works well in the feature extraction. As a result, we demonstrate that the proposed SICA algorithm gives good features with high separability as compared with principal component analysis and a conventional ICA. In the second experiment, the recognition performance of features extracted by the proposed SICA is evaluated using the three data sets of UCI Machine Learning Repository. From the results, we show that the better recognition accuracy is obtained using our proposed SICA. Furthermore, we show that pattern features extracted by SICA are better than those extracted by only maximizing the Mahalanobis distance.

キーワード：独立成分分析, 教師あり学習, 特徴抽出, パターン認識

Keywords: independent component analysis, supervised learning, feature extraction, pattern recognition

1. はじめに

入力信号に対する高次統計量を無相関化し、独立な成分に分離する手法として独立成分分析 (Independent Component Analysis; ICA) が注目されている⁽¹⁾⁽²⁾。この手法は、これまで主にブラインド信号分離問題に適用されてきたが、パターン認識の特徴抽出手法としても、一部の問題においてその有効性が示されつつある⁽³⁾⁻⁽⁶⁾。

例えば、Bartlettらは、ICAによって顔画像から抽出された特徴量が主成分分析 (Principal Component Analysis; PCA) によって抽出された特徴量に比べて、回転や平行移動などの変動に対してロバストであることを示している⁽⁷⁾。

PCAが2次統計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、この結果は、高次特徴量の無相関化が特徴抽出に有効であることを示唆している。しかしながら、これまでの我々の研究では、このような独立な特徴量の有効性は問題に大きく依存することがわかっている⁽⁸⁾。つまり、クラス情報を利用しないで単に独立性を高めることが特徴量の分離性を高めることになるかどうかは問題 (パターンの分布) に依存する。いうまでもなく、この原因の一つとして、ICAアルゴリズムがクラス情報を利用しない「教師なし学習」であることが考えられる。

ICAの教師あり学習アルゴリズムとしては、クラス情報から作り出した制御信号 (教師信号) との相関を高める梅山らの手法がある⁽⁹⁾。梅山らは、この手法を人物の表情解析などに適用し、その有効性を示している。また、指紋認識⁽¹⁰⁾や数字認識⁽¹¹⁾の特徴抽出に適用され、パターン認識への応用も検討されている。しかし、この手法は独立性

* 神戸大学大学院自然科学研究科
〒657-8501 神戸市灘区六甲台町 1-1
Graduate School of Science and Technology, Kobe University
1-1, Rokkodaicho, Nada-ku, Kobe 657-8501

を高めると同時に制御信号と基底ベクトルの相関を高めるという手法であるため、望ましい基底ベクトルがどのようなものかが明確になっている必要がある。パターンの特徴抽出に適用する場合、特徴量のクラス分離性が高い基底ベクトルがどのようなものかが事前にわかっていることはあまりなく、優れた特徴量を得るには、一般に制御信号の選定に試行錯誤を必要とする。

本稿では、特徴量の分離性を向上するような基底ベクトルを自動的に見つけるため、パターンの独立性とクラス間マハラノビス距離を同時に最大化する手法を新たに提案する。第2節では、特徴抽出法に用いられる従来の教師なしICAアルゴリズムについて述べ、第3節では提案する教師ありICAアルゴリズムを示す。第4節では2次元人工データとUCI Machine Learning Repository⁽¹²⁾に含まれるベンチマークデータを使った提案手法の性能評価を行ない、第5節で結論と今後の課題について述べる。

2. 独立成分分析による特徴抽出

k 番目に与えられる L 次元の入力パターンを $x(k) = [x_1(k), \dots, x_L(k)]^T$ で表す。ここで、 T は行列またはベクトルの転置を表す。このパターン $x(k)$ は、統計的に独立な M 個の基本成分 $s(k) = [s_1(k), \dots, s_M(k)]^T$ ($L \geq M$) が式 (1) のように線形に混合されたものと仮定する。

$$x(k) = As(k) \dots \dots \dots (1)$$

ここで、行列 A は、入力パターン (観測パターン) がどのような基本成分 $s(k)$ で構成されているかを表す $L \times M$ の混合行列である。ICA では、行列 A が未知である条件のもとで、入力パターン $x(k)$ から独立成分ベクトル $s(k)$ の推定を行う (推定値を $\tilde{s}(k)$ で表す)。そして、これによって得られた $\tilde{s}(k)$ を特徴量としてパターン認識に利用する。

本研究では、特徴抽出に用いる ICA アルゴリズムとして Oja が提案した bigradient ICA アルゴリズム⁽¹³⁾を採用する。このアルゴリズムでは、まず、入力パターン $x(k)$ が PCA によって、以下のように白色化される。

$$v(k) = D^{-1/2}U^T x(k) \dots \dots \dots (2)$$

ここで D および U は、 λ_i を共分散行列 $E\{x(k)x(k)^T\}$ の i 番目の固有値、 u_i を i 番目の固有ベクトルとしたとき、 $D = \text{diag}[\lambda_1, \dots, \lambda_M]$ および $U = [u_1, \dots, u_M]$ で与えられる。式 (2) で得られた白色化ベクトル $v(k)$ を用いて、独立成分ベクトル $\tilde{s}(k)$ は式 (3) で与えられる。

$$\tilde{s}(k) = Wv(k) \dots \dots \dots (3)$$

ここで、 $W = [w_1, \dots, w_M]^T$ は分離行列と呼ばれ、逐次的な学習を通して得られる。式 (2)-(3) から、ICA による特徴抽出における入出力関係は式 (4) で与えられる⁽⁸⁾。

$$\tilde{s}(k) = WD^{-1/2}U^T x(k) = Bx(k) \dots \dots \dots (4)$$

式 (4) で表される ICA による特徴抽出の過程を図 1 に示

図 1 ICA を用いた特徴抽出の過程

Fig. 1. Process of feature extraction using ICA.

図 2 基底の回転によって分離性が向上する例

Fig. 2. An example case that the separability is enhanced by rotating bases.

す。ここで、行列 B の行ベクトル b_i ($i = 1, 2, \dots, M$) は、 L 次元特徴空間を張る基底ベクトル (ICA 基底と呼ぶ) に相当し、 $\tilde{s}_i(k)$ はこれら ICA 基底への射影値で与えられる。Oja のアルゴリズムでは、この射影値が非ガウス分布になるように 4 次の統計量である kurtosis の最大化または最小化を行う。Oja のアルゴリズムにおける W の更新式は以下で与えられる⁽¹³⁾。

$$W_{k+1} = W_k + \mu(\tanh \tilde{s}(k))v(k)^T + \gamma W_k(I - W_k W_k^T)W_k \dots \dots \dots (5)$$

ここで、 $\tanh \tilde{s}(k)$ は $\tilde{s}$ の各成分を $\tanh \tilde{s}_i(k)$ で求めることを表す。

パターン認識において、異なるクラスに属する特徴量は、可能な限りオーバーラップが少なくなるよう分離されることが望ましい。しかし、ICA アルゴリズムはクラス情報を利用しない教師なし学習であるため、抽出された特徴量が必ずしも分離性に優れているとは限らない。これを説明するため、図 2 のようなケースを考えてみる。図 2 における 2 つの楕円は、それぞれクラス l とクラス m のパターンが存在する領域を表している。ICA では、一般に各基底への射影値 (特徴量) が非ガウス分布になるよう軸回転が行われる。そこで、いま図 2 の b_{ICA} が基底として得られたと考えよう。このとき、特徴量の分布は (a) で表されるようになり、2 つのクラスは線形分離可能であるものの、最も近接した特徴量同士の距離はかなり小さい。しかしながら、 b_{ICA} を少し時計方向に回転してやれば、2 クラス間の距離は大きくなることがわかる。つまり、単に独立性を上げるだけで高い分離性が保証されるわけでない。よって、 b_{SICA} のような軸を得るためには、パターンのクラス情報を利用し、クラス分離性に対する評価を学習に反映する必要がある。

そこで、異なるクラスに属する特徴量に対し、その分離性をクラス間マハラノビス距離で評価し、独立性とともにこ

の距離を最大化するような評価関数を次節で考えることにする。さらに、この評価関数から分離行列 W の学習アルゴリズムを導出する。以下では、導出される教師あり ICA アルゴリズムを Supervised Independent Component Analysis by maximizing Mahalanobis Distance (SICA-MD) と呼ぶことにする。

3. クラス間距離を最大化する独立成分分析

まず、異なるクラスに属する特徴量に対し、クラス間マハラノビス距離を定義する。クラス l, m の入力パターンから求められた白色化ベクトル v のクラス内共分散行列 $\Sigma_{W_{lm}}$ とクラス間共分散行列 $\Sigma_{B_{lm}}$ は以下で与えられる。

$$\Sigma_{W_{lm}} = \sum_{c=l,m} p(c) \frac{1}{n_c} \sum_{v \in \chi_c} (v - d_c)(v - d_c)^T \dots \dots \dots (6)$$

$$\Sigma_{B_{lm}} = \sum_{c=l,m} p(c)(d_c - d)(d_c - d)^T \dots \dots \dots (7)$$

($l, m = 1, \dots, C; l \neq m$)

ここで $\sum_{c=l,m}$ は、 c に l と m を代入した項をたし合わせることを表す。また C はクラス数、 χ_c はクラス c に属するパターンの集合、 n_c は各クラスのパターン数、 $p(c)$ はクラス c の事前確率である。また、 d と d_c はそれぞれ全てのパターンの平均ベクトルとクラス c に属するパターンの平均ベクトルである。

一方、基底 b_i への射影値 (特徴量) に対するクラス間共分散 $\hat{\Sigma}_{B_{lm}}$ とクラス内共分散 $\hat{\Sigma}_{W_{lm}}$ は式 (6)-(7) を用いて以下で与えられる。

$$\hat{\Sigma}_{W_{lm}} = \sum_{c=l,m} \frac{p(c)}{n_c} \sum_{v \in \chi_c} w_i^T (v - d_c)(v - d_c)^T w_i = w_i^T \Sigma_{W_{lm}} w_i \dots \dots \dots (8)$$

$$\hat{\Sigma}_{B_{lm}} = \sum_{c=l,m} p(c) w_i^T (d_c - d)(d_c - d)^T w_i = w_i^T \Sigma_{B_{lm}} w_i \dots \dots \dots (9)$$

クラス l とクラス m 間のマハラノビス距離 $\phi_{lm}(w_i)$ は、これら $\hat{\Sigma}_{W_{lm}}$ および $\hat{\Sigma}_{B_{lm}}$ を用いて式 (10) で与えられる⁽¹⁴⁾。

$$\phi_{lm}(w_i) = \frac{\hat{\Sigma}_{B_{lm}}}{\hat{\Sigma}_{W_{lm}}} = \frac{w_i^T \Sigma_{B_{lm}} w_i}{w_i^T \Sigma_{W_{lm}} w_i} \dots \dots \dots (10)$$

特徴量のクラス分離性を高める方法として、すべてのクラス間 (l, m) で式 (10) のマハラノビス距離 $\phi_{lm}(w_i)$ を最大化することが考えられる。そこで、2 クラス間のマハラノビス距離を分離性に対する評価項としてそのまま利用し、その総和を評価関数として定義する。ただし、非常に近接している特徴量については、その分離性を他に比べて高める必要がある。そこで、単純に評価項の総和をとるのではなく、2 クラス間のマハラノビス距離に応じて評価項を重み付けした評価関数を導入する。重み付けの方法には様々なものが考えられるが、本稿ではクラス l とクラス m の評

価項に対して、重み係数を $1/\phi_{lm}^2(w_{i0})$ と設定する。ここで、 w_{i0} は i 番目の初期基底ベクトルであり、本論文では PCA で求めた基底ベクトルとする。以上より、クラス分離性を高めるための評価関数は次式で表される。

$$\Phi(w_i) = \sum_{l=1}^{c-1} \sum_{m=l+1}^c \frac{1}{\phi_{lm}^2(w_{i0})} \phi_{lm}(w_i) \dots \dots \dots (11)$$

提案する SICA-MD の評価関数は、独立性に対する評価関数に式 (11) を加えたもので定義される。分離行列 W の学習アルゴリズムは、この評価関数を w_i で偏微分することによって求められるが、これは以下に示す式 (11) の勾配ベクトル ψ_i を式 (5) に加えたものに相当する。

$$\psi_i = \frac{\partial \Phi(w_i)}{\partial w_i} = \sum_{l=1}^{c-1} \sum_{m=l+1}^c \frac{1}{\phi_{lm}^2(w_{i0})} \frac{\partial \phi_{lm}(w_i)}{\partial w_i} \dots \dots \dots (12)$$

($i = 1, \dots, M$)

ただし、

$$\frac{\partial \phi_{lm}(w_i)}{\partial w_i} = \frac{2\{(w_i^T \Sigma_{W_{lm}} w_i) \Sigma_{B_{lm}} w_i - (w_i^T \Sigma_{B_{lm}} w_i) \Sigma_{W_{lm}} w_i\}}{(w_i^T \Sigma_{W_{lm}} w_i)^2} \dots \dots \dots (13)$$

である。式 (12) を Oja の bigradient ICA アルゴリズムに加えることで、SICA-MD における分離行列 W の更新式が次式で与えられる。

$$W_{k+1} = W_k + \mu(\tanh \tilde{s}(k)) v(k)^T + \gamma W_k (I - W_k W_k^T) W_k + \alpha \Psi \dots \dots \dots (14)$$

ここで、 $\Psi = [\psi_1, \psi_2, \dots, \psi_M]^T$ 、 α は正の定数である。

4. 特徴抽出への適用

4.1 特徴変換に対する考察 SICA-MD における特徴変換の様子を視覚的に示すため、クラスター状に分布した 2 次元の人工データ (x, y) に提案手法 (SICA-MD) を適用した。また、PCA や Oja の bigradient ICA アルゴリズムにも適用し、得られた基底ベクトルの比較を行った。

図 3(a)-(c) に 2 次元人工データの分布を表す。図中の点は人工データに相当し、これらが 2000 個集まって 2 つのクラスターを形成している。各クラスターをクラス 1 とクラス 2 とラベル付けし、データ数はそれぞれ 1000 とした。なお、各クラスのデータは中心と分散が異なる正規乱数で生成している。また、図 3(a)-(c) における直線は PCA と ICA, SICA-MD で得られた基底を表しており、それぞれを $b_1 \sim b_6$ で示している。

図 4(a)-(f) は、PCA と ICA, SICA-MD の 3 手法で得られた特徴量の度数分布を折れ線グラフで示している。図 3(a) と図 4(a) からわかるように、基底 b_1 で変換されたク

図3 2クラス人工データの分布と (a) PCA , (b) ICA, (c) ICA-MD で得られた基底 $b_1 \sim b_6$

Fig. 3. A distribution of 2-class artificial data and the calculated base vectors $b_1 \sim b_6$ using (a) PCA , (b) ICA, and (c) ICA-MD.

図4 2クラス人工データで得られた特徴量の度数分布

Fig. 4. The frequency distributions of features obtained from 2-class artificial data.

ラス1とクラス2の特徴量はオーバーラップを有しており、この基底だけでは2つのクラスを完全に分離できない。一方、ICAについては、図4(c)の基底 b_3 に対する度数分布から、PCAより優れた分離性を有した特徴量が求まることがわかる。つまり、この例は独立性を向上することでクラス分離性が向上するデータ分布の一例といえる。クラス分離性を陽に最大化する SICA-MD では、図4(e)の基底 b_5 に対する度数分布から、クラス境界付近において存在しているデータ数がわずかではあるが減少していることがわかる。このことから、独立性とクラス分離性を同時に最大化することで、優れた特徴量が得られているといえる。

次に、図5に示すようなクラス数が3となる人工データにも適用した。人工データはクラスターごとに1000個ずつ、正規乱数を用いて生成した。前の実験と同様、PCAと

図5 3クラス人工データの分布と (a) PCA , (b) ICA, (c) ICA-MD で得られた基底 $b_1 \sim b_6$

Fig. 5. A distribution of 3-class artificial data and the calculated base vectors $b_1 \sim b_6$ using (a) PCA , (b) ICA, and (c) ICA-MD.

図6 3クラス人工データで得られた特徴量の度数分布

Fig. 6. The frequency distributions of features obtained from 3-class artificial data.

ICA, SICA-MDによって基底を求め、それらを図5(a)-(c)の直線で表す。また、図6(a)-(f)に各手法で得られた特徴量の度数分布を折れ線グラフで示す。図6(a)-(d)からわかるように、PCAでは比較的分離性に優れた基底 b_2 が求まっているのに対して、ICAではどちらの基底であっても3クラスの特徴量が完全にオーバーラップしてしまっている。つまり、この例は独立性を向上することでクラス分離性が逆に劣化するデータ分布の一例といえる。一方、図6(f)からわかるように、SICA-MDではPCAよりもオーバーラップの少ない、分離性に優れた基底 b_6 が得られている。このことから、独立性を向上することが分離性を劣化させるケースであっても、SICA-MDではクラス分離性の最大化を優先し、優れた基底を見つけることが可能といえる。

4.2 特徴抽出に対する性能評価

4・2・1 評価データおよび実験条件 多次元多クラスのデータとして、UCI Machine Learning Repository⁽¹²⁾から breast-cancer-wisconsin データと segmentation データ、waveform データの 3 種類を選び、これらを性能評価に利用する。各データのクラス数はそれぞれ 2, 7, 3 であり、属性数はそれぞれ 9, 19, 21 である。segmentation データは訓練データが 210 個、テストデータが 2100 個で構成されている。一方、breast-cancer-wisconsin データ (データ数 683) と waveform データ (データ数 5000) に関しては、訓練データとテストデータに分かれていないため、交差検定を行うことによって、その平均性能を求める。

式 (14) のパラメータ μ と γ は、経験的に -0.001 と 0.5 と設定した。ただし、クラス分離性と独立性とのバランスを調整するパラメータ α は、訓練データの一部を検査用データとして利用し、数値計算上の発散が生じない範囲で最適化を行う[†]。つまり、いくつかの α に対して学習を行い、検査用データに対する認識率をそれぞれ求める。そして、その中で認識率が最も高かった α を採用し、そのときに得られた基底を用いて、テストデータに対する性能評価を行う。なお、検査用データは訓練データから取り出したものであり、テストデータとは異なることに注意する。また、訓練データ数の少ない breast-cancer-wisconsin データと segmentation データに関しては、訓練データの選び方で性能がばらつきやすいため、4 通りの異なった選び方で性能を平均して評価した。ICA および SICA で求める特徴ベクトルの次元は、最も高い認識率が得られた PCA 特徴ベクトルと同じ次元に設定することとし、前実験の結果から breast-cancer-wisconsin データと segmentation データ、waveform データの次元はそれぞれ 8, 14, 20 となった。

独立性とクラス間マハラノビス距離を同時に最大化することの有効性を確認するため、式 (14) の μ の値を 0 としたクラス間距離のみを最大化する学習で得られる特徴量も評価する。また、他の教師あり ICA と比較するため、梅山らの手法との比較も行う。本稿では、便宜上、この梅山らの教師あり ICA を Supervised Independent Component Analysis by Maximizing Correlation (SICA-MC) と呼ぶ。SICA-MC の制御信号には、各基底が特定クラスの共通した特徴を抽出するよう、各クラスの平均パターンを与えることとした⁽¹¹⁾。ただし、クラス数が特徴量の次元より小さいデータに関しては、一部の基底でクラスの情報が利用できなくなる。そこで、SICA-MC が不利とならないよう、複数の基底に同じ制御信号を割り当てることにする。SICA-MC では、制御信号との相関を最大化するだけでなく、特徴量の独立性も最大化するため、複数の基底に同じ制御信号が与えられても、それらがまったく同じ基底に収束することはない。breast-cancer-wisconsin データの次元数は 8 であるため、異なる 4 つの基底に対して同じ制御信

号を割り当てる。また、segmentation データの次元数は 14 であるため、異なる 2 つの基底に対して同じ制御信号を割り当てる。waveform データの次元数は 20 であるため、異なる 6 つの基底に対して同じ制御信号を割り当て、残りの 2 つの基底に関しては 3 つ制御信号のうち適当に 2 つを選び、その制御信号を割り当てる。

4・2・2 新しい距離尺度の提案 Oja の bigradient ICA アルゴリズムでは分離行列 W を正規直交行列として独立成分を求めるため、入力パターン $x(k)$ の白色化を行う。このため、PCA 基底 u_i に射影された特徴量のばらつきが正規化され、このばらつきが識別する上で重要な情報となっているデータセットでは、ICA による認識率の向上はあまり望めない。これが ICA で抽出された特徴量の認識率がデータセットに大きく依存する原因となっている可能性がある。そこで、PCA で得られた各基底 u_j の固有値 λ_j に基づき、ICA 基底 b_i の重要度を以下で定義する。

式 (4) から、ICA 基底 b_i と PCA 基底 u_j との関係は次式で表される。

$$b_i = \frac{w_{i1}}{\sqrt{\lambda_1}} u_1 + \cdots + \frac{w_{iM}}{\sqrt{\lambda_M}} u_M \cdots \cdots \cdots (15)$$

この ICA 基底は以下のように正規化できる。

$$\hat{b}_i = (\frac{w_{i1}}{\sqrt{\lambda_1}} u_1 + \cdots + \frac{w_{iM}}{\sqrt{\lambda_M}} u_M) / C_i \cdots \cdots \cdots (16)$$

ただし、

$$C_i = \sqrt{(\frac{w_{i1}}{\sqrt{\lambda_1}})^2 + \cdots + (\frac{w_{iM}}{\sqrt{\lambda_M}})^2} \cdots \cdots \cdots (17)$$

である。 $\hat{b}_i$ と固有値 λ_j の PCA 基底 u_j との間の類似度は $\hat{b}_i^T u_j$ で与えられ、この類似度が大きければ大きいほど b_i の重要度 ν_i が PCA 基底 u_j の固有値 λ_j と同じ値を取るよう、 b_i の重要度 ν_i を次式で定義する。

$$\begin{aligned} \nu_i^2 &= \lambda_1^2 (\hat{b}_i^T u_1)^2 + \cdots + \lambda_M^2 (\hat{b}_i^T u_M)^2 \\ &= \frac{1}{C_i^2} \{ \lambda_1^2 (\frac{w_{i1}}{\sqrt{\lambda_1}})^2 + \cdots + \lambda_M^2 (\frac{w_{iM}}{\sqrt{\lambda_M}})^2 \} \\ &= \frac{\lambda_1 w_{i1}^2 + \cdots + \lambda_M w_{iM}^2}{C_i^2} \cdots \cdots \cdots (18) \end{aligned}$$

この重要度 ν_i によって ICA 特徴ベクトルの各成分 $\tilde{s}_i(k)$ を重み付けし、 $\tilde{s}'(k) = [\nu_1 s_1(k), \dots, \nu_M s_M(k)]^T$ に変換する。訓練データから得られた特徴ベクトルを参照ベクトルとし、テストデータから得られた特徴ベクトルとの間の距離尺度には方向余弦を用いる。このとき、ICA 特徴ベクトル $\tilde{s}(k)$ をそのまま用いたときの方向余弦を距離 A とし、各成分を ν_i で重み付けして得られる特徴ベクトル $\tilde{s}'(k)$ を用いたときの方向余弦を距離 B とする。これらの距離尺度を適用し、 k 近傍 (k -NN) 法を用いて認識を行う。なお、 k は 1, 3, 5 とし、その中の最高の認識率で評価する。

4・2・3 実験結果および考察 3 種類の多次元多クラスデータに対する認識結果を表 1(a)-(c) に示す。表 1(a)-(c) における距離 A の結果では、PCA 以外は手法による変動

[†] 式 (14) における Ψ の大きさがクラス数やクラス間距離などによって変わるため、 α の最適値はデータセットによってかなり異なる。

表 1 多次元多クラスデータに対する認識率 [%]

Table 1. Recognition rate [%] for multi-dimensional and multi-class data.

(a) breast-cancer-wisconsin		
	Distance A	Distance B
PCA	89.15	—
ICA	89.04	89.71
SICA-MC	89.23	89.37
SICA-MD	89.04	90.00
SICA-MD ($\mu = 0$)	89.04	89.01

(b) segmentation		
	Distance A	Distance B
PCA	78.94	—
ICA	86.33	86.42
SICA-MC	86.31	82.10
SICA-MD	86.36	87.85
SICA-MD ($\mu = 0$)	86.33	82.00

(c) waveform		
	Distance A	Distance B
PCA	77.86	—
ICA	64.12	65.08
SICA-MC	66.46	72.94
SICA-MD	64.14	78.28
SICA-MD ($\mu = 0$)	64.24	77.68

がほとんど見られないことがわかる。一方，距離 B の結果では，提案手法である SICA-MD で抽出された特徴量が最もよい認識性能を示した。これらの結果から，パターンのばらつきに基づいて基底を求める PCA に比べ，SICA-MD の方が有効であるといえる。また，式 (14) の μ を 0 としてクラス間距離のみを最大化したときと比べても，SICA-MD の有効性は明らかである。つまり，特徴量の独立性と分離性を同時に最大化することの有効性が示されたといえる。

次に，特徴量の独立性を高める効果に関して考察を行うため，パラメータ μ をいくらか変化し，特徴量の分離性にどのような影響を及ぼしているのか調べる。特徴量の独立性を評価するには，抽出された特徴量 $\tilde{s} = [\tilde{s}_1, \dots, \tilde{s}_M]$ の kurtosis を計算する。ここでは，次式で与えられる kurtosis の絶対値を独立性の評価値とする。

$$\text{kurt}(\tilde{s}) = \frac{1}{M} \sum_{i=1}^M \left| \frac{E[\tilde{s}_i^4]}{\{E[\tilde{s}_i^2]\}^2} - 3 \right|. \dots\dots\dots (19)$$

また，特徴量のクラス分離性の評価には，式 (11) の評価関数値を用いる。 μ の変化に伴う特徴量の特性に関するグラフを図 7 に示す。グラフの横軸には式 (19) で与えられる kurtosis の絶対値を示し，縦軸には距離 B を用いたときの認識率とクラス分離性を示した。グラフ中の点 (× または +) は右にあるものほど大きな μ に対する結果を表しており，ほとんどのケースで μ を大きくすれば kurtosis も大きくなる。図 7 の結果から，breast-cancer-wisconsin 以外のデータに関しては，独立性を高めることがクラス分離性の向上につながり，それが認識率の向上につながっている。ただし，独立性を高め過ぎると認識率は悪化し，クラ

図 7 独立性に対する認識率とクラス分離性の変化

Fig.7. Variations in recognition accuracy and class separability for independence.

表 2 基底の学習に要した CPU 時間 [sec.]

Table 2. CPU time [sec.] for learning base vectors.

	PCA	ICA	SICA-MC	SICA-MD
breast-cancer-wisconsin	0.1	8.1	61.2	52.2
segmentation	0.2	1.3	10.7	1460
wave	0.3	43.8	445	1692

ス分離性と独立性のバランスが重要であることがわかる。

最後に，基底の学習に要した CPU 時間を表 2 に示す。計算機の CPU は Pentium 4 (2.2GHz) である。表 2 からわかるように，segmentation と wave に対して SICA-MD の計算時間がかなり長くなっている。これは，式 (14) における Ψ の計算が訓練データ数や入力次元数に依存して多くなることと，3 種類の評価関数を同時に最小化するのに多くの繰り返しを要するためである。ただし，オンライン学習が必要な場合を除いて，基底の学習に多くの時間を要することが実用上問題になることは少ないと考えられる。なお，認識に必要な CPU 時間はどの手法でも同じである。

5. おわりに

ICA は高次統計量を無相関化することで独立な成分に分離する手法であり、パターン認識の特徴抽出法としても期待されている。しかし、ICA によって抽出された特徴量の有効性はデータセットに大きく依存するという問題点があった。これは ICA が教師なし学習であり、クラス情報を基底の学習に用いていないことが原因の一つと考えられる。そこで本論文では、独立性を高めると同時に特徴量の分離性を高める教師あり独立成分分析 (SICA-MD) を提案した。

提案手法の有効性を検証するため、まず人工データを用い、得られる基底ベクトルとその射影値の分布を調べた。その結果、SICA-MD によってクラス分離性に優れた軸が見つけれられることを確認した。また、UCI Machine Learning Repository⁽¹²⁾ に含まれる 3 種類のデータセットに対して特徴量の評価を行った。その結果、SICA-MD で得られた特徴量が他の特徴抽出法で得られた特徴量に比べて有効であることを確認した。また、独立性の評価である kurtosis の変化が特徴量の分離性にどのような影響を与えているかについて調べた。その結果、3 つのデータのうち 2 つのデータに関しては独立性を高めることによってクラス分離性に優れた特徴量を抽出できることを示した。

本稿で提案した SICA-MD は、今のところ、あらゆるデータセットに対して有効性を示している訳ではなく、ICA で得られる特徴量がデータセットに大きく依存する問題を完全に解決できたわけではない。しかしながら、教師を導入することで有効性を示せるデータセットが増えたことは事実であり、クラス情報の与え方や評価関数の改良によって、今後さらに改善が見込まれる。また、特徴量の独立性を高めることが、どのようなデータセットに対して有効なのかを明らかにすることが重要であると考えている。

本研究に関して、日頃から有益な議論をして頂いている神戸大学阿部重夫教授に感謝します。

(平成 14 年 10 月 23 日受付、平成 15 年 6 月 24 日再受付)

文 献

- (1) M. Girolami: Self-organizing neural networks, Springer, (1999)
- (2) A. Hyvärinen, J. Karhunen, and E. Oja: Independent component analysis, Wiley-Interscience, (2001)
- (3) J. Karhunen, A. Hyvärinen, R. Vigar, J. Hurri, and E. Oja: "Applications of neural blind separation to signal and image processing", *Proc. IEEE Int. Conf. Acoust. Speech Signal Processing*, pp.131-134 (1997)
- (4) M. Kotani, S. Maekawa, S. Ozawa, and K. Akazawa: "Signal processing of speech using independent component analysis based on information maximization algorithm", *T. SICE*, Vol.36, No.5, pp.456-458 (2000) (in Japanese)
小谷 学・前川 聡・小澤誠一・赤澤堅造:「情報量最大化アルゴリズムに基づく独立成分分析を用いた音声信号処理」, 計測自動制御学会論文集, 36, 5, pp.456-458 (2000)
- (5) M. Kotani, Y. Shirata, S. Maekawa, S. Ozawa, and K. Akazawa: "Representations of speech by sparse coding algorithm", *IEEE Trans. EIS*, Vol.120, No.12, pp.1996-2002 (2000-12) (in Japanese)
小谷 学・白田康伸・前川 聡・小澤誠一・赤澤堅造:「スパースコーディングによる音声の表現」, 電学論 C, 120, 12, pp.1996-2002 (2000-12)
- (6) M. Kotani and S. Ozawa: "Feature extraction of digit patterns using independent component analysis", *IEEE Trans. EIS*, Vol.122, No.3, pp.465-470 (2002-3) (in Japanese)
小谷 学・小澤誠一:「独立成分分析を用いた数字パターンの特徴抽出」, 電学論 C, 122, 3, pp.465-470 (2002-3)
- (7) 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)
- (8) S. Ozawa and M. Kotani: "A study of feature extraction and selection using independent component analysis", *Proc. of 7th Int. Conf. on Neural Info. Processing*, I, pp.369-374 (2000)
- (9) S. Umeyama, S. Akaho, and Y. Sugase: "Supervised independent component analysis and its applications to face image analysis", *Tech. Report of IEICE, NC99-2*, pp.9-16 (1999) (in Japanese)
梅山伸二・赤穂昭太郎・菅生康子:「教師付き独立成分分析の提案とその顔画像分析への応用について」, 信学技報, NC99-2, pp.9-16 (1999)
- (10) Y. Watanabe, M. Hirahara, and T. Nagano: "Feature extraction of palm prints using supervised independent component analysis", *CD-ROM Proc. of the 7th Int. Conf. on Neural Info. Processing*, (2000)
- (11) S. Ozawa, Y. Sakaguchi, and M. Kotani: "A study of feature extraction using supervised independent component analysis", *Proc. of Int. Conf. on Neural Networks*, pp.2958-2963 (2001)
- (12) <http://www.ics.uci.edu/mllearn/MLRepository.html>
- (13) J. Karhunen and E. Oja: "A class of networks for independent component analysis", *IEEE Trans. on Neural Networks*, 8, pp.486-503 (1997)
- (14) 石井健一郎・他: パターン認識, 共立出版, (1998)

坂 口 善 規 (非会員) 2001 年, 神戸大学工学部電気電子工学科卒業, 同年 4 月神戸大学自然科学研究科博士前期課程電気電子工学専攻入学, 現在に至る。パターン認識の特徴抽出に関する研究に従事。

小 澤 誠 一 (正員) 1989 年神戸大学大学院工学研究科修士課程 (計測工学専攻) 修了。同年奈良工業高等専門学校情報工学科助手, 1993 年大阪教育大学教育学部教養学科情報科学専攻助手, 同講師を経て, 2000 年 1 月より神戸大学大学院自然科学研究科助教授となり, 現在に至る。ニューラルネットによる知的情報処理の研究に従事。博士 (工学) 電子情報通信学会, 計測自動制御学会, システム制御情報学会, IEEE, INNS の会員。

小 谷 学 (正員) 1983 年神戸大学大学院工学研究科修士課程 (計測工学専攻) 修了。同年 (株) 神戸製鋼所入社。1989 年神戸大学工学部学計測工学科助手を経て, 同情報知能工学科助教授となり, 現在に至る。電子的計測, 知能システムなどの研究に従事。博士 (工学) 電子情報通信学会, 計測自動制御学会, システム制御情報学会, 日本神経回路学会等の会員。