

# ベイズアプローチによる最適識別系の有限標本効果に関する考察 ——学習標本の大きさがクラス間で異なる場合——

韓 雪仙<sup>†</sup> 若林 哲史<sup>†</sup> 木村 文隆<sup>†</sup> 三宅 康二<sup>†</sup>

On the Limited Sample Effect of the Optimum Classifier by Bayesian Approach  
——The Case of Independent Sample Size for Each Class——

Xuexian HAN<sup>†</sup>, Tetsushi WAKABAYASHI<sup>†</sup>, Fumitaka KIMURA<sup>†</sup>, and Yasuji MIYAKE<sup>†</sup>

あらまし 学習標本の大きさが有限の場合、2次識別関数の識別性能は特徴ベクトルの次元数の増加に伴って劣化しピーキング現象を起こすことが知られている。このような問題に対処するために、ベイズアプローチによって導かれる未知の分布に対する最適識別関数の平均誤り確率等を理論的・実験的に評価して、学習標本の大きさがクラス間で等しい場合と異なる場合の最適識別系の有効性を明らかにする。その結果、学習標本の大きさがほぼ等しい場合は、最適識別関数及び最適識別関数から導かれる擬似ベイズ識別関数、改良投影距離、投影距離を用いる識別系ではピーキング現象が起きないことがわかった。また、学習標本の大きさが異なる場合は、最適識別関数及び擬似ベイズ識別関数ではピーキング現象が起きないが、2次識別関数、改良投影距離、投影距離を用いる識別系ではピーキング現象が起きることがわかった。

キーワード ベイズアプローチ、ベイズ学習、最適識別関数、投影距離、平均誤り確率、手書き数字認識

## 1. ま え が き

18世紀の中ごろに発見されたベイズの定理を利用して、観測結果からその結果をもたらした母集団の性質を推測する古典的推測法(ベイズアプローチ)が生まれた[1]。その後顧みられない時代があったが、今世紀の後半になってコンピュータによる人工知能やパターン認識の研究が始まるとベイズ統計学が見直されるようになった。学習標本を用いて事前分布から事後分布を逐次的に求める方法は、ベイズ学習と呼ばれる学習の一つのモデルとなった。

本論文では、対象とする分布が正規分布であることがわかっているが、その共分散行列が未知の場合について考察する。このような場合に用いられるのは、学習標本を用いて母集団の平均ベクトルと共分散行列を推定し、それを用いて2次識別関数を構成する最大ゆがみアプローチである。しかし、学習標本の大きさが一定の場合、2次識別関数の識別性能は、特徴ベクトルの次元数の増加に伴って劣化しピーキング現象を起

こすことが知られている。一方、ベイズアプローチによる最適識別系はピーキング現象を起こさないことが証明されている[2]~[5]。

本論文では、有限標本に基づくベイズアプローチによって導かれる未知の分布に対する最適識別関数[6]の平均誤り確率等を理論的・実験的に評価して、学習標本の大きさがクラス間で等しい場合と異なる場合の最適識別系の有効性を明らかにする。

ベイズアプローチと最大ゆがみアプローチの有効性については、AitchisonとDunsmoreによる希少疾患のスクリーニングにおける小規模な比較研究等があるが、一般にベイズアプローチの有効性は十分に認められておらず[7]、文献[8]には、学習標本の大きさがクラス間で異なる場合には、文献[6]の最適識別関数は有効でないと記述されている。

## 2. ベイズアプローチと最大ゆがみアプローチ

### 2.1 ベイズアプローチ

ベイズアプローチでは、未知のパラメータ $\theta$ (平均や分散・共分散)を確率変数として扱い、次式を用い

<sup>†</sup> 三重大学工学部情報工学科、津市

Faculty of Engineering, Mie University, Tsu-shi, 514-8507  
Japan

て、未知パラメータを消去する．

$$p(X|\chi) = \int p(X|\theta)p(\theta|\chi)d\theta \quad (1)$$

ここで、 $p(X|\chi)$  は標本  $\chi$  が与えられたときの特徴ベクトル  $X$  の密度関数、 $p(X|\theta)$  は未知パラメータ  $\theta$  を含む  $X$  の密度関数、 $p(\theta|\chi)$  は  $\theta$  の密度関数である． $p(\theta|\chi)$  はパラメータ  $\theta$  の事前分布  $p(\theta)$  を仮定し、ベイズ学習（逐次ベイズ推定）により求める [9]．

ベイズアプローチには次のような長所と短所がある．

(1) パラメータが未知の分布に対する最適識別系が構成できる．

(2) パラメータの事前分布  $p(\theta)$  を仮定する必要がある．

### 2.2 最大ゆう度アプローチ

統計的パターン認識で一般的に用いられるのは最大ゆう度アプローチである．最大ゆう度アプローチは、未知のパラメータを未知の定数（確率変数ではない）と考え、未知パラメータをその最ゆう推定量で置き換えて、特徴ベクトル  $X$  の密度関数を求める．

$$p(X) \approx p(X|\theta)_{\theta \approx \hat{\theta}(\chi)} \quad (2)$$

ここで、 $\theta$  は未知パラメータ、 $\hat{\theta}(\chi)$  は標本  $\chi$  に基づく、 $\theta$  の最ゆう推定量である．

最大ゆう度アプローチには次のような長所と短所がある．

(1) パラメータの事前分布を仮定する必要がない．

(2) 最ゆう推定量  $\hat{\theta}(\chi)$  の推定誤差のために、特徴ベクトル  $X$  の次元数がある次元数を超えると認識性能が劣化する（ピーキング現象）．

(3) 学習標本の大きさがクラス（字種）によって異なる場合にも性能が劣化する [10]．

## 3. ベイズアプローチから導かれる識別関数

### 3.1 最適識別関数

#### 3.1.1 正規母集団の条件付密度関数

母集団の平均ベクトルが既知で、共分散行列が未知の場合、有限標本が与えられた条件下での母集団の条件付確率密度関数は（多変量） $t$  分布となる．

(1) 1次元の場合

式 (1) において、 $p(X|\theta)$  が正規分布、 $p(\theta|\chi)$  がカイ 2 乗分布となり、 $p(X|\chi)$  は  $t$  分布となる．

(2) 多次元の場合

同様に  $p(X|\theta)$  が正規分布、 $p(\theta|\chi)$  がウィシャート

分布となり、 $p(X|\chi)$  は次式で与えられる [6]．

$$p(X|\chi) = (N_N \pi)^{-\frac{n}{2}} |\Sigma_N|^{-\frac{1}{2}} \frac{\Gamma(\frac{N_N+1}{2})}{\Gamma(\frac{N_N-n+1}{2})} \left\{ 1 + \frac{(X-M)^T \Sigma_N^{-1} (X-M)}{N_N} \right\}^{-\frac{N_N+1}{2}} \quad (3)$$

ただし、

$$N_N = N + N_0$$

$$\Sigma_N = (1 - \alpha)\Sigma + \alpha\Sigma_0$$

$$\alpha = \frac{N_0}{N + N_0}$$

ここで、 $X$  は未知パターンの特徴ベクトル、 $N$  は各クラスの学習標本の大きさ、 $n$  は特徴ベクトルの次元数、 $M$  は母集団の平均ベクトル、 $\Sigma$  は標本共分散行列、 $\Sigma_0$  は  $X$  の事前分布の共分散行列で、 $N_0$  は  $\Sigma_0$  の信頼度を表す定数であり、 $\Gamma$  はガンマ関数である． $p(X|\chi)$  は簡単な変数変換により多変量  $t$  分布となる．

#### 3.1.2 最適識別関数

式 (3) より、最適識別関数は次式で与えられる．

$$g(X) = -2 \ln p(X) P(\omega) = (N_N + 1) \ln \left\{ 1 + \frac{(X-M)^T \Sigma_N^{-1} (X-M)}{N_N} \right\} + \ln |\Sigma_N| - 2 \ln D - 2 \ln P(\omega) \quad (4)$$

ただし、

$$D = (N_N \pi)^{-\frac{n}{2}} \frac{\Gamma(\frac{N_N+1}{2})}{\Gamma(\frac{N_N-n+1}{2})}$$

ここで、 $P(\omega)$  は、クラス  $\omega$  の事前確率である．

最適識別関数において、形式的に  $N_0 = 0$  とすると、 $N_N = N$ 、 $\Sigma_N = \Sigma$  となって  $\Sigma_0$  は不要となる． $N_0 \neq 0$  の場合には  $\Sigma_0$  が必要となるが、その場合は、特徴ベクトル  $X$  の事前分布として、球状分布を仮定して  $\Sigma_0 = \sigma^2 I$  とする． $I$  は単位行列であり、 $\sigma^2$  は球状分布の分散である．4.1、4.2 の性能評価では  $N_0 = 0$  とする．また 4.3 の性能評価では  $N_0 \neq 0$  として、 $\sigma^2$  の値として全字種、全固有値の平均を用いる．また、 $M$  は標本平均ベクトルで代用する．

#### 3.2 擬次ベイズ識別関数

計量量・記憶容量を削減するために式 (4) を変形す

ると、次式の擬似ベイズ識別関数 [11] が得られる [付録 1.] .

$$\begin{aligned}
 g(X) &= (N + N_0 + 1) \ln \left\{ 1 + \frac{1}{N_0 \sigma^2} \left[ \|X - M\|^2 \right. \right. \\
 &\quad \left. \left. - \sum_{i=1}^k \frac{(1 - \alpha) \lambda_i}{(1 - \alpha) \lambda_i + \alpha \sigma^2} \left[ \Phi_i^T (X - M) \right]^2 \right] \right\} \\
 &\quad + \sum_{i=1}^k \ln \left( (1 - \alpha) \lambda_i + \alpha \sigma^2 \right) - 2 \ln P(\omega)
 \end{aligned} \tag{5}$$

ここで、 $\lambda_i$ 、 $\Phi_i$  はそれぞれ共分散行列  $\Sigma$  の第  $i$  固有値と第  $i$  固有ベクトル、 $k$  は識別に用いる固有ベクトルの数である .

### 3.3 改良投影距離・投影距離法

すべてのクラスで  $|\Sigma_N|$ 、 $P(\omega)$ 、 $N$ 、 $N_0$  が等しいと仮定すると、式 (5) の第 2 項、第 3 項は定数項となり、第 1 項の係数  $(N + N_0 + 1)$  はクラス間で共通となるので、次式で定義される改良投影距離 (擬似マハラノビス距離) によって、擬似ベイズ識別関数を近似することができる .

$$\begin{aligned}
 g(X) &= \|X - M\|^2 \\
 &\quad - \sum_{i=1}^k \frac{(1 - \alpha) \lambda_i}{(1 - \alpha) \lambda_i + \alpha \sigma^2} \left\{ \Phi_i^T (X - M) \right\}^2
 \end{aligned} \tag{6}$$

更に  $i \leq k$  においては、大部分の固有値が  $\lambda_i \gg (N_0/N)\sigma^2$  を満たすため式 (6) は次式で定義される投影距離 [12] で近似できる .

$$\begin{aligned}
 g(X) &= \sum_{i=k+1}^n \left\{ \Phi_i^T (X - M) \right\}^2 \\
 &= \|X - M\|^2 - \sum_{i=1}^k \left\{ \Phi_i^T (X - M) \right\}^2 \tag{7}
 \end{aligned}$$

この識別関数は特徴ベクトル  $X$  を共分散行列の最初の  $k$  個の主要固有ベクトルで K-L 展開したときの 2 乗誤差であり、 $X$  からこれらの固有ベクトルで張られる超平面への距離の 2 乗である . この超平面は各クラスの標本の分布を近似する最小 2 乗誤差超平面となっている .

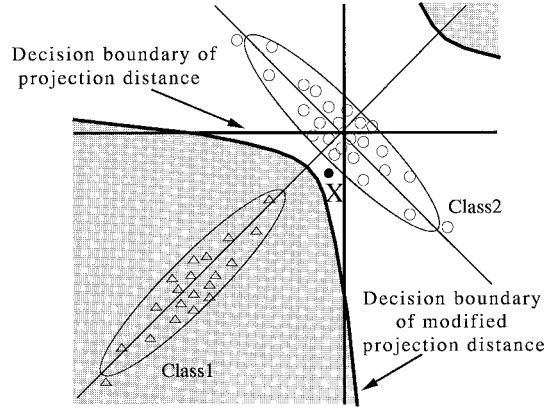


図 1 投影距離と改良投影距離の決定境界

Fig.1 Decision boundaries of projection distance and modified projection distance.

図 1 は、投影距離と改良投影距離の決定境界の例を示す . この図では特徴空間の次元数が 2 で各クラスの分布はその長軸 (1 次元の超平面) で近似されている . 投影距離による分類においては、未知パターン  $X$  は各々のクラスの長軸までの距離の短い方のクラスに分類される . その結果、投影距離の決定境界は図の直線のようになり、クラス 2 の一部が誤認識される . 改良投影距離の決定境界は固有値によって決まる双曲線になり、超平面の交差部分 (共通部分空間) とその近傍における誤認識を避けることができる .

## 4. 性能評価

本章では、最適識別関数の平均誤り確率等を理論的・実験的に評価して、学習標本の大きさがクラス間で等しい場合と異なる場合の最適識別系の有効性を明らかにする . 4.1 では 1 変数の最適識別関数と 2 次識別関数の平均誤り確率を理論的に評価・比較し、4.2 では多変量 (8 変量) の場合に、モンテカルロシミュレーションによる評価・比較を行う . 更に 4.3 では、手書き数字の認識実験により 3. に述べた各識別関数・手法の評価・比較を行う .

最大ゆう度アプローチから導かれる識別関数としては、次式で与えられる 2 次識別関数を用いる .

$$\begin{aligned}
 g(X) &= (X - M)^T \Sigma^{-1} (X - M) \\
 &\quad + \ln |\Sigma| - 2 \ln P(\omega)
 \end{aligned} \tag{8}$$

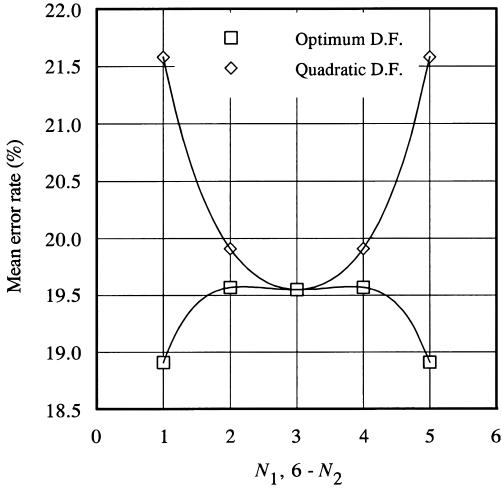


図2 総標本の大きさが一定の場合の平均誤り確率の理論値 (%) ( $N_1 + N_2 = 6, \sigma_1^2 = \sigma_2^2 = 1.0$ , 1変数の場合)

Fig. 2 Theoretical mean error rate (%) v.s. sample size with fixed total sample size ( $N_1 + N_2 = 6, \sigma_1^2 = \sigma_2^2 = 1.0$ , univariate case).

4.1 1次元の場合

図2, 図3に1変数の場合の平均誤り確率の理論値と標本の大きさの関係を示す。図2は2クラスの標本分散が等しい場合 ( $\sigma_1^2 = \sigma_2^2 = 1.0$ ), 図3は2クラスの標本分散が異なる場合 ( $\sigma_1^2 = 4.0, \sigma_2^2 = 0.25$ ) の平均誤り確率を示す。2クラスの事前確率は等しく, 平均はそれぞれ  $-1.0, 1.0$  とした。  $N_1, N_2$  は各クラスの標本の大きさで, 総標本の大きさは  $6 (N_1 + N_2 = 6)$  である。平均誤り確率は  $t$  分布の積分によって求めた理論値である [付録2.]。

標本分散と標本の大きさがともに等しい場合は最適識別関数と2次識別関数は同じ結果を与えるが, それ以外の場合では, 最適識別関数の平均誤り確率は2次識別関数より常に小さい。2クラスの標本分散が等しい場合, 2次識別関数では2クラスの標本の大きさが等しい場合に最も平均誤り確率が小さく, 最適識別関数では逆に2クラスの標本の大きさが等しい場合に最も平均誤り確率が大きい。2クラスの標本分散が異なる場合には, 2次識別関数, 最適識別関数ともに標本分散の小さいクラスの標本の大きさがより多い場合に平均誤り確率が小さくなる。

4.2 多次元の場合

図4に2クラスの標本共分散行列が異なる場合の最

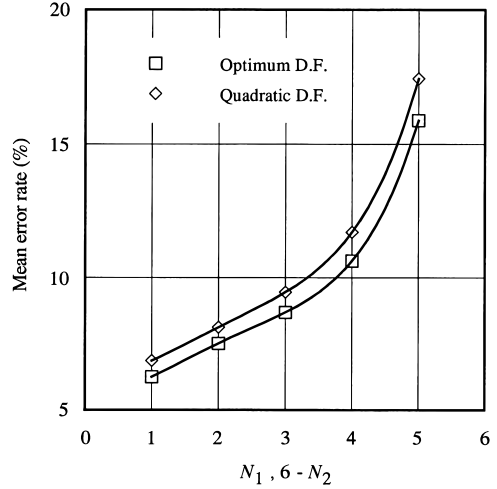


図3 総標本の大きさが一定の場合の平均誤り確率の理論値 (%) ( $N_1 + N_2 = 6, \sigma_1^2 = 4.0, \sigma_2^2 = 0.25$ , 1変数の場合)

Fig. 3 Theoretical mean error rate (%) v.s. sample size with fixed total sample size ( $N_1 + N_2 = 6, \sigma_1^2 = 4.0, \sigma_2^2 = 0.25$ , univariate case).

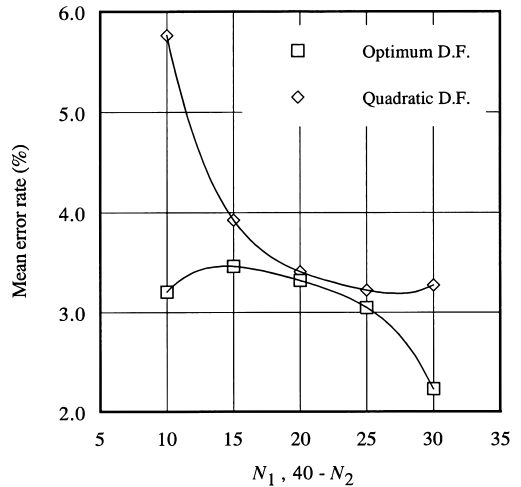


図4 総標本の大きさが一定の場合の平均誤り確率 (%) ( $N_1 + N_2 = 40$ , 8変数の場合)

Fig. 4 Mean error rate (%) v.s. sample size with fixed total sample size ( $N_1 + N_2 = 40$ , 8-variate case).

適識別関数と2次識別関数の平均誤り確率をモンテカルロ法により求めた結果を示す。テスト標本の大きさは5000, テストの反復回数は1000とし, 2クラスの事前確率は等しいものとする。クラス1の標本共分散行列は  $8 \times 8$  の単位行列, クラス2の標本共分散行列

は  $8 \times 8$  の対角行列でその対角要素は、

$$\begin{aligned} \text{diag}\Sigma_2 \\ = (8.41, 12.06, 0.12, 0.22, 1.49, 1.77, 0.35, 2.73) \end{aligned} \tag{9}$$

である。平均ベクトルは次式で与えられる 8 次元ベクトルである。

$$\begin{aligned} M_1 &= (0, 0, 0, \dots, 0), \\ M_2 &= (3.86, 3.10, 0.84, 0.84, 1.64, 1.08, 0.26, 0.01) \end{aligned} \tag{10}$$

$N_1, N_2$  は各クラスの標本の大きさで、総標本の大きさは  $40 (N_1 + N_2 = 40)$  である。1 次元の場合と同様に、最適識別関数の平均誤り確率は 2 次識別関数に比べて常に小さく、標本の大きさの差が大きいほど平均誤り確率の差も大きくなる。

### 4.3 文字認識実験による性能評価

本節では、3. で述べた各識別関数・手法により手書き数字の認識率と特徴ベクトルの次元数の関係を計算機実験で調べる。比較実験は学習標本の大きさがほぼ等しい場合と学習標本の大きさが異なる場合とに分けられる。

以下に比較実験について詳しく説明する。

#### 4.3.1 特徴ベクトル

比較実験では、実際の郵便物から収集した郵便番号の手書き数字に対して、濃度こう配から求めた 400 次元の原特徴ベクトル [11] とその特徴ベクトルを主成分分析により次元削減した 3, 6, 9, 12, 20, 32, 48, 64, 100, 144, 196, 256 次元の特徴ベクトルを用いる。

以下に、画像の周辺分布と文字の位置情報を用いる切出しを行った後の数字データに対する 400 次元の原特徴ベクトルの抽出手順を簡単に示す。

- (1) 前処理として、外接枠・重心合せによる位置・大きさの正規化を行い、 $72 \times 72$  画素の画像を作成する。
- (2) 上記の画像に  $2 \times 2$  の平均値フィルタを用いる平滑化処理を  $r (r=5)$  回反復して濃淡画像を求める。
- (3) 濃淡画像の濃度値の平均が 0、最大値が 1 となるように濃度値を正規化する。
- (4) 正規化画像に対して Roberts フィルタを適用し、濃淡画像の濃度値のこう配の向きと強さを求める。
- (5) こう配の向きを  $\pi / 16$  刻みで、32 方向に量子化する。

表 1 各クラスの標本の大きさ (学習標本の大きさがほぼ等しい場合)

Table 1 Sample size of each class (Case of nearly common learning sample size).

クラス	学習標本の大きさ	認識用標本の大きさ
0	3202	1752
1	3100	1678
2	2891	1386
3	3088	1556
4	3298	1380
5	2937	1546
6	2725	1650
7	2805	1481
8	2886	1291
9	3945	1241
Total	29877	14961

表 2 各クラスの標本の大きさ (学習標本の大きさが異なる場合)

Table 2 Sample size of each class (Case of independent learning sample size).

クラス	学習標本の大きさ	認識用標本の大きさ
0	2000	1752
1	200	1678
2	20	1386
3	2000	1556
4	200	1380
5	20	1546
6	2000	1650
7	200	1481
8	20	1291
9	2000	1241
Total	8660	14961

(6) 文字の外接枠を  $9 \times 9$  の 81 のブロックに分割し、各ブロック内において (5) で量子化した 32 の方向別にこう配の強さを加算して局所方向ヒストグラムを作成する。

(7)  $[14641]$  の加重フィルタにより、32 方向のヒストグラムを、16 方向に再量子化する。更に方向別に、 $5 \times 5$  の 2 次元ガウスフィルタを縦横 1 ブロックおきに施して再標本化し、ブロック数を  $9 \times 9$  から  $5 \times 5$  に削減して 400 次元の特徴ベクトルを求める。

(8) 変数変換 ( $y = x^u, u = 0.4$ ) により特徴ベクトルの分布形を正規分布に近づける。

次に、主成分分析法を用いて、400 次元の原特徴ベクトルを  $n'$  次元 ( $n' = 3, 6, 9, 12, 20, 32, 48, 64, 100, 144, 196, 256, 400$ ) に次元削減する。

#### 4.3.2 実験データ

実験には、第 1 回及び第 2 回郵政省郵政研究所文字

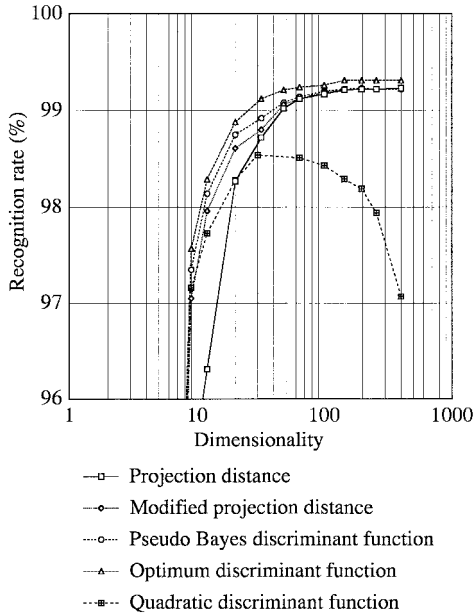


図5 手書き数字認識の認識率  
(学習標本の大きさがほぼ等しい場合)

Fig. 5 Recognition rate of handwritten numeral recognition (Case of nearly common learning sample size).

認識技術コンテストで集められた学習用・評価用の標本を用いる [11], [13], [14]. この標本は実際の郵便はがきから収集された 3 けたの郵便番号からなり、郵便番号の総数は 14,946 (44,838 文字) である. そのうち、学習用に 29,877 文字、評価用に残り 14,961 文字を用いた. 字種ごとの文字数の内訳を表 1 と表 2 に示す.

#### 4.3.3 学習標本の大きさがほぼ等しい場合

図 5 は学習標本の大きさがほぼ等しい場合の各識別関数・手法による認識率のグラフである.

信頼度定数  $N_0$  は全クラスで共通とし、次式で決定した.

$$N_0 = \frac{\alpha}{1-\alpha} \bar{N} \quad (11)$$

ここで、 $\bar{N}$  は各クラスの学習標本の大きさの平均で、 $\alpha$  は実験的に認識率が最も高くなる値を選ぶ.

図 5 では、横軸は特徴抽出後の次元数 (対数表示)、縦軸は認識率を表す. このグラフから、最適識別関数及び最適識別関数から導かれる、擬似ベイズ識別関数、改良投影距離、投影距離を用いる識別系ではピーキング現象が起きていないことがわかる. また、最適識別

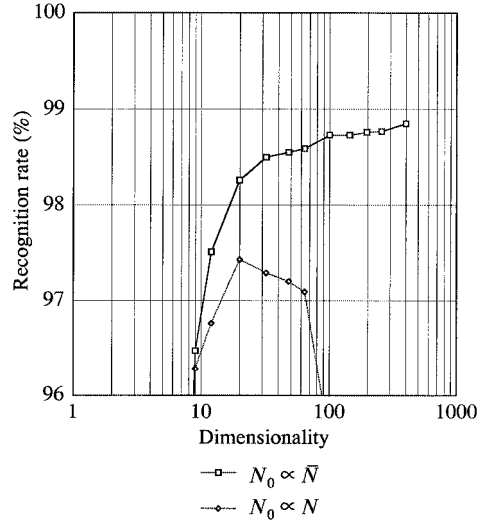


図6 最適識別関数の認識率  
(学習標本の大きさが異なる場合)

Fig. 6 Recognition rate of optimum discriminant function (Case of independent learning sample size).

関数の認識率が最も高く、32 次元の特徴ベクトルで、99.0%以上の認識率が得られ、144 次元以上の特徴ベクトルでは 99.31%の認識率が得られている. 2 次識別関数では、3 次元から 32 次元までの特徴ベクトルに対して認識率が向上しているが、48 次元以上では認識率が下がっている. また、最適次元数における 2 次識別関数の認識率は、その次元における擬似ベイズ識別関数の認識率より低く、必要な計算量、記憶容量も多い. 2 次識別関数の性能劣化の原因としては、学習文字数の不足が考えられる. 2 次識別関数の学習標本の大きさは特徴ベクトルの次元数の 2 次のオーダーで増加させる必要があることが指摘されている [15].

#### 4.3.4 学習標本の大きさが異なる場合

学習標本の大きさが異なる場合の特徴ベクトルの次元数と認識率の関係を図 6、図 7 に示す.

図 6 に、信頼度定数  $N_0$  が全クラス共通の場合と  $N_0$  が学習標本の大きさに比例する場合の最適識別関数の認識率のグラフを示す.  $N_0$  が全クラス共通の場合には、特徴ベクトルの次元数の増加に伴って認識率が高くなり、 $N_0$  が学習標本の大きさに比例する場合には、ピーキング現象が起きている. 一般に、異なる次元数の特徴ベクトルを用いる最適識別系の性能を比較する場合、「比較可能」の条件が満たされていないと見かけのピーキング現象が生じることがわかってい

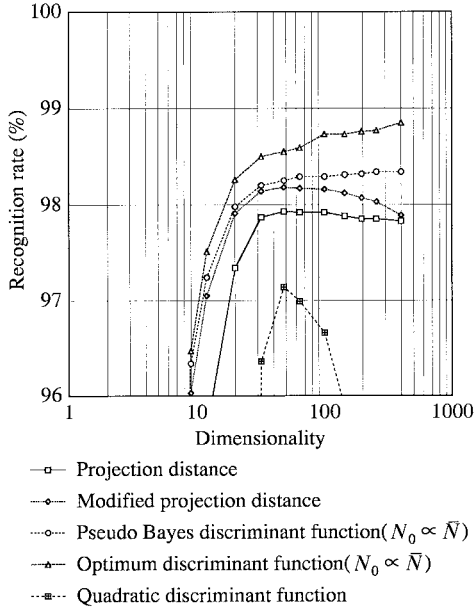


図7 手書き数字認識の認識率  
(学習標本の大きさが異なる場合)

Fig.7 Recognition rate of handwritten numeral recognition (Case of independent learning sample size).

る [2], [16]. 以上の結果から, 信頼度定数を学習標本の大きさとは無関係に, クラスで共通の定数にした場合には「比較可能」の条件が満たされ, ピーキング現象が生じなかったものと考えられる.

図7は, 学習標本の大きさが異なる場合の各識別関数・手法による認識率のグラフである. 最適識別関数の認識率が最も高く, 改良投影距離, 投影距離, 2次識別関数ではピーキング現象が起きている.

投影距離のピーキング現象が顕著でないのは, 投影距離においては, 用いる特徴ベクトルの次元数が低いほど, 3.3で述べた部分空間の共有による認識率の低下が大きくなって, 改良投影距離に比べて顕著なピークが現れにくいからである.

#### 4.3.5 計算量の比

表3に, 擬似バイズ識別関数の計算量を1とした場合の各手法のおよその計算量を示す. 最適識別関数の計算量は10, 投影距離法の計算量は0.625, 改良投影距離法の計算量は1, 2次識別関数の計算量は10になる.

表3 計算量の比  
Table 3 Ratio of computation cost.

最適識別関数・手法	計算量
最適識別関数	10
擬似バイズ識別関数	1
投影距離	0.625
改良投影距離	1
2次識別関数	10

## 5. まとめ

平均誤り確率の理論及びモンテカルロ法による評価の結果を以下に要約する.

- (1) 最適識別関数の平均誤り確率が2次識別関数の平均誤り確率を超えることはなかった.
- (2) 標本の大きさの差が大きいほど最適識別関数と2次識別関数の性能の差が顕著になる傾向が認められた.
- (3) 2クラスの分散に差がある場合, 分散の小さなクラスの標本が大きいほど平均誤り確率が減少する傾向が認められた.

また, 手書き数字認識実験の結果を要約すると以下のようなになる.

- (1) 最適識別関数の認識率が最も高かった.
- (2) 最適識別関数と擬似バイズ識別関数は標本の大きさが等しい場合も異なる場合もピーキング現象が起きなかった.
- (3) 改良投影距離と投影距離は学習標本の大きさが等しい場合にはピーキング現象が起きなかったが, 学習標本の大きさが異なる場合にはピーキング現象が起きた.
- (4) 2次識別関数はどちらの場合でもピーキング現象が起きた.
- (5) 信頼度定数  $N_0$  は学習標本の大きさとは無関係に, クラスで共通の定数にした方がよい.

学習標本の大きさがほぼ等しい場合にバイズアプローチから導かれる識別関数・手法でピーキング現象が起きなかったのは, 事前分布の共分散行列  $\Sigma_0 (= \sigma^2 I)$  を用いているためと考えられる. また, 学習標本の大きさが異なる場合に最適識別関数と擬似バイズ識別関数でのみピーキング現象が起きなかったのは, これらの識別関数には, 学習標本の大きさ  $N$  がパラメータとして含まれているからである. 特に最適識別関数では学習標本の大きさの違いが多変量  $t$  分布の自由度の違いとして密度関数の形に反映されて決定境界の位置

が最適化されるためにピーキング現象が生じないものと考えられる。

手書き数字認識実験で用いた特徴ベクトルの分布は、変数変換によって正規性が改善されているが完全な正規分布ではない[11]。このことは、最適識別関数が正規分布からのずれに対してある程度の頑健性を備えていることを示しているが、この問題については今後更に検討する予定である。

## 6. む す び

今後の課題として、

- (1) 信頼度定数  $N_0$  が 0 でない場合の平均誤り確率の、理論及びモンテカルロ法による評価、
  - (2) 4.3 で行った手書き数字認識実験のモンテカルロ法によるモデル化と評価、
  - (3) 信頼度定数の決め方とその根拠に関する考察、
  - (4) 事前分布の分散の決め方とその根拠に関する考察、
  - (5) 正規分布からのずれによって最適識別関数が受ける影響に関する考察、
- が考えられる。

謝辞 本研究に関し、手書き郵便番号データを作成・配布された郵政省郵政研究所の皆様、原稿の清書に協力をしていただいた研究室の中嶋孝氏に感謝致します。

## 文 献

- [1] 蓑谷千鳳彦, “推定と検定のはなし,” 東京図書館, 1988.
- [2] J.M. Van Campenhout, “On the peaking of Hughes mean recognition accuracy: The resolution of an apparant paradox,” IEEE Trans. Syst., Man & Cybern., vol.SMC-8, no.5, pp.390-395, May 1978.
- [3] W.G. Waller and A.K. Jain, “On the monotonicity of the performance of Bayesian classifiers,” IEEE Trans. Info. Theory, vol.IT-24, pp.392-394, 1978.
- [4] J.M. Van Campenhout, “Topics in Measurement Selection,” Handbook of Statistics, vol.2, North-Holland Publishing Company, pp.793-803, 1982.
- [5] D. Lindley, “The Bayesian approach,” Scand. J. Statist., vol.5, pp.1-26, 1978.
- [6] D.G. Keehn, “A note on learning for Gaussian properties,” IEEE Trans. Inform. Theory, vol.IT-11, no.1, pp.126-132, Jan. 1965.
- [7] B.D. Ripley, “Pattern Recognition and Neural Networks,” p.52, Cambridge University Press, 1996.
- [8] S.J. Raudys and A.K. Jain, “Small sample size effects in statistical pattern recognition: Recommendations for practitioners,” IEEE Trans. Pattern Analysis & Machine Intelligence, vol.13, no.3, pp.252-264, March 1991.
- [9] R.O. Duda and P.E. Hart, “Pattern Classification

and Scene Analysis,” p.52, John Wiley & Sons, Inc., New York, 1973.

- [10] 浜本義彦, “パターン認識理論の最近の動向,” 信学誌, vol.77, no.8, pp.853-864, Aug. 1994.
- [11] 若林哲史, 鶴岡信治, 木村文隆, 三宅康二, “特徴量の次元数増加による手書き数字認識の高精度化,” 信学論(D-II), vol.J77-D-II, no.10, pp.2046-2053, Oct. 1994.
- [12] 池田正幸, 田中英彦, 元岡 達, “手書き文字認識における投影距離法,” 情処学論, vol.24, no.1, pp.106-112, Jan. 1983.
- [13] 松井俊弘, 山下郁生, 若厚 徹, 吉室 誠, “文字認識アルゴリズム複合化の検討, 第一回文字認識技術コンテストの結果より,” 信学技報, PRU92-33, Sept. 1992.
- [14] 能見 正, 松井俊弘, 山下郁生, 若厚 徹, 吉室 誠, “手書き数字認識における誤読・リジェクトパターンの分析,” 信学技報, PRU-93-46, Sept. 1993.
- [15] K. Fukunaga and R.R. Hayes, “Effects of sample size in classifier design,” IEEE Trans. Pattern Analysis & Machine Intelligence, vol.PAMI-11, no.8, pp.873-885, Aug. 1989.
- [16] G.F. Huges, “On the mean accuracy of statistical pattern recognizers,” IEEE Trans. Info. Theory, vol.IT-14, no.1, pp.55-63, Jan. 1968.

## 付 録

### 1. 擬次ベイズ識別関数の導出

事前分布として球状分布を仮定して  $\Sigma_0 = \sigma^2 I$  とすると、

$$\Sigma_N = (1 - \alpha)\Sigma + \alpha\sigma^2 I \quad (\text{A-1})$$

が得られる。

ここで、

$$\begin{aligned} \{(1 - \alpha)\Sigma + \alpha\sigma^2 I\}\Phi_i &= (1 - \alpha)\Sigma\Phi_i + \alpha\sigma^2\Phi_i \\ &= \{(1 - \alpha)\lambda_i + \alpha\sigma^2\}\Phi_i \\ &\quad (i = 1, 2, \dots, n) \end{aligned} \quad (\text{A-2})$$

となるため、 $\Sigma_N$  の固有値、固有ベクトルはそれぞれ  $(1 - \alpha)\lambda_i + \alpha\sigma^2$ 、 $\Phi_i (i = 1, 2, \dots, n)$  となることから、その結果、

$$\begin{aligned} Y &= (X - M)^T \Sigma_N^{-1} (X - M) \\ &= \sum_{i=1}^n \frac{1}{(1 - \alpha)\lambda_i + \alpha\sigma^2} \{\Phi_i^T (X - M)\}^2 \end{aligned} \quad (\text{A-3})$$

となる。

ここで、ある定数  $k$  に対して、 $i > k$  の場合に  $(1 - \alpha)\lambda_i \ll \alpha\sigma^2$  を仮定すると式(A-3)は次式のよう



に近似される .

$$Y \approx \sum_{i=1}^k \frac{1}{(1-\alpha)\lambda_i + \alpha\sigma^2} \{\Phi_i^T(X-M)\}^2 + \sum_{i=k+1}^n \frac{1}{\alpha\sigma^2} \{\Phi_i^T(X-M)\}^2 \quad (\text{A}\cdot 4)$$

以下の関係 ,

$$\sum_{i=k+1}^n \{\Phi_i^T(X-M)\}^2 = \|X-M\|^2 - \sum_{i=1}^k \{\Phi_i^T(X-M)\}^2 \quad (\text{A}\cdot 5)$$

を用いて式 (A.4) を変形すれば ,

$$Y \approx \frac{1}{\alpha\sigma^2} \left[ \|X-M\|^2 - \sum_{i=1}^k \frac{(1-\alpha)\lambda_i}{(1-\alpha)\lambda_i + \alpha\sigma^2} \{\Phi_i^T(X-M)\}^2 \right] \quad (\text{A}\cdot 6)$$

が得られる . 一方 ,

$$\begin{aligned} \ln|\Sigma_N| &= \sum_{i=1}^n \ln\{(1-\alpha)\lambda_i + \alpha\sigma^2\} \\ &\approx \sum_{i=1}^k \ln\{(1-\alpha)\lambda_i + \alpha\sigma^2\} + \sum_{i=k+1}^n \ln(\alpha\sigma^2) \end{aligned} \quad (\text{A}\cdot 7)$$

となり , 第 2 項はクラスに共通の定数項となる .

式 (4) に式 (A.6) , (A.7) と  $\alpha = N_0/(N + N_0)$  を代入し , 定数項を省略すると式 (5) の擬似バイズ識別関数を得る .

## 2. 平均誤り確率の計算

### 2.1 標本の大きさが等しい場合

標本の大きさが等しい場合識別関数は 2 次関数となり , 1 次元の場合は容易に決定境界を求め平均誤り確率を計算することができる . ここでは簡単のため事前確率が等しい場合を考えるが , 事前確率が異なる場合も同様に計算できる .

識別関数は式 (4) より次式となる .

$$g_i(x) = \sigma_{N_i}^{\frac{2}{N_i+1}} \left\{ 1 + \frac{1}{N_N} \left( \frac{x-m_i}{\sigma_{N_i}} \right)^2 \right\} \quad (i=1,2) \quad (\text{A}\cdot 8)$$

決定境界は次式の  $h(x)$  を 0 とおいて求められる .

$$\begin{aligned} h(x) &= g_1(x) - g_2(x) \\ &= \sigma_{N_1}^{\frac{2}{N_1+1}} \left\{ 1 + \frac{1}{N_N} \left( \frac{x-m_1}{\sigma_{N_1}} \right)^2 \right\} \\ &\quad - \sigma_{N_2}^{\frac{2}{N_2+1}} \left\{ 1 + \frac{1}{N_N} \left( \frac{x-m_2}{\sigma_{N_2}} \right)^2 \right\} \\ &= (a-b)x^2 - 2(am_1 - bm_2)x + am_1^2 - bm_2^2 + c \\ a &= \frac{1}{N_N} \sigma_{N_1}^{\frac{2}{N_1+1}-2}, \\ b &= \frac{1}{N_N} \sigma_{N_2}^{\frac{2}{N_2+1}-2}, \\ c &= \sigma_{N_1}^{\frac{2}{N_1+1}} - \sigma_{N_2}^{\frac{2}{N_2+1}} \end{aligned} \quad (\text{A}\cdot 9)$$

$h(x) = 0$  より

$$\begin{aligned} \alpha &= \beta = \frac{m_1 + m_2}{2} \quad (\sigma_{N_1} = \sigma_{N_2}) \\ \alpha, \beta &= \frac{am_1 - bm_2 \mp \sqrt{(m_1 - m_2)^2 ab - (a-b)c}}{a-b} \quad (\sigma_{N_1} \neq \sigma_{N_2}) \end{aligned} \quad (\text{A}\cdot 10)$$

となる .  $\sigma_{N_1} \geq \sigma_{N_2}$  の場合 , 平均誤り確率は次式で与えられる .

$$\begin{aligned} \varepsilon &= P(\omega_1)\varepsilon_1 + P(\omega_2)\varepsilon_2 \\ &= P(\omega_1)P(\text{error}|\chi, \omega_1) + P(\omega_2)P(\text{error}|\chi, \omega_2) \\ &= B + A + C \\ A &= \int_{-\infty}^{\alpha} p(x|\chi, \omega_2)P(\omega_2)dx \\ &= \frac{1}{2}\Phi_{N_N} \left( \frac{\alpha - m_2}{\sigma_{N_2}} \right) \\ B &= \int_{\alpha}^{\beta} p(x|\chi, \omega_1)P(\omega_1)dx \\ &= \frac{1}{2}\Phi_{N_N} \left( \frac{\beta - m_1}{\sigma_{N_1}} \right) - \frac{1}{2}\Phi_{N_N} \left( \frac{\alpha - m_1}{\sigma_{N_1}} \right) \\ C &= \int_{\beta}^{\infty} p(x|\chi, \omega_2)P(\omega_2)dx \\ &= \frac{1}{2} \left\{ 1 - \Phi_{N_N} \left( \frac{\beta - m_2}{\sigma_{N_2}} \right) \right\} \end{aligned} \quad (\text{A}\cdot 11)$$

$\Phi_N(x_0)$  は , 自由度  $N$  の  $t$  分布  $t_N(x)$  を用いると ,

$$\Phi_N(x_0) = \int_{-\infty}^{x_0} t_N(x)dx \quad (\text{A}\cdot 12)$$

と表すことができる .

## 2.2 標本の大きさが異なる場合

標本の大きさが異なる場合の識別関数は次式で与えられる。ここでは事前確率が等しい場合を考えるが、事前確率が異なる場合も同様に計算できる。識別関数は、

$$g_i(x) = \left(\frac{\sigma_{N_i}}{D_i}\right)^2 \left\{ 1 + \frac{1}{N_{N_i}} \left(\frac{x - m_i}{\sigma_{N_i}}\right)^2 \right\}^{N_{N_i}+1}$$

$$D_i = (N_{N_i} \pi)^{-\frac{1}{2}} \frac{\Gamma\left(\frac{N_{N_i}+1}{2}\right)}{\Gamma\left(\frac{N_{N_i}}{2}\right)} \quad (i = 1, 2) \quad (\text{A.13})$$

となる。この場合は、 $h(x) = 0$  を次式の漸化式を用いて Newton 法で解き決定境界を求める。

$$x_{k+1} = x_k - \frac{h(x_k)}{h'(x_k)}$$

$$h(x) = \left(\frac{\sigma_{N_1}}{D_1}\right)^2 \left\{ 1 + \frac{1}{N_{N_1}} \left(\frac{x - m_1}{\sigma_{N_1}}\right)^2 \right\}^{N_{N_1}+1}$$

$$- \left(\frac{\sigma_{N_2}}{D_2}\right)^2 \left\{ 1 + \frac{1}{N_{N_2}} \left(\frac{x - m_2}{\sigma_{N_2}}\right)^2 \right\}^{N_{N_2}+1}$$

$$h'(x) = \frac{2(N_{N_1} + 1)}{N_{N_1} \sigma_{N_1}} \left(\frac{x - m_1}{\sigma_{N_1}}\right) \left(\frac{\sigma_{N_1}}{D_1}\right)^2$$

$$\cdot \left\{ 1 + \frac{1}{N_{N_1}} \left(\frac{x - m_1}{\sigma_{N_1}}\right)^2 \right\}^{N_{N_1}}$$

$$- \frac{2(N_{N_2} + 1)}{N_{N_2} \sigma_{N_2}} \left(\frac{x - m_2}{\sigma_{N_2}}\right) \left(\frac{\sigma_{N_2}}{D_2}\right)^2$$

$$\cdot \left\{ 1 + \frac{1}{N_{N_2}} \left(\frac{x - m_2}{\sigma_{N_2}}\right)^2 \right\}^{N_{N_2}} \quad (\text{A.14})$$

平均誤り確率は式 (A.11) と同様に計算する。

(平成 10 年 7 月 16 日受付, 11 月 16 日再受付)



韓 雪仙

1987 中国・西安電子科技大・工・電子卒。  
1994 中国・江蘇理工大・大学院修士課程了。  
平 9 三重大・工・大学院博士前期課程了。  
平 10 同大学院博士後期課程入学、現在に至る。  
文字・パターン認識、コンピュータグラフィックスの研究に従事。



若林 哲史 (正員)

昭 60 三重大・工・電子卒。昭 62 同大学院修士課程了。平 3 三重大・工・助手。平 10 同助教授、現在に至る。手書き文字認識、文書理解、画像処理、コンピュータグラフィックスの研究に従事。工博。



木村 文隆 (正員)

昭 48 名大・工・電気卒。昭 53 同大学院博士課程了。同年同大・工・助手。昭 58 三重大・工・助教授。平 10 同教授、現在に至る。その間平 1~3 ミシガン大客員助教授を務める。文字・パターン認識、画像処理、コンピュータグラフィックスの研究に従事。情報処理学会、日本 ME 学会、人工知能学会各会員。工博。



三宅 康二 (正員)

昭 35 名大・工・電気卒。昭 40 同大学院博士課程了。同年同大・工・助手、昭 43 同講師。昭 43 同助教授。昭 53 三重大・工・教授、現在に至る。コンピュータグラフィックス及び医用画像処理の研究に従事。情報処理学会、日本 ME 学会、日本人工知能学会、日本ロボット学会、電気学会、日本シミュレーション学会各会員。工博。