

## 付加情報割当ての一般理論と誤認識率の解析

岩村 雅一<sup>†</sup> 古谷 嘉男<sup>†</sup> 黄瀬 浩一<sup>†</sup> 大町真一郎<sup>††</sup> 内田 誠一<sup>†††</sup><sup>†</sup> 大阪府立大学大学院工学研究科 〒 599-8531 堺市中区学園町 1-1<sup>††</sup> 東北大学大学院工学研究科 〒 980-8579 仙台市青葉区荒巻字青葉 6-6-05<sup>†††</sup> 九州大学大学院システム情報科学研究所 〒 819-0395 福岡市西区元岡 744E-mail: <sup>†</sup>{masa,kise}@cs.osakafu-u.ac.jp, <sup>††</sup>machi@ecei.tohoku.ac.jp, <sup>†††</sup>uchida@is.kyushu-u.ac.jp

あらまし 通常のパターン認識とは異なる, 付加情報を用いるパターン認識という枠組みが提案されている. これは通常のパターンと同時に, それを補助する情報(付加情報)をパターンとは別に用意し, それらを同時に用いることにより誤認識率を減少させる方式である. 付加情報が誤らないという条件の下で誤認識率を最小にする条件がこれまでに導かれているが, 実際には観測誤差が生じるために付加情報が誤らないという条件はいささか強過ぎる. そこで本研究では, 付加情報に誤りがあるという条件において, どのような付加情報が誤認識率の減少に寄与するかについて考察する. さらにサンプルが正規分布に従い, 各クラスの共分散行列が等しい場合について近似最適解を求め, 誤認識率が減少することを検証する.

キーワード 付加情報, 観測誤差, 誤認識, 正規分布, マハラノビス距離

## A General Theory of Supplementary Information Assignment and Analysis of Error Rates

Masakazu IWAMURA<sup>†</sup>, Yoshio FURUYA<sup>†</sup>, Koichi KISE<sup>†</sup>,Shinichiro OMACHI<sup>††</sup>, and Seiichi UCHIDA<sup>†††</sup><sup>†</sup> Graduate School of Engineering, Osaka Prefecture University, 1-1 Gakuencho, Naka, Sakai, 599-8531 Japan<sup>††</sup> Graduate School of Engineering, Tohoku University, 6-6-05 Aoba, Aramaki, Aoba, Sendai, 980-8579 Japan<sup>†††</sup> Faculty of Information Science and Electrical Engineering, Kyushu University

744 Motoooka, Nishi, Fukuoka, 819-0395 Japan

E-mail: <sup>†</sup>{masa,kise}@cs.osakafu-u.ac.jp, <sup>††</sup>machi@ecei.tohoku.ac.jp, <sup>†††</sup>uchida@is.kyushu-u.ac.jp

**Abstract** *Pattern recognition with supplementary information* is a new pattern recognition framework that determines an output class by combining a feature vector extracted from the pattern and *supplementary information* suggesting the true class. Under the condition that supplementary information does not contain error, a theory that reduces error rates have been proposed. However, in the real world, we cannot observe any measure without error. Thus, in this paper, we discuss how to reduce error rates using the erroneous supplementary information, and confirm the effect experimentally using artificial samples following the normal distribution with a common covariance matrix.

**Key words** supplementary information, observation error, misrecognition, normal distribution, Mahalanobis distance

## 1. はじめに

パターン認識には本質的に解決不可能な誤認識が存在する. これは異なるクラスに属するサンプルから類似の特徴量が抽出されることによって生じ, バイズエラーと呼ばれる. バイズエラーを減少させるためには特徴量を工夫することが王道であ

るが, これにも限界がある. 例えば, アルファベット大文字の“P”(アイ)と小文字の“p”(エル)は, フォントによっては完全に同一の外見を持ち, 見た目では判断することは不可能になる.

このような限界に挑むため, 付加情報を用いるパターン認識という方式が提案されている [1]. この方式では通常のパターン認識を行うと同時に, パターンとは別に用意した正解クラスに

関する何らかの情報を観測し、両者を組み合わせることで認識を行う。これにより、特徴量によってクラスが本質的に分離不可能な場合であっても、付加情報を上手く割り当てることで誤認識率を減少させることができる。文献[1]では付加情報の種類と誤認識率の関係について論じている。

ところが、付加情報を実際に用いようとするとき、観測誤差が問題となる。文献[2]～[4]では幾何学的不変量を用いて文字に付加情報を埋め込む方法が検討されているが、いずれの方法も観測誤差の影響を回避できない。そのため、付加情報に誤差がないとする文献[1]の理論には限界があると考えられる。そこで本研究では付加情報が誤ることを前提に、誤認識率を最小にする付加情報の割り当て方法と達成可能な認識率について考察する。また、提案する割り当て方法を用いることで、文献[1]の方法に比べて誤認識率が減少することを実験により確認する。

## 2. 付加情報を用いる認識の理論

本題に入る前に、付加情報の理論体系における本研究の位置付けを確認しておく。

図1は3クラス問題を模式的に表している。図1(a)は通常のパターン認識である。通常のパターン認識ではサンプルがどのクラスに属しているかを認識する際に、認識対象から抽出した特徴量のみを用いる。したがって、図のように特徴量の分布に重なりがあると誤認識が起これ、これを避けることはできない。

これに対して、図1(b)は文献[1]の付加情報を用いるパターン認識を表している。文献[1]では特徴量のみでは避けることができない誤認識を回避するために、誤りのない付加情報を用いる。図1(b)の例ではクラスAとCのサンプルに0の付加情報を、クラスBのサンプルに1の付加情報を与えている。付加情報が誤らないとしているため、認識の際に付加情報を調べることは図のように空間を分割し、候補クラスを限定することに相当する。したがって、誤認識が起これり易い2クラスに別々の付加情報を与えておくことで、特徴量のみでは不可避な誤りを回避することができる。空間分割後は候補クラスを絞った上で通常のパターン認識を行う。

図1(c)は本研究で検討する付加情報の誤りを考慮する場合である。この場合は付加情報を図1(b)と同様に与えても、付加情報の誤りのために特徴空間を完全に分割することができず、図1(b)では起これらない誤認識が起これる。ここで生じる疑問は、付加情報が誤らないことを前提とした図1(b)の割り当て方法が付加情報が誤る図1(c)においても良い割り当て方法であるのか、ということである。本研究ではこの点について議論し、付加情報が誤る場合における「付加情報の良い割り当て」がどのようなものかを明らかにする。

## 3. 誤らない付加情報の割り当て方法 [1]

付加情報が誤る場合を考える前に、その準備として付加情報が誤らない場合の割り当て方法の概略を述べる。

まず、ある3クラス問題の特徴量の分布が図2(a)であるとする。付加情報を用いない通常のパターン認識を行う場合、これは図1(a)のモデルに相当する。図中の灰色の領域は不可避

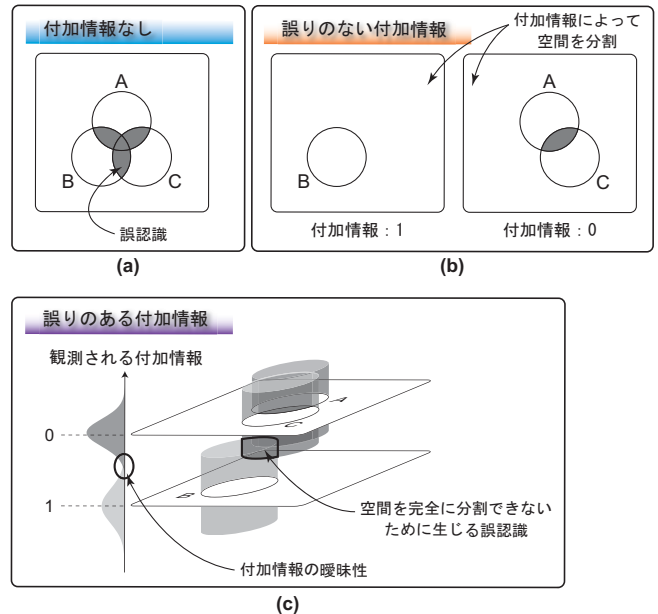
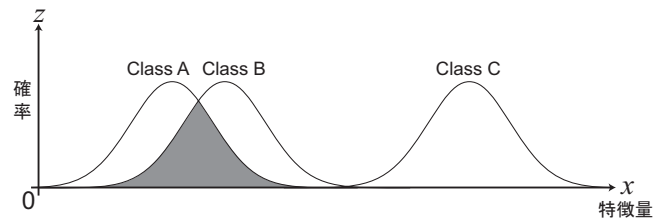
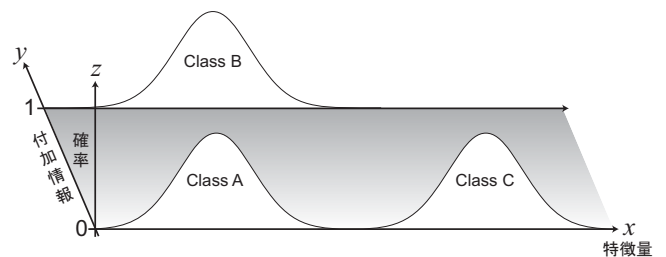


図1 付加情報を用いるパターン認識の理論体系



(a) 通常のパターン認識。



(b) 文献[1]の付加情報の割り当て。

図2 各クラスの特徴量の分布と付加情報。

な誤認識を表し、その面積が誤認識率を表す。

次に、この問題に付加情報を導入する。図1(b)で述べたように付加情報が誤らないとすれば、このモデルでは付加情報の種類が認識性能を左右する。付加情報が1種類の場合は通常のパターン認識に相当し、付加情報の種類が多くなるほど誤認識率が減少する。この例の場合は3種類で誤認識率が起これなくなり、そのときの割り当ては一意に定まる。一方、2種類の場合は3通りの割り当て方法が考えられる。文献[1]では付加情報を用いたときの誤認識モデルから誤認識率を最小にする割り

当て方法を与えている<sup>(注1)</sup>。

図 2(a) に 2 種類の付加情報を与える場合、誤認識率が最小になるのはクラス間の分布の重なりが最小になる割り当て、すなわち図 2(b) の割り当てである。図 2(b) には図 2(a) に付加情報を表す  $y$  軸が加えてある。付加情報が誤らないとしているため、 $y$  軸方向の分散が 0 であり、図 1(b) のように空間が完全に分割されている。

本稿ではここで述べた文献 [1] の手法を従来手法と呼び、付加情報が誤る場合の考察の出発点とする。

#### 4. 誤差を含む付加情報の割り当て方法

前節では付加情報が誤らない場合を図 2(b) のような  $y$  軸方向には分散を持たない分布で表した。一方、付加情報が誤差を含む場合は  $y$  軸方向には分散を持つ分布になる。この場合、異なる付加情報を割り当てた分布でも重なることがあるため、誤認識が生じる。本節では、この状況において、どのように付加情報を割り当てれば誤認識率が最小の「良い割り当て」になるのかについて検討する。

##### 4.1 問題設定

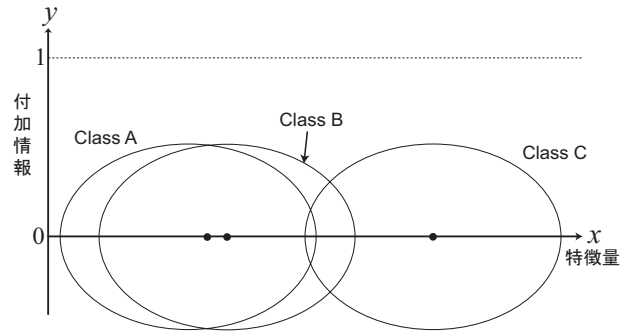
提案手法でどのような問題を考えるのかを整理するために、設計者ができることとできないことを述べる。

- 割り当てる付加情報 [変更可能]：文献 [1] の場合と同様に、設計者は付加情報を自由に割り当てることができる。
- 付加情報の分布 [変更不可能]：付加情報の分布とは付加情報の誤りを表す指標である。前項で述べたように、設計者は割り当てる付加情報を自由に変更できる。しかし、誤りがあるために割り当てた値を観測できる訳ではない。割り当てた値に対して観測される値の分布を付加情報の分布と呼ぶことにする。誤りの傾向は環境の影響を受けるため、観測することはできても、設計者が変更することができない<sup>(注2)</sup>。
- 特徴量の分布 [変更不可能]：特徴量は、サンプルを観測することによって得られる。設計者はどの特徴を使うかの選択はできるが、観測結果を左右することはできない。したがって、特徴量の分布を設計者は変更することができない。

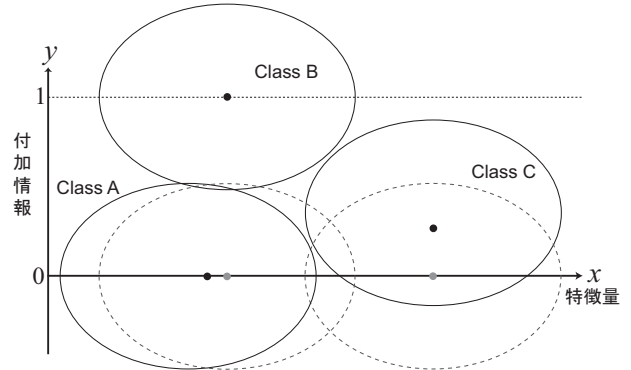
以上をまとめると、設計者は割り当てる付加情報の値のみを任意に設定できる。唯一任意に設定できる付加情報も、設定できる範囲は有限である。これは物理的な制約に起因する。文献 [2]～[4] では幾何学的不変量である複比や面積比を用いているが、通常これらの値は無限大にできない。一般に無限大の付加情報を使用することはできないと考えられるため、有限と考えるのが道理である。本研究では簡単のために付加情報を割り当てる範囲を 0～1 の間とする。もし実際に割り当てることができる範囲が 0～2 であれば 2 倍すればいいだけなので、このことによって一般性は失われない。

##### 4.2 付加情報の与え方

本論文の主題である、どのような付加情報が誤認識率を最小



(a) 全クラスに同じ付加情報を割り当てた場合。



(b) 最適な付加情報を割り当てた場合。

図 3 各クラスの特徴量と誤りがある付加情報の分布。楕円は確率分布の等高線。

にするのかについて検討する。図 3(a) は図 2(a) の 3 クラス問題を  $z$  軸の正の方向から見た図である。図中の楕円はサンプルの発生確率の等高線を表している。ただし、図 2(a) と違い、付加情報が誤るため、付加情報の分布は  $z$  軸方向の広がりを持っている。図 3(a) では付加情報がすべて同じであるため、誤認識率は図 2(a) と同じである。なお、付加情報は任意に与えるため、サンプルの特徴量  $x$  と付加情報  $y$  は独立である。したがって、図 3 の楕円は  $x$ - $y$  平面で傾かない。

図 3(a) の問題の誤認識率を減少させるため、割り当てる付加情報の値を変更することを考える。そこで図 3(a) を眺めてみると、付加情報を表す  $y$  軸と特徴量を表す  $x$  軸に大きな違いがないことに気付く。すなわち、誤りのある付加情報は、設計の余地が残された特徴量と解釈できる。したがって、誤りのある付加情報を用いた場合の誤認識率は、通常のパターン認識のそれと類似の方法で計算できると考えられる。以上をまとめると、誤りのある付加情報を用いたときの誤認識率を最小にするには、「付加情報も特徴量とみなして計算した誤認識率が最小になるような付加情報を割り当てれば良い」ということになる。ここで誤りのある付加情報を特徴量と解釈できるのは誤認識率の計算においてであり、任意の値を設定できるという意味では通常の特徴量と異なる点に注意していただきたい。

##### 4.3 問題の定式化

$N$  クラス  $\omega_1, \dots, \omega_N$  を分類する問題を考え、前節で得られ

(注1)：文献 [1] では誤認識率が最小の割り当てを求める方法を与えているが、この問題は NP 困難であるため、実験では欲張り法による近似解を求めている。

(注2)：付加情報としてどの値を割り当てるかは設計者が決定できるため、厳密な意味では、設計者は付加情報の分布を平行移動することのみ可能である。

た知見を定式化する。

クラス  $\omega_i$  のサンプルが発生する確率を  $P(\omega_i)$  とおく。サンプルの特徴量を  $\mathbf{x}$ , 付加情報を  $\mathbf{y}$  とおき,  $\mathbf{x}$  と  $\mathbf{y}$  はそれぞれ  $p$  次元と  $q$  次元のベクトルとする。そして,  $\mathbf{x}$  がクラス  $\omega_i$  から発生する確率を  $P_i^x \equiv P(\mathbf{x}|\omega_i)$  とおき, 同様にクラス  $\omega_i$  の付加情報が発生する確率を  $P_i^{y|\nu_i} \equiv P(\mathbf{y}|\nu_i; \omega_i)$  とおく。ここで  $\nu_i$  は割り当てる付加情報である。  $\mathbf{x}$  と  $\mathbf{y}$  を合成したベクトル  $\mathbf{z} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}$  を認識して得られる誤認識率は,  $\mathbf{x}$  と  $\mathbf{y}$  の独立性より  $\mathbf{z}$  の確率密度関数を  $P_i^z = P_i^x P_i^{y|\nu_i}$  とおくととき,

$$\begin{aligned} 1 - \int_{\mathbf{z}} \max_i \{P(\omega_i) P_i^z\} d\mathbf{z} \\ = 1 - \int_{\mathbf{x}} \int_{\mathbf{y}} \max_i \{P(\omega_i) P_i^x P_i^{y|\nu_i}\} d\mathbf{y} d\mathbf{x} \quad (1) \\ \equiv R_{\text{error}}(P_i^x, P_i^{y|\nu_i}) \end{aligned}$$

とおける。この誤認識率を最小にするために解くべき問題は

$$(\nu_1, \dots, \nu_N) = \underset{\nu_1, \dots, \nu_N}{\operatorname{argmin}} R_{\text{error}}(P_i^x, P_i^{y|\nu_i}) \quad (2)$$

を満たす  $\nu_1, \dots, \nu_N$  の組を求めることに等しい。

式 (2) で定式化した問題を解くには,  $\nu_1, \dots, \nu_N$  の空間を探索する必要がある。効率の良い探索法についての考察は本稿では扱わない。その代わりに, 性質の良い単純な問題を考えて, 付加情報の誤りを考慮することで誤認識率が減少することを次節以降で確認する。

## 5. サンプルの分布が正規分布で共分散行列が等しい場合の最急降下法

本節では誤りのある付加情報を最急降下法を用いて割り当てる方法について述べ, 次節の実験ではこれを提案手法として用いる。最急降下法を用いるためには, 誤認識率を解析的に求めて微分する必要がある。そこで, 特徴量と付加情報の誤りの分布 (前節の  $\mathbf{z}$  に相当) が正規分布に従い, 全クラスの共分散行列が等しく, 認識にマハラノビス距離を用いるという理想的な場合を考える。誤認識率を求める際の方針は以下の通りである。(1) 全体の誤認識率を各 2 クラスの誤認識率を組み合わせることにより求める。(2) 2 クラスの誤認識率を誤差関数で表し, それを logistic 関数で近似する。

以下では, 4.3 節で述べた  $\mathbf{z}$  の分布を多次元正規分布として扱い, 付加情報  $\mathbf{y}$  の次元数を  $q = 1$  とする。

### 5.1 2 クラスの誤認識率

最初にクラス  $\omega_i$  と  $\omega_j$  の 2 クラスの誤認識率  $J_{ij}$  を求める。多次元空間で誤認識率を直接計算することは容易でないため, 1 次元の軸に射影して計算する。図 4 は 2 クラスの識別境界と直交する 1 次元の軸  $w_{ij}$  に分布を射影して, 1 次元正規分布を得る様子を表した模式図である。本節で考える問題は識別境界が超平面になるため, 図のように射影すれば識別境界が軸上の 1 点に射影される。このとき, 1 次元正規分布における誤認識率 (図中の灰色の領域の面積) が求める誤認識率  $J_{ij}$  になる。

次に, この 1 次元正規分布を求める。クラス  $\omega_i$  の多次元正

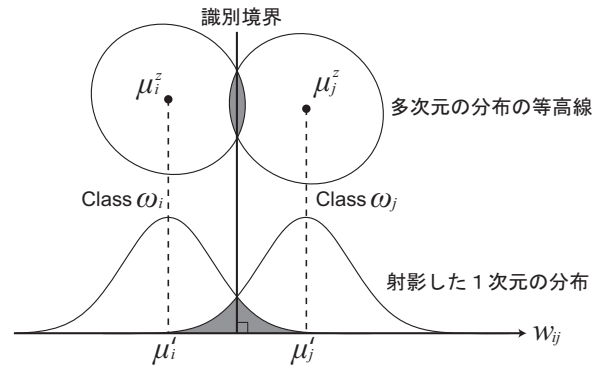


図 4 多次元正規分布を 1 次元に射影する様子。

規分布を平均  $\mu_i^z$ , 全クラス共通の共分散行列を  $\Sigma$  とおく。クラス  $\omega_i$  と  $\omega_j$  の識別境界に垂直な軸は  $w_{ij} = \Sigma^{-1}(\mu_j^z - \mu_i^z)$  で求めることができる [5]。  $w_{ij}$  のノルムを 1 に正規化したベクトルを  $w'_{ij}$  とおくと,  $w'_{ij} = \frac{w_{ij}}{|w_{ij}|} = \frac{\Sigma^{-1}(\mu_j^z - \mu_i^z)}{|\Sigma^{-1}(\mu_j^z - \mu_i^z)|}$  と表すことができる。この軸に分布を射影すれば, 軸上の 1 次元正規分布の平均と分散はそれぞれ  $\mu_i^w = w_{ij}^T \mu_i^z$ ,  $\sigma^w = w_{ij}^T \Sigma w_{ij}$  となる。

2 クラスの誤認識率は誤差関数で表せる。ここで正規分布の累積密度関数が logistic 関数  $l(x) = \frac{1}{1 + \exp(-\alpha x)}$  で近似できること [6] を利用し, 誤差関数を logistic 関数で近似すれば,

$$J_{ij} = 2 \left[ 1 - \frac{1}{1 + \exp \left\{ -\frac{\alpha}{2\sqrt{2}} \sqrt{(\mu_j^z - \mu_i^z)^T \Sigma^{-1} (\mu_j^z - \mu_i^z)} \right\}} \right] \quad (3)$$

が得られる。本稿では  $\alpha = 2.4$  とした。詳細は紙面の都合で省略するが,  $\alpha = 2.4$  とすることで精度の良い近似が得られる。

### 5.2 2 クラスの最急降下法

前節で算出した 2 クラス間の誤認識率  $J_{ij}$  が最小になるように, 割り当てる付加情報の値を最急降下法で求める。ここでのポイントは, 割り当てる付加情報の範囲が有限であるということである。最急降下法で値を更新する際には値の範囲を制限しないほうが容易であるため, 本稿では付加情報の値を更新する際に付加情報の範囲を  $-\infty \sim +\infty$  とし, 値を出力する際に logistic 関数を用いて  $0 \sim 1$  の範囲に変換することにする。すなわち,  $-\infty \sim +\infty$  で定義される値  $\xi_i$  を用いて,  $\nu_i = l(\xi_i)$  とおく。  $\nu_i$  は実際に割り当てられる 1 次元付加情報で,  $0 \sim 1$  の値をとる。これにより,  $\mu_i^z = [\mu_{i1}^x \dots \mu_{ip}^x \ l(\xi_i)]^T$  とおける。最急降下法を用いるために, 式 (3) の誤認識率  $J_{ij}$  を変数  $\xi_j$  により微分する。紙面の制約により結果のみを示すと,

$$\begin{aligned} \frac{dJ_{ij}}{d\xi_j} = & - \frac{\exp \left\{ -\frac{\alpha}{2\sqrt{2}} \sqrt{(\mu_j^z - \mu_i^z)^T \Sigma^{-1} (\mu_j^z - \mu_i^z)} \right\}}{\left[ 1 + \exp \left\{ -\frac{\alpha}{2\sqrt{2}} \sqrt{(\mu_j^z - \mu_i^z)^T \Sigma^{-1} (\mu_j^z - \mu_i^z)} \right\} \right]^2} \\ & \cdot \frac{\alpha}{\sqrt{2}} \frac{\alpha \exp(-\alpha \xi_j)}{\{1 + \exp(-\alpha \xi_j)\}^2} \frac{\sigma_{xx} \{l(\xi_j) - l(\xi_i)\}}{\sqrt{(\mu_j^z - \mu_i^z)^T \Sigma^{-1} (\mu_j^z - \mu_i^z)}} \quad (4) \end{aligned}$$

となる。

### 5.3 $N$ クラス問題の最急降下法

前節までで2クラス問題における最急降下法を求めたが、本節ではこれを  $N$  クラス問題に拡張する。そのために  $N$  クラス問題の誤認識率を求める。  $k$  クラス  $\omega_1, \dots, \omega_k$  の分布の重なりを  $\pi_1 \cap \dots \cap \pi_k$  と表す。このとき、2クラス  $\omega_i$  と  $\omega_j$  の分布の重なり  $\pi_i \cap \pi_j$  は2クラスの誤認識率になるので  $J_{ij}$  と等価である。このようにおけば、 $N$  クラスの誤認識率  $J$  は

$$J = \sum_{i_1} \sum_{i_2 \neq i_1} \pi_{i_1} \cap \pi_{i_2} - 2 \sum_{i_1} \sum_{i_2 \neq i_1} \sum_{i_3 \neq i_1, i_2} \pi_{i_1} \cap \pi_{i_2} \cap \pi_{i_3} + \dots + (-1)^n (n-1) \underbrace{\sum_{i_1} \dots \sum_{i_N \neq i_1, \dots, i_{N-1}} \pi_{i_1} \cap \dots \cap \pi_{i_N}}_N \quad (5)$$

により求めることができる。したがって、最急降下法を用いるためには  $J$  を微分すればよい。ただし、3クラス以上の分布の重なりを解析的に求めることは困難であるため、式(5)の第1項のみで近似して用いることが多い。しかし、この方法で実験したところ、これでは近似精度が不十分であった。そこで本研究では  $\pi_{i_1} \cap \dots \cap \pi_{i_N}$  で重なっているクラスの中で最も離れた2クラス  $\omega_x, \omega_y$  の分布の重なり(誤認識率)  $J_{xy}$  を用いて3クラス以上の分布の重なりを近似する。この近似については付録で考察する。

## 6. 実験と考察

付加情報の誤りを考慮することで誤認識率が減少する可能性を確認するために、従来手法(文献[1])と提案手法を比較する。

従来手法では付加情報が誤らないという前提があったため、付加情報としては単に異なる整数を割り当てていた。一方、提案手法では0~1の実数を付加情報として割り当てる。しかし、このままでは付加情報が1以下の提案手法と1より大きくなる従来手法を公平に比較できない。そこで、まず付加情報を割り当てる範囲を揃えるため、従来手法の付加情報が0~1の範囲に収まるように定めた。すなわち、 $n$  種類の付加情報を割り当てる場合、 $0, 1, 2, \dots, n-2, n-1$  と割り当てる代わりに  $0, 1/(n-1), 2/(n-1), \dots, (n-2)/(n-1), 1$  と割り当てる。このような付加情報の等間隔割り当ては比較的無難な割り当てであり、付加情報の標準偏差が小さければ付加情報は高い確率で判別可能であるため、一定の認識性能が見込まれる。しかし、付加情報を割り当てる際に順番を考慮していないため、必ずしも誤認識率が最低になるとは限らない。例として図3の場合を考えると、クラスAとBが最も紛らわしい2クラスであるため、なるべく離れた値の付加情報を割り当てるべきである。しかし、クラス番号順に割り当てればこれらは隣り合ってしまうため、図3(b)の最適な付加情報にはならない。そこで提案手法では下記の方法でこの問題を解決する。まず、誤認識率が小さくなるように、前節で述べた最急降下法を用いて付加情報を割り当てる。ただし、最急降下法は初期値として与えた付加情報の順番に大きく依存し、前述の問題が解決できないため、付加情報の初期値を複数与え、最良のものを選択することにする。

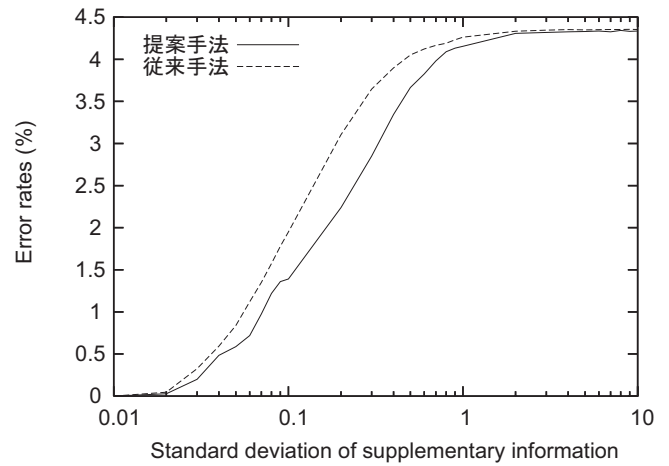


図5 実験結果。

提案手法の初期値はランダムな値100通りと従来手法と同じ等間隔割り当て1通りの合わせて101通りを試す。

特徴量として、NIST Special Database 19 [7]の数字サンプルを  $64 \times 64$  の大きさに非線形正規化 [8] した後、196次元の方向線素特徴量 [9] を抽出した。特徴ベクトルの平均と共分散行列をクラス毎に推定し、これらをパラメータとする正規乱数を発生させて、正規分布に従う人工サンプルをクラス毎に作成した。付加情報の分布は、付加情報として割り当てた値を平均とし、あらかじめ定めた値を標準偏差とする1次元の正規分布とした。したがって、認識に用いる特徴ベクトル ( $z$  に相当) は、サンプルの特徴量を表す196次元 ( $x$ ) と付加情報の1次元 ( $y$ ) を合わせた197次元の特徴ベクトルとなる。認識にはマハラノビス距離を用い、各クラスの共分散行列が等しいという理想的な場合を考えるため、全クラスの共分散行列の平均を全クラス共通の共分散行列として用いる。

使用したサンプル数は従来手法、提案手法共にクラス当たり5,500である。従来手法の内訳は学習用に4,500、テスト用に1,000である。提案手法の場合は初期値の良さを評価するために1,000サンプルを評価用として用いるため、学習用に3,500、テスト用に1,000とした。評価用サンプルは101通りの初期値の中で最良のものを決定する際にテストサンプルの代わりに用い、誤認識率が最小になる付加情報を選んだ。

付加情報の標準偏差と誤認識率の関係を図5に示す。図中の誤認識率は10回の実験の平均値である。提案手法と従来手法を比べると、付加情報の標準偏差が0.01~1の範囲では、提案手法の誤認識率が従来手法のそれを下回り、標準偏差が0.2のときに提案手法の改善効果が最大であった。一方、付加情報の標準偏差が2~10までの範囲では、提案手法と従来手法の誤認識率はほとんど差がなかった。

以下では実験結果の詳細について考察する。標準偏差が0.01~0.03の範囲では、提案手法の初期値を等間隔にしたとき、初期値をランダムにしたときの最小の誤認識率が同程度であった。一方、標準偏差が0.04~1までの間は、初期値をランダムにした方が誤認識率が減少した。その理由として、付加情報がほとんど誤らなければ従来方法で十分であるが、誤りの程度がある

程度大きくなると、従来方法では対応できず、提案手法のように誤りを考慮した付加情報の割り当てが必要になることが考えられる。それに対して、提案手法の効果が確認できなかった2~10の範囲については、付加情報の標準偏差が大きくなり過ぎたため、付加情報から得られる情報量が減少し、割り当て方法の違いが意味を持たなくなっていると考えられる。

ここで一つ注意すべきなのは、付加情報の標準偏差が大きくなったときに誤認識率が通常のパターン認識の誤認識率に収束しているということである。これは割り当て方法によらず、付加情報が誤認識率に悪影響を与えないことを示している。

## 7. まとめ

特徴量のみでは本質的に避けることができない誤認識を回避するために、付加情報を用いるパターン認識という枠組みが提案されている。この枠組みにおいては、これまで誤差を含まない付加情報のみが考慮されてきた。しかし、実世界において付加情報が誤らないという仮定は強過ぎるため、本研究では誤差を考慮した付加情報の割り当て方法について考察し、実験により、誤認識率が減少することを検証した。

実験結果から、以下の事がわかった。(1) 標準偏差が小さい(本実験では0.01~0.03)場合は、付加情報の誤りを考慮する必要がほとんど無いため、文献[1]の手法でも十分である。(2) 標準偏差が中程度(0.04~1)の場合は、付加情報の誤りを考慮することにより、誤認識率が大幅に向上する。(3) 標準偏差が大きい(2~10)場合は、付加情報自体の意味がなくなってしまうため、どのように割り当てても大差がない。

誤りのある付加情報に関して、本稿で触れられなかったことがある。誤りのない付加情報が最初に提案された文献[10]において、「何ビットの情報量を加えれば何%の認識性能が得られるのか?」という情報量と認識性能の関係に関する素朴な疑問が付加情報というアイデアの種になっていた。誤りのない付加情報の理論では付加情報の種類が情報量を表していたが、本稿において付加情報の誤りを考慮する段になって、情報量という考え方が陽に出て来なくなっている。しかし、これは誤りのある付加情報において、情報量という考え方が不要になった事を意味していない。

直感的に述べれば、2つの付加情報を類似した値を割り当てた場合と離れた値を割り当てた場合では、付加情報の誤りに対して頑健なのは後者である。したがって、後者のほうが大きな情報量を持つと考えられのだろうか?同様に分散の変化に対しては、付加情報の分散が小さければ2つの付加情報の区別が容易であるが、分散が大きければ区別が困難になる。この場合は分散が小さいほうが大きな情報量を持つと考えられのだろうか?このような点についての考察が今後の課題である。

## 文 献

- [1] 岩村雅一, 内田誠一, 大町真一郎, 黄瀬浩一, “付加情報を用いるパターン認識,” 信学論 D, vol.J90-D, no.2, pp.460-470, Feb. 2007.
- [2] 内田誠一, 岩村雅一, 大町真一郎, 黄瀬浩一, “カメラによる文字認識のためのカテゴリ情報の埋込に関する検討,” 信学論 D, vol.J89-D, no.2, pp.344-352, Feb. 2006.

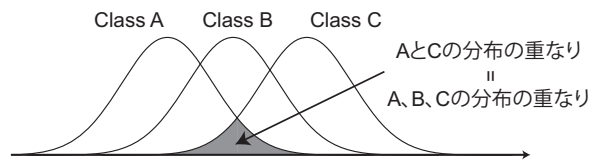


図 A-1 1次元の分布の重なり

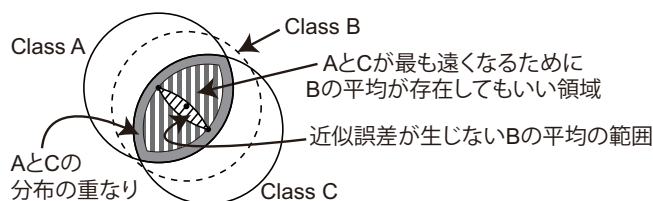


図 A-2 近似誤差が生じない領域. 図中の円は等高線.

- [3] 内田誠一, 酒井恵, 岩村雅一, 大町真一郎, 黄瀬浩一, “オートマトン制御付き最適セグメンテーション法とその実環境文字認識への応用,” 信学論 D, vol.J90-D, no.8, pp.1966-1976, Aug. 2007.
- [4] 大町真一郎, 岩村雅一, 内田誠一, 黄瀬浩一, “実環境文字認識のための面積比による付加情報埋込,” 信学論 D, vol.J90-D, no.12, pp.3246-3256, Dec. 2007.
- [5] 鳥脇純一郎, 認識工学—パターン認識とその応用—, コロナ社, 東京, 1993.
- [6] A. Birnbaum and J. Dudman, “Logistic order statistics,” The Annals of Mathematical Statistics, vol.34, no.2, pp.658-663, June 1963.
- [7] P. J. Grother, “NIST special database 19 — handprinted forms and characters database, Technical report,” National Institute of Standards and Technology, March 1995.
- [8] 山田博三, 斉藤泰一, 山本和彦, “線密度イコライゼーション—相関法のための非線形正規化法,” 信学論 (D), vol.J67-D, no.11, pp.1379-1383, Nov. 1984.
- [9] 孫寧, 田原透, 阿曾弘具, 木村正行, “方向線素特徴量を用いた高精度文字認識,” 信学論 (D-II), vol.J74-D-II, no.3, pp.330-339, March 1991.
- [10] 岩村雅一, 内田誠一, 大町真一郎, 黄瀬浩一, “認識が与える情報量—認識率100%まであと何ビット?—,” 信学技報 PRMU-742, March 2005.

## 付 録

### A. 5.3 節の近似について

本実験のように各クラスの共分散行列が等しいとき、5.3 節の近似が、近似ではなく厳密な値となる条件を示す。

まず1次元の場合について考える。図 A-1 は3クラスの分布を表している。図から両端のクラスの分布の重なりが3クラスの分布の重なりと等しいことがわかる。これは両端のクラスが重なる領域を真中のクラスが完全に覆っていることによる。

次に多次元の場合について考える。1次元のときと同様に考えれば、最も遠い2クラスの重なり領域を他のクラスが完全に覆っていれば良いということになる。図 A-2 は多次元の3クラスの分布を表しており、クラス A と C が最も遠い2クラスであるとする。この条件を満たすために、クラス B の平均は図中の縦線の領域になくなくてはならない。さらに、近似誤差が起こらないためには、クラス B がクラス A と C の重なる領域を完全に覆う位置にある必要がある。この条件を満たすためにはクラス B の平均が図中の斜線の領域になくなくてはならない。