グラフィクスとＣＡＤ　48－8
（1990. 12. 14）

マハラノビス距離を用いた判別ベクトルの組合せによる特徴抽出法

松浦　豊 [††]　浜本　義彦 [†]　内村　俊二 [†††]
金岡　泰保 [†]　富田　真吾 [†]

[†] 山口大学　工学部　[††] 日本電気（株）　[†††] 大島商船高等専門学校

あらまし　本論文では、特徴の組合せを考慮に入れた特徴抽出法を論じた。本手法は、予めｎ個の正規直交判別ベクトルを求め、それらの中から選択される正規直交判別ベクトルによって生成される系をマハラノビス距離の大きさに基づいて順序づけ、上位の系についてのみエラー確率の評価を行い、最小エラー確率となる特徴空間を求めるものである。実在データに対するシミュレーション結果から、本手法が正規直交判別ベクトル法、Karhunen-Loeve展開よりエラー確率の低い特徴空間を与えることが示された。

Feature Extraction Method based on the Combination of Discriminant Vectors Using the Mahalanobis distance

[††] Yutaka Matsuura　[†] Yoshihiko Hamamoto　[†††] Shunji Uchimura
[†] Taiho Kanaoka　[†] Shingo Tomita

[†] Yamaguchi University
Ube, 755

[††] NEC Corporation
5-33-1, Shiba, Minato-ku, 108

[†††] Oshima National College of Maritime Technology
Oshima, 742-21

Abstract　In this paper, we propose a feature extraction method based on the combination of discriminant vectors. One way of evaluating the combination is the well-known exhaustive search. From the standpoint of computational complexity, however, exhaustive search is impractical in general. The Mahalanobis distance plays an important role in reducing computational complexity due to the evaluation of the combination. From experimental results, we show that the proposed method is more powerful than the Karhunen-Loeve expansion and orthonormal discriminant vector method in terms of the error probability.

1.まえがき

パターン認識系における特徴抽出系の目的は、一般に高次元であるパターン空間から低次元である特徴空間への次元圧縮をできるだけ識別能力を損なわないように行うことである。特徴抽出系の設計法は、特徴選択法と特徴抽出法とに大別される[5]。特徴選択法は与えられた特徴の中から識別に有用な少数の特徴を選択するもので、特徴抽出法はパターン空間とほぼ同等な識別能力を有する特徴空間を生成する基底（特徴）を求めるものである。特徴選択法、特徴抽出法いずれの手法も、特徴の評価関数の設定およびその最適化という二つの問題を解かなければならない。特徴選択法における評価関数の最適化問題に関して、Cover[1]は単独で最も認識率の高い特徴を組み合わせても最適な2次元特徴空間が得られない例を示した。更に、Coverら[2]は逐次形特徴選択法の限界を理論的に明らかにし、最適解を得るためにはすべての特徴の組合せを評価しなければならないことを示した。一方、特徴抽出法については、これまで評価関数の最適化という観点からの十分な議論はなされていない。実験的立場から、評価関数の最適化を逐次的に行う正規直交判別ベクトル法[3]を通して逐次形特徴抽出法の問題点を示す結果が報告された[4]。文献（4）の手法は、修正されたプラスeマイナスf法により特徴の組合せを考慮に入れて特徴抽出を行うもので、正規直交判別ベクトル法よりエラー確率の低い特徴空間を与えた。従って、特徴の組合せを考慮に入れることにより逐次形特徴抽出法を改善することができるものと考えられる。

本論文は、文献（4）とは異なる立場から特徴の組合せを考慮に入れるもので、組合せの評価に要する計算量をマハラノビス距離に基づいて低減する特徴抽出法を提案している。

2.で特徴を評価する確率的距離として知られている、マハラノビス距離、ダイバージェンス、バッタチャヤ距離の優劣をエラー確率の観点から調べ、マハラノビス距離が三つの距離の中で有効であることを実験的に示す。3.では、マハラノビス距離に基づいて特徴の組合せを考慮に入れた特徴抽出法を提案し、その有効性を他の特徴抽出法との比較を通して示す。

2．確率的距離の選定

2．1 準備

まず、諸記号を以下のように定義する。

R^n：n次元パターン空間

m：クラス数

$p(i)$：クラスiの事前確率

$\boldsymbol{x}$：n次元パターンベクトル

$\boldsymbol{\mu}_i^*$：クラスiの平均ベクトル

W_i^*：クラスiのクラス内共分散行列

本手法は、予めn個の正規直交判別ベクトル（以下、判別ベクトルと呼ぶ）を求め、それらの中から選択される判別ベクトルによって生成される系の中でエラー確率の最も低い系（特徴空間）を求めるものである。その際、パターン空間の次元数と特徴空間の次元数によっては判別ベクトルの組合せの数が膨大となるため、生成されるすべての系に対してエラー確率の評価を行うことは計算時間の観点から非実用的となる。本論文では、確率的距離の大きさに基づいて系を順序づけ、上位の系についてのみエラー確率の評価を行うことにより計算時間を低減する手法を提案する。

代表的な確率的距離として、マハラノビス距離、ダイバージェンス、バッタチャヤ距離が知られている[5]。確率的距離は、ある種のクラス間距離であり、パターン分布間の差異を識別情報として表す。従って、確率的距離が大きい系ほど識別が容易であると考えられる。しかしながら、確率的距離とエラー確率との関係は一意的でないため、確率的距離の最も大きい系が最小のエラー確率を与えるとは限らない。これまで確率的距離とエラー確率との関係は、上限または下限という立場からあるいは2クラス問題で共分散行列が等しい場合など特定の分布に対する立場からしか議論されていない[6]。

理論的な立場からの議論は不十分であるが、これまでの実験結果から一般に最小エラー確率を与える系の順位は上位に位置する。従って、いずれの確率的距離を用いるかは最小エラー確率を与える系の順位に着目することによってなされるべきである。本論文では、実験的な立場からこの問題を考察する。

2．2 実験1

三つの確率的距離の優劣を調べるため、以下の実験を行った。まず、実験方法を述べる。いま、s個の判別ベクトル $\{\boldsymbol{d}k_1, \boldsymbol{d}k_2, \cdots, \boldsymbol{d}k_s\}$ によって生成されるs次元の系をR_sと表す。このR_sは次

式によって定義される。

$$R_s = \{\hat{\boldsymbol{x}} \mid \hat{\boldsymbol{x}} = \Phi^T \boldsymbol{x},\ \boldsymbol{x} \in R^n\} \tag{1}$$

ここで、変換行列Φは

$$\Phi = [\boldsymbol{d}k_1, \boldsymbol{d}k_2, \cdots, \boldsymbol{d}k_s],\quad 1 \leqq k_1 < \cdots < k_s \leqq n$$

で与えられ、$\Phi^T\Phi = I$なる関係を満たす。但し、Iは単位行列である。

このとき、この系では、$\boldsymbol{\mu}_i^*$、W_i^*はそれぞれ

$$\boldsymbol{\mu}_i = \Phi^T \boldsymbol{\mu}_i^* \tag{2}$$

$$W_i = \Phi^T W_i^* \Phi \tag{3}$$

で与えられる。正規分布を仮定すると上述のマハラノビス距離J_M、ダイバージェンスJ_D、バッタチャヤ距離J_Bはそれぞれ

$$J_M = (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T (p(1)W_1 + p(2)W_2)^{-1} \cdot (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) \tag{4}$$

$$J_D = 0.5(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T (W_1^{-1} + W_2^{-1})(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) + 0.5\,tr(W_1^{-1}W_2 + W_2^{-1}W_1 - 2I) \tag{5}$$

$$J_B = 0.25(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T (W_1 + W_2)^{-1} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) + 0.5\ln\left[\frac{|0.5 \cdot (W_1 + W_2)|}{(|W_1||W_2|)^{0.5}}\right] \tag{6}$$

で与えられる[5]。

n個の中からs個の判別ベクトルを選択することから${}_nC_s$個のs次元の系R_sが候補として考えられ、それぞれの系に対して確率的距離を計算し、得られる確率的距離の値に基づいて系を順序づける。ここで、第i番目の系のマハラノビス距離の値を$J_M(s、i)$と表し、同様にダイバージェンスの値を$J_D(s、i)$、バッタチャヤ距離の値を$J_B(s、i)$と表す。確率的距離の値は用いるデータや特徴空間の次元数sに依存するため、次のような正規化を行う。

$$J_M^*(s, i) = J_M(s, i) / J_M(s, 1) \tag{7}$$

$$J_D^*(s, i) = J_D(s, i) / J_D(s, 1) \tag{8}$$

$$J_B^*(s, i) = J_B(s, i) / J_B(s, 1) \tag{9}$$

これにより、確率的距離の値は

$$0 \leqq J_M^*(s, i) \leqq 1 \tag{10}$$

$$0 \leqq J_D^*(s, i) \leqq 1 \tag{11}$$

$$0 \leqq J_B^*(s, i) \leqq 1 \tag{12}$$

となる。以下、正規化された確率的距離のみを用いる。マハラノビス距離により順序づけられる第i番目の系を$R_s(J_M^*、i)$、同様に $R_s(J_D^*、i)$、$R_s(J_B^*、i)$と表す。

エラー確率の推定法として種々の手法が提案されているが、推定されたエラー確率の不偏性と識別系の多クラス問題への対応の容易さとから、識別系として最近傍則を用いたLeave-one-out法[5]、[6]を採用した。

次に、用いたデータについて述べる。本実験では、実在データとしてFossilデータ[7]とCh.データ[8]を、人工データとしてMarill-Greenデータ[9]を用いた。以下、各データの説明を簡単に述べる。まずFossilデータは、3クラスの6次元データで、各クラスのデータ数はクラス1が40個、クラス2が34個、クラス3が13個である。このデータは、特徴抽出法の評価データとして用いられてきた。Ch.データは、3クラスの6次元データで、各クラスのデータ数はクラス1が21個、クラス2が31個、クラス3が22個である。このデータは、識別関数の評価データとして用いられてきた。Marill-Greenデータは、4クラスの8次元データで、各クラスのデータ数は200個である。このデータは、種々の文献で引用されてきたものである。

実在データに対し、$n=6$、$s=3$の実験結果を図1，2に示す。${}_6C_3$であるから20個の系を評価した。図中の黒点は最小エラー確率を与える系を示す。 Fossilデータでは、20個の系の中でマハラノビス距離が最小エラー確率の系を4位に順序づけているのに対し、ダイバージェンス、バッタチャヤ距離はそれぞれ14位、3位であった。Ch.データについては、20個の系の中でマハラノビス距離が1位、ダイバージェンス、バッタチャヤ距離も1位であった。一方、Marill-Greenデータについては人工データであるため、モンテカルロ法によりシミュレーションを行った。すなわち、120回独立にデータを発生させ、それぞれについて$n=8$、$s=4$で確率的距離とエラー確率を求めた。その結果を表1に示す。${}_8C_4$であるから70個の系が評価され、マハラノビス距離は、最小エラー確率の系を最高1位、最低6位と順序づけ、順位の平均値が1．192であった。これに対し、ダイバージェンスは、最高1位、最低10位、順位の平均値が2．183であり、バッタチャヤ距離は、最高1位、最低8位、順位の平均値が1．250であった。

以上の実験結果から、三つの確率的距離の中でマハラノビス距離が有効であることが分かった。

図1－A　Fossilデータに対する結果

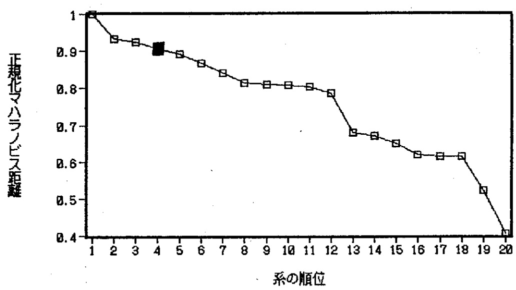


図2－A　Ch.データに対する結果

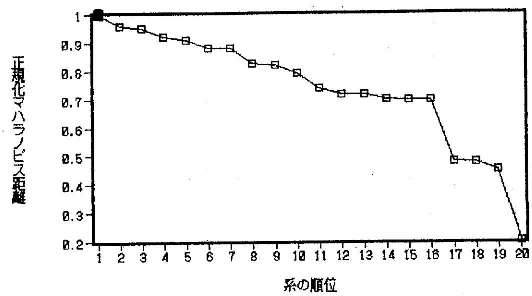


図1－B　Fossilデータに対する結果

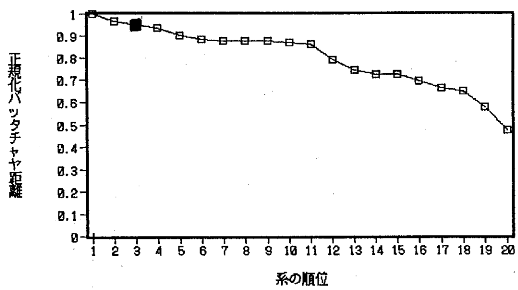


図2－B　Ch.データに対する結果

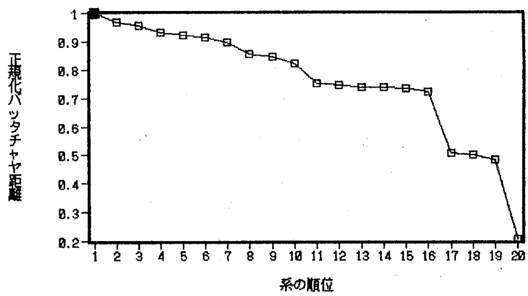


図1－C　Fossilデータに対する結果

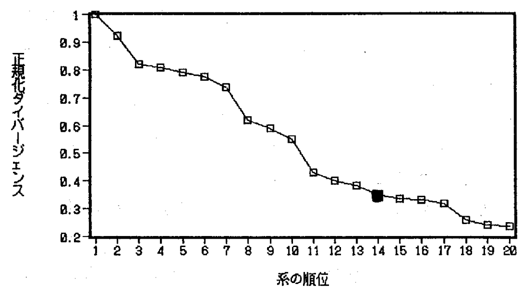


図2－C　Ch.データに対する結果

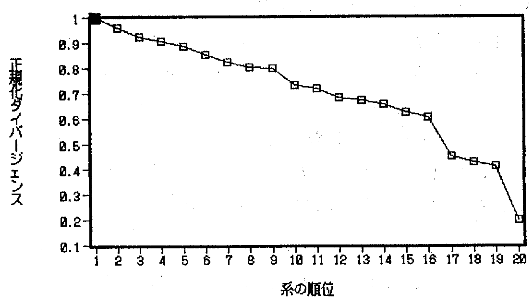


表1　Marill-Greenデータに対する結果

	最高順位	最低順位	順位の平均値
マハラノビス距離	1	6	1.192
バッタチャヤ距離	1	8	1.250
ダイバージェンス	1	10	2.183

3．マハラノビス距離を用いた特徴抽出法

3．1　アルゴリズム

以下、本手法のアルゴリズムを示す。

手順1：正規直交判別ベクトル法により抽出されるn個の判別ベクトル $\{\boldsymbol{d}_1, \boldsymbol{d}_2, \cdots, \boldsymbol{d}_n\}$ を特徴候補とする。

手順2：特徴候補から選択されたs個の判別ベクトル $\{\boldsymbol{d}_{k_1}, \boldsymbol{d}_{k_2}, \cdots, \boldsymbol{d}_{k_s}\}$ で生成される系のマハラノビス距離を計算する。

手順3：得られたマハラノビス距離を正規化する。

手順4：正規化されたマハラノビス距離に基づいて系を順序づける。

手順5：$J_M^*(s, i) \geq J_M^*(s, 1) - \varepsilon$ を満たす系 $R_s(J_M^*, i)$ についてエラー確率を推定する。

手順6：推定された中で最小エラー確率となる系を求め、それを生成するs個の判別ベクトルを第1特徴から第s特徴 $\{\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \cdots, \boldsymbol{\tau}_s\}$ とする。

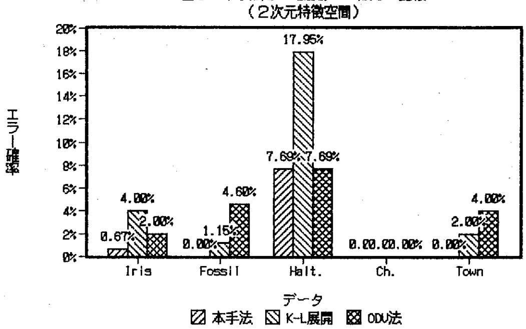
パラメータ ε $(0 \leq \varepsilon \leq 1)$ の値が小さくなると、一般に計算時間は減少するが、すべての特徴の組合せの中で最小エラー確率となる系を選択できない場合がある。それゆえ、ε の値は重要である。２.２で述べた実在データと人工データの実験結果から、$\varepsilon = 0.2$ とすれば、すべての特徴の組合せの中で最小エラー確率となる系を選択することができる。従って、本論文では $\varepsilon = 0.2$ とした。

3．2　実験2

本手法の有効性を文献（３）の手法（以下ODV法と呼ぶ）、Karhunen-Loeve展開（K-L展開）[11]との比較を通して実験的に示す。

用いたデータは、Irisデータ[10]、Fossilデータ、Halt.データ[8]、Ch.データ、Townデータ[7]で、これらはすべて実在データである。Irisデータは、よく知られた３クラスの４次元データである。各クラスのデータ数は５０個である。Halt.データは、２クラスの４次元データで、各クラスのデータ数はクラス１が１９個でクラス２が２０個である。Townデータは、４クラスの４次元データで、各クラスのデータ数はクラス１が１０個、クラス２が１０個、クラス３が１６個、クラス４が１４個である。これら実在データを２次元へ次元圧縮し、２次元空間上でエラー確率の比較を行った。その結果を図３に示す。

図3　本手法とK-L展開，ODV法との比較（2次元特徴空間）



いずれの実験結果も本手法が他の特徴抽出法と比べ優れていることを示している。ここで、全データについて本手法がすべての特徴の組合せの中で最小エラー確率となる系を選択したということを述べておく。これにより、$\varepsilon = 0.2$ の妥当性が示された。

本手法は、特徴抽出問題で指摘された逐次形特徴抽出法の問題点を改善するもので、まず特徴のすべての組合せから得られる系をマハラノビス距離により順序づけ、次に上位の系についてのみエラー確率により詳細に系の有効性を評価している。これは、手書き漢字認識手法でよく用いられている、大分類、詳細識別の２段階で認識を行うというアプローチを特徴抽出問題に適用したものである。計算時間は、ほとんどがエラー確率の推定に要する時間で、特徴空間の次元数sの増加とともに急速に増加する。従って、本手法は低次元特徴空間の設計に用いられるものと考えられる。例えば、本手法は、対話形パターン解析システムの２次元ディスプレイ面を生成する座標軸を提供する。この対話形パターン解析システムは、多次元データを２次元ディスプレイ面上に表示し、設計者がその２次元パターン分布を見ながらパターン識別器の設計またはデータ解析を行うことができるものである。設計者の知見がよく反映されるため、有効なシステムとして知られている。

4．むすび

本論文では、特徴の組合せを考慮に入れた特徴抽出法を論じた。まず、マハラノビス距離、ダイバージェンス、バッタチャヤ距離の優劣を実在データと人工データから得られたエラー確率を基に検討し、三つの確率的距離の中でマハラノビス距離が有効であることを示した。この結果を基に、特徴の組合せ

の評価に要する計算量をマハラノビス距離により低減する特徴抽出法を提案した。実在データを用いたシミュレーション結果から、本手法は逐次形特徴抽出法より低いエラー確率を与える特徴空間を生成する、ということを示した。

本論文では、特徴として判別ベクトルを用いて議論を進めたが、他の特徴についても同様な議論を行うことができる。特徴の組合せの有効性を理論的な立場から論じることは、今後の課題である。

文献

(1)T.M.Cover:" The best two independent measurements are not the two best", IEEE Trans. Syst. Man & Cybern., SMC-4, 1, pp. 116-117(1974).

(2)T.M.Cover and J.M.Van Campenhout:" On the possible orderings in the measurement selection problem", IEEE Trans. Syst. Man & Cybern., SMC-7, 9, pp.657-661(1977).

(3)岡田、富田：”正規直交判別ベクトルによる特徴抽出論”、信学論（Ａ）、 Ｊ６５－Ａ、8, pp.767-771(1982).

(4)Y.Hamamoto, Y.Matsuura, T.Kanaoka and S. Tomita:" A note on the orthonormal discriminant vector method for feature extraction", to be published in Pattern Recognition.

(5)A.K.Jain:" Advances in statistical pattern recognition", in Pattern Recognition Theory and Applications, ed., P.A.Devijver, J.Kittler, pp.1-19(1987).

(6)K.Fukunaga:" Introduction to statistical pattern recognition", Academic Press(1972)

(7)Y.Chien:" Interactive pattern recognition", Marcel Dekker INC.(1978).

(8)A.A.Lubischew:" On the use of discriminant functions in taxonomy", Biometrics, pp.455-477(1962).

(9)T.Marill and D.M.Green:" On the effectiveness of receptors in recognition systems", IEEE Trans. Information Theory IT-9, pp.11-27(1963).

(10)R.A.Fisher:" The use of multiple measurements in taxonomic problems" Annals of Eugenics, Vol.7, pp.179-188(1936).

(11)S.Watanabe:" Karhunen-Loeve expansion and factor analysis", Trans. 4th Prague Conf. on Information Theory(1965).