

付加情報の一般的な割当

岩村 雅一^{†a)} 古谷 嘉男[†] 黄瀬 浩一[†] 大町真一郎^{††}
 内田 誠一^{†††}

A General Assignment of Supplementary Information

Masakazu IWAMURA^{†a)}, Yoshio FURUYA[†], Koichi KISE[†], Shinichiro OMACHI^{††},
 and Seiichi UCHIDA^{†††}

あらまし 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。この方式では、パターン認識を行う際に、付加情報と呼ばれるクラスの決定を補助する少量の情報を特徴量と同時に用いて認識性能の改善を目指す。付加情報は自由に設定でき、通常は誤認識率が最小になるように設定する。ここで問題となるのは、誤認識率が最小になる付加情報の設定方法である。常に正しい付加情報が得られるという理想的な条件においては既に問題が定式化され、付加情報の割当方法が導かれている。しかし、実環境での使用を考えると、付加情報に生じる観測誤差を考慮した割当方法が求められる。そこで本論文では付加情報の観測誤差を考慮に入れて、問題を新たに定式化する。これは付加情報が誤らない場合にも有効な一般的なものである。本論文で導いた割当方法が有効に機能することをマハラノビス距離を用いた実験で例示する。

キーワード バイズエラー，誤差のある付加情報の割当，最急降下法，正規分布，マハラノビス距離

1. ま え が き

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同一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“l”（アイ）と小文字の“l”（エル）は、フォントによっては極めて類似の外見をもち、見た目で判断することは困難である。

この問題に対処するアイデアの一つとして、パターン認識を行う際に正解クラスに関する少しのヒン

トをもらい、そのヒントを考慮に入れてどのクラスかを判断することが考えられる。前述の例の場合、「この文字は大文字である」といったヒントだけで認識可能になる。このようなヒントに相当する情報は「付加情報」と呼ばれており、付加情報を利用して認識を行う「付加情報を用いるパターン認識」という新しいパターン認識の枠組みが提案されている [1]~[5]。

付加情報を実際に付与するには、認識対象に埋め込む方法と認識対象とは全く異なるチャネルで伝送する方法が考えられる。文字認識を例にとると、前者には図 1 のように文字を加工して文字そのものに付加情報を埋め込む方法があり、後者には、文字認識率を向上させるために、帳票に記入される字種の情報をあらかじめ付加情報として帳票の枠に重畳しておく研究 [6] がある。後者の他の例としては、図 2 のようにバーコードによって付加情報を表す方法も考えられる。

このように付加情報には様々な付加方法が考えられるが、すべてに共通するのは「情報を自由に設定できる」ことである。認識対象から新たな特徴量を観測する場合は自由に設定できないため、これは付加情報とは呼べない。最初に述べた文字の例のように、外観が完全に同一であれば、新たに観測した特徴量も同一に

[†] 大阪府立大学大学院工学研究科，堺市
 Graduate School of Engineering, Osaka Prefecture University, 1-1 Gakuencho, Naka-ku, Sakai-shi, 599-8531 Japan

^{††} 東北大学大学院工学研究科，仙台市
 Graduate School of Engineering, Tohoku University, 6-6-05 Aoba, Aramaki, Aoba, Sendai-shi, 980-8579 Japan

^{†††} 九州大学大学院システム情報科学研究院，福岡市
 Faculty of Information Science and Electrical Engineering, Kyushu University, 744 Motoooka, Nishi-ku, Fukuoka-shi, 819-0395 Japan

a) E-mail: masa@cs.osakafu-u.ac.jp

なってしまう懸念があるが、自由に設定できる付加情報ならばその心配はない。つまり、認識に有効な特徴量を抽出する努力を尽くしておお認識性能を向上する可能性を提供するという点に付加情報の存在価値がある。

実際に、付加情報を用いたパターン認識を実現する場合、以下で述べる二つの問題が発生する。

1 番目の問題は、物理的な制約などによって、付加情報の情報量が制限されることである。例えば、図 2 のバーコードは情報量に応じて大きさが決まるため、どれだけの情報量を埋め込めるかは紙面の大きさやレイアウトの制約などで決まってしまう。そのため、一般に付加情報のみですべてのクラスを区別することはできず、限られた情報量でなるべく高い認識性能を達成できる付加情報の割当方法が求められる。この問題に対しては、付加情報の情報量に応じて達成可能な認識性能が定式化され、認識性能が最大となる付加情報の割当が導かれている [1]。

2 番目の問題は、付加情報も誤る可能性があること

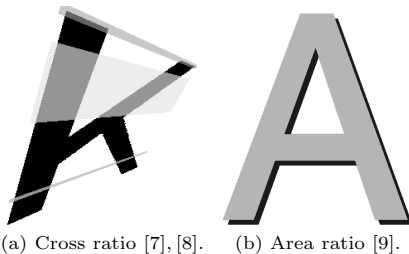


図 1 文字への付加情報の埋込例。(a) 縞模様の複比を利用した付加情報 [7], [8]。(b) 文字領域と影領域の面積比を利用した付加情報 [9]

Fig. 1 An example of supplementary information embedded in a character.

Supplementary information

あらかし、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。この方式では、パターン認識を行う際に、付加情報と呼ばれるクラスの決定を補助する少量の情報を特徴量と同時に用いて認識性能の改善を目指す。付加情報は自由に設定でき、通常は誤認識率が最小になるように設定する。ここで問題となるのは、誤認識率が最小になる付加情報の設定方法である。常に正しい付加情報が得られるという理想的な条件においては既に問題が定式化され、付加情報の割り当て方法が導かれている。しかし、実環境での使用を考えると、付加情報に生じる観測誤差を考慮した割り当て方法が求められる。そこで本論文では付加情報の観測誤差を考慮に入れて、問題を新たに定式化する。これは付加情報が誤らない場合にも有効な一般的なものである。本論文で導いた割り当て方法が有効に機能する

Recognition target

図 2 バーコードによる付加情報の例
Fig. 2 An example of supplementary information given by a barcode.

である。付加情報として複比や面積比を用いる図 1 が良い例であるが、付加情報を特定の値に設定しても、たいていは誤差の影響により、設定した値と若干異なる値が得られる。値の変動が小さければ、観測値からもとの値を推定できるため、認識性能への影響はほとんどないと考えられるが、大きな変動に対しては付加情報の誤差を考慮する必要が生じる。前述した文献 [1] ではこのような変動を一切考慮せず、付加情報は常に正しいとしているため、実際にはその理論が適用できない場合も多かった。

そこで本研究では、2 番目の問題を解決するために付加情報が誤ることを前提として理論の構築を行い、誤認識率を最小にする最適な付加情報の割当方法を定式化する。そして、提案する割当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マハラノビス距離を用いた実験で例示する。

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 3 は 3 クラス問題を模式的に表している。図 3 (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

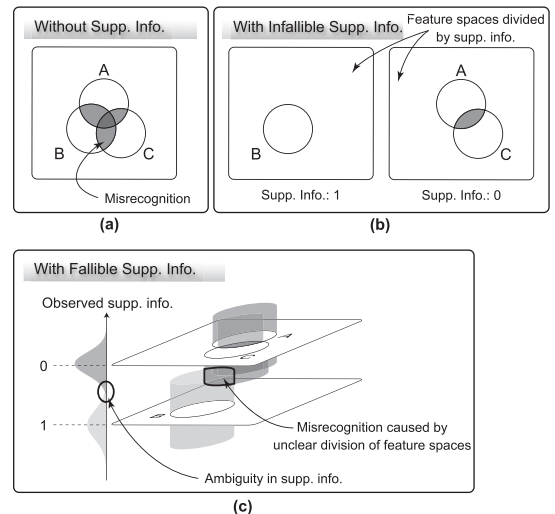


図 3 付加情報を用いるパターン認識の理論体系

Fig. 3 Pattern recognition models. (a) Without supplementary information. (b) With infallible supplementary information [1] (conventional method). (c) With fallible supplementary information (the proposed model).

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。これはすなわち、ベイズエラーである。

これに対して、図 3 (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 (ベイズエラー) を回避するために、誤差のない付加情報を用いる。図 3 (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、認識に付加情報を用い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、誤認識が起これやすい 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 (ベイズエラー) を回避できる。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 3 (c) は本研究で検討する付加情報の誤差を考慮する場合である。この場合は付加情報を図 3 (b) と同様に与えても、付加情報の誤差のために特徴空間を完全に分割できず、図 3 (b) では起これ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 3 (b) の割当方法が付加情報が誤る図 3 (c) においても良い割当方法であるのか、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「付加情報の良い割当」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当方法 [1]

付加情報が誤る場合を考える準備として、付加情報が誤らない場合の割当方法 [1] の概略を述べる。本論文ではここで述べる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。

まず、ある 3 クラス問題の特徴量の分布が図 4 (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 3 (a) のモデルに相当する。図中の灰色の領域は不可避な誤認識 (ベイズエラー) を表し、その面積が誤認識率を表す。

次に、図 3 (b) のように、この問題に付加情報を導入する。付加情報が誤らないとすれば、このモデルでは何種類の付加情報を用いるかが認識性能を左右する。割り当てる付加情報が 1 種類の場合 (すべてのクラスに同じ付加情報を与える場合) 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

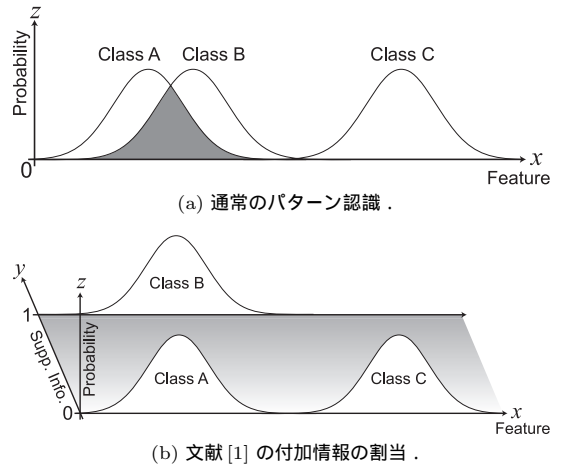


図 4 各クラスの特徴量の分布と付加情報
Fig. 4 (a) A three-class problem and (b) the best assignment of infallible supplementary information given in [1].

減少する。図 4 の場合は 3 種類で誤認識が起これなくなり、そのときの割当は一意に定まる。

一方、2 種類の場合は 3 通りの割当方法が考えられる。このうち誤認識率が最小になるのは付加情報を含めたクラスの分布の重なりが最小になる割当であり、すなわちこれは図 4 (b) の割当である。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当方法を与えている^(注1)。図 4 (b) は、図 4 (a) に付加情報を表す y 軸を追加して得られたものである。付加情報が誤らないとしているため、 y 軸方向の分布の分散が 0 であり、図 3 (b) のように空間が完全に分割されている。

4. 誤差を含む付加情報の割当方法

前章の図 4 (b) では付加情報が誤らないことを y 軸方向に分散をもたない (幅がない) 分布で表した。一方、付加情報が誤差を含む場合は y 軸方向に分散をもつ分布になる。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、誤認識が生じ得る。本章では、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「良い割当」になるのかについて検討する。

4.1 問題設定

提案手法で考える問題を明らかにするために、設計

(注1): 文献 [1] では誤認識率が最小の割当の求め方を与えているが、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。

者ができることとできないことを整理しておく．

- 割り当てる付加情報 [自由に設定可能]: 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．

- 付加情報の分布形状 [設定不可能]: 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，観測誤差のために，割り当てた値をそのまま観測できるわけではない．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．この分布は環境によって決まるため，観測することはできても，設計者が変更することはできない．

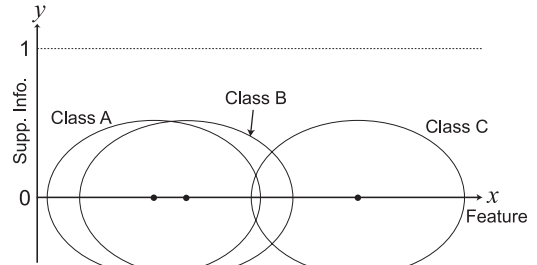
- 特徴量の分布 [設定不可能]: 特徴量とは，通常のパターン認識で用いる特徴量を意味する．特徴量は，サンプルを観測することによって得られる．設計者はどの特徴を使うかの選択はできるが，観測の結果得られる特徴量そのものを変更することはできない．したがって，特徴量の分布を設計者は変更できない．

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 [7] ~ [9] 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．そのため，一般に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 $0 \sim 1$ の間とする．もし実際に割り当てることができる範囲が $0 \sim 2$ であれば 2 倍すればいいので，このことによって一般性は失われない．

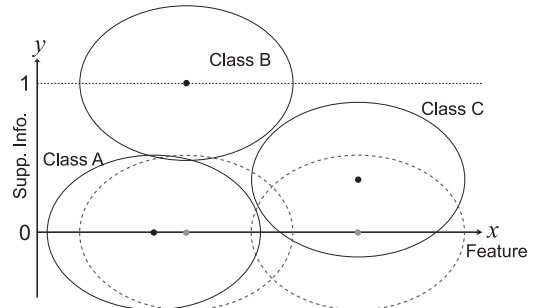
4.2 付加情報の与え方

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にするのかについて検討する．図 5 (a) は図 4 (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 4 (a) と違い，付加情報が誤るため，付加情報の分布は y 軸方向の広がりをもっている．図 5 (a) では付加情報がすべて同じであるため，誤認識率は図 4 (a) と同じである．なお，サンプルの特徴量 x と付加情報 y はクラスが決まれば独立となる (条件付独立) と仮定すれば，図 5 の楕円は x - y 平面で傾かない．

図 5 (a) において付加情報の値が決まったときの誤認識率について考えると， y 軸の付加情報と x 軸の特徴量のどちらも特徴量とみなすことができるので，通常のパターン認識の誤認識率と類似の方法で計算できると考えられる．したがって，誤りのある付加情報を



(a) The conventional model (the same value is assigned as the supplementary information).



(b) The best assignment of the supplementary information.

図 5 各クラスの特徴量と誤差がある付加情報の分布．楕円は確率分布の等高線．(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

Fig.5 Assignments of supplementary information and class distributions. The contours in the figure are equiprobable lines.

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よい」ということになる．ただし，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のみであり，「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されたい．

4.3 問題の定式化

前節で得られた知見を定式化する．まず，特徴量について考える．認識対象の N クラスを $\omega_1, \dots, \omega_N$ とおき，クラス ω_i のサンプルが生成される確率を $P(\omega_i)$ とおく．観測されるサンプルの特徴量を p 次元ベクトル x で表せば， x は ω_i に依存して決まるので， ω_i が与えられたときの x の確率密度は $p(x|\omega_i)$ と書ける．次に，付加情報について考える．各クラス ω_i のサンプルに割り当てる付加情報を ν_i とする．観測される付加情報 y は割り当てる付加情報の値 ν_i に依存して決まるので，以上より， ω_i が与えられたときの y の確

率密度は $p(y|\omega_i; \nu_i)$ と書ける .

以上のように定義すれば , ベイズの定理より

$$P(\omega_i|x, y; \nu_i) = \frac{p(x|\omega_i)p(y|\omega_i; \nu_i)P(\omega_i)}{p(x, y)} \quad (1)$$

が成り立ち , 誤認識率 $E(\nu_1, \dots, \nu_N)$ は次式で与えられる .

$$\begin{aligned} E(\nu_1, \dots, \nu_N) \\ = 1 - \int_y \int_x \max_i \{P(\omega_i|x, y; \nu_i)p(x, y)\} dx dy \end{aligned} \quad (2)$$

この誤認識率を最小にするために解くべき問題は

$$(\nu_1, \dots, \nu_N) = \underset{\nu_1, \dots, \nu_N}{\operatorname{argmin}} E(\nu_1, \dots, \nu_N) \quad (3)$$

を満たす ν_1, \dots, ν_N の組を求めることに等しい .

式 (3) で定式化した問題を解くには , ν_1, \dots, ν_N の空間を探索する必要がある . 本論文では見通しの良い問題を考えて , 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章以降で確認する .

5. 最急降下法による解法—正規分布で共分散行列が共通な場合—

本章では最急降下法を用いて誤りのある付加情報を割り当てる方法について述べ , 次章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 x, y に相当) がともに正規分布に従い , 全クラスの共分散行列が等しい場合について , マハラノビス距離^(注2)を用いた認識実験を行う . 各クラスの事前確率は等しい ($P(\omega_i) = 1/N$) とする .

誤認識率を求める手順はおおよそ以下のとおりである . (1) 2 クラスの誤認識率を求める . (2) N クラスの誤認識率を各 2 クラスの誤認識率を組み合わせることにより求める . 以下では , $z \equiv (x, y)$ の分布を多次元正規分布として扱う .

5.1 2 クラスの誤認識率

最初にクラス ω_i と ω_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^(注3) . 図 6 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

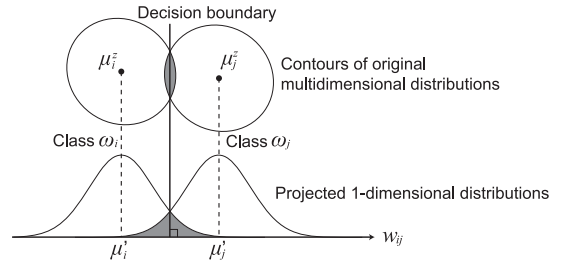


図 6 多次元正規分布を 1 次元に射影する様子
Fig. 6 Projection of a multidimensional normal distribution onto an axis.

た模式図である . 本節で考える問題で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 ω_i の多次元正規分布の平均を μ_i^z , 全クラス共通の共分散行列を Σ とおく . クラス ω_i と ω_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 (文献 [10] の p.120) . 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 , $w'_{ij} = \frac{w_{ij}}{|w_{ij}|}$ 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 $\mu'_i = w'^T_{ij} \mu_i^z$, $(\sigma')^2 = w'^T_{ij} \Sigma w'_{ij}$ となる^(注4) .

各クラスの事前確率は 1/2 であるので ,

$$F(t) = \int \frac{1}{\sqrt{2\pi(\sigma')^2}} \exp\left(-\frac{t^2}{2(\sigma')^2}\right) dt \quad (4)$$

$$M^z_{ij} = \sqrt{(\mu_j^z - \mu_i^z)^T \Sigma^{-1} (\mu_j^z - \mu_i^z)} \quad (5)$$

とおけば ,

$$\begin{aligned} J_{ij} &= 2 \int_{\frac{\mu'_j - \mu'_i}{2}}^{\infty} P(z|\omega_i) P(\omega_i) dt \\ &= F(\infty) - F\left(\frac{(M^z_{ij})^2}{2 \|\Sigma^{-1}(\mu_j^z - \mu_i^z)\|}\right) \end{aligned} \quad (6)$$

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

(注2): マハラノビス距離は , ベイズ決定則において , データの分布が正規分布であり , 全クラスの共分散行列が等しいという条件のもとで導出される (文献 [10] の p.39) .

(注3): 文献 [11] , pp.85-87 と同様の手順である .

(注4): 同様にクラス ω_j の軸上の分布の平均は $\mu'_j = w'^T_{ij} \mu_j^z$ となる .

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方が容易であるため、本論文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、値を出力する際にロジスティック関数 $l(x) = \frac{1}{1+\exp(-x)}$ を用いて $0 \sim 1$ の範囲に変換することにする。すなわち、実際に割り当てられる 1 次元の付加情報 ($0 \sim 1$ の値をとる) ν_i を $\nu_i = l(\xi_i)$ とおき、 ξ_i は $-\infty \sim +\infty$ で定義される値とする。これにより、 $\mu_i^z = [\mu_{i1}^z \ \dots \ \mu_{ip}^z \ l(\xi_i)]^T$ とおける。ここで μ_{il}^z は、クラス ω_i における l 番目の特徴量の平均である。そして、式 (6) の誤認識率 J_{ij} を変数 ξ_j により微分して得られる

$$\begin{aligned} \frac{dJ_{ij}}{d\xi_j} = & -\frac{2\tau^2 \|\Sigma^{-1}(\mu_j^z - \mu_i^z)\|^2 - (M_{ij}^z)^2}{2\sqrt{2\pi} (M_{ij}^z)^2 \tau^4 \|\Sigma^{-1}(\mu_j^z - \mu_i^z)\|^2} \\ & \cdot \{l(\xi_j) - l(\xi_i)\} \frac{\exp(-\xi_j)}{\{1 + \exp(-\xi_j)\}^2} \\ & \cdot \exp\left(-\frac{(M_{ij}^z)^2}{8}\right) \end{aligned} \quad (7)$$

を最急降下法で用いる。ここで τ^2 は付加情報の分散である。

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。このようにおけば、 N クラスの誤認識率 J は

$$\begin{aligned} J = & \sum_{i_1} \sum_{i_2 > i_1} \pi_{i_1} \cap \pi_{i_2} \\ & - \sum_{i_1} \sum_{i_2 > i_1} \sum_{i_3 > i_2} \pi_{i_1} \cap \pi_{i_2} \cap \pi_{i_3} + \dots \\ & + (-1)^N \underbrace{\sum_{i_1} \dots \sum_{i_N > i_{N-1}} \pi_{i_1} \cap \dots \cap \pi_{i_N}}_N \end{aligned} \quad (8)$$

により求めることができるので、最急降下法を用いるためには式 (8) を ξ_j で微分すればよい。

2 クラス ω_i と ω_j の分布の重なり $\pi_i \cap \pi_j$ は 2 クラスの誤認識率 J_{ij} に事前確率の比である $2/N$ を乗じ

たものであり、式 (7) が使える。しかし、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。そのため、通常は式 (8) を J_{ij} に関する第 1 項のみで近似して用いることが多い。本研究では更に精度を上げるために $\pi_{i_1} \cap \dots \cap \pi_{i_N}$ で重なっているクラスの中でマハラノビス距離において最も離れた 2 クラス ω_s, ω_t の分布の重なり (誤認識率) J_{st} を用いて、3 クラス以上の分布の重なりを

$$\pi_{i_1} \cap \dots \cap \pi_{i_N} \sim \frac{2}{N} \min_{s,t \in \{i_1, \dots, i_N\}, s \neq t} J_{st} \quad (9)$$

と近似する。この近似については付録で考察する。式 (8) に近似を施し、変数 ξ_j により微分すれば

$$\begin{aligned} \frac{dJ}{d\xi_j} = & \frac{2}{N} \sum_{i_1} \sum_{i_2 > i_1} \frac{dJ_{i_1 i_2}}{d\xi_j} \\ & - \frac{2}{N} \sum_{i_1} \sum_{i_2 > i_1} \sum_{i_3 > i_2} \\ & \cdot \frac{d}{d\xi_j} \left\{ \min_{s,t \in \{i_1, i_2, i_3\}, s \neq t} J_{st} \right\} + \dots \\ & + (-1)^N \frac{2}{N} \underbrace{\sum_{i_1} \dots \sum_{i_N > i_{N-1}}}_N \\ & \cdot \frac{d}{d\xi_j} \left\{ \min_{s,t \in \{i_1, \dots, i_N\}, s \neq t} J_{st} \right\} \end{aligned} \quad (10)$$

が得られる。次章の実験ではこの式を用いる。

6. 実験と考察

付加情報の誤りを考慮することで文献 [1] の手法 (従来手法) よりも本論文で提案した手法 (提案手法) の誤認識率が減少することを示す。実験では、付加情報の標準偏差 σ と誤認識率の関係を調べた。

準備として、従来手法と提案手法で割り当てる付加情報の範囲をそろえる。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クラス番号の順に割り当てていた。一方、提案手法では $0 \sim 1$ の実数を割り当てる。付加情報の範囲が違う両者を公平に比較できないため、従来手法の付加情報を $0 \sim 1$ の範囲に変更する。具体的には、割り当てる付加情報が n 種類の場合には、 $0, 1, 2, \dots, n-2, n-1$ 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当は比較的無難な割当であるため、従来手法でも一定の認識性能が

見込まれる。

一方の提案手法では、前節で導いた最急降下法を用いて誤認識率が小さくなるように付加情報を割り当てるので、従来手法より適切に割り当てることができる。ただし、最急降下法で得られる付加情報は、初期値として与えた付加情報の順番に大きく依存するため、より良い値に収束するためには初期値を複数与え、最良のものを選択する必要がある。そのため、提案手法の初期値として、ランダムな値 100 通りと従来手法と同じ等間隔割当 1 通りの合計 101 通りを試す。

特徴量として正規分布に従う人工サンプルを用いる。作成方法は、まず NIST Special Database 19 [12] の数字サンプルを 64×64 の大きさに非線形正規化 [13] し、更に 16×16 の大きさに縮小した画像の画素値を要素とする 256 次元ベクトルを生成する。そして、クラスごとに特徴ベクトルの平均を推定し、全クラスのサンプルから共通の (pooled) 共分散行列を推定した後、これらをパラメータとする正規乱数を発生させる。推定にはクラスごとに 36,000 サンプルを用いる。これ以降、ここで作成した人工サンプルのみを実験に用いる。

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、 τ 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 256 次元 (x) と付加情報の 1 次元 (y) を合わせた 257 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、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、提案手法ともにクラス当たり 10,000 である。従来手法の内訳は学習用に 9,000、テスト用に 1,000 である。提案手法についても同様に、学習用に 9,000、テスト用に 1,000 とした。ただし学習用のうち 1,000 サンプルについては、101 回分の収束結果から最良のもの一つを選出することのみを用いた。

付加情報の標準偏差と誤認識率の関係を図 7 に示す。図中の誤認識率 (縦軸) は 10 回の実験の平均値とその標準偏差を示す。参考のために、付加情報を用いない場合の誤認識率も示す。提案手法と従来手法を比較すると、付加情報の標準偏差に応じていくつかの傾向が見て取れる。まず、付加情報の標準偏差が中程度のとき (0.08 ~ 1 の範囲) は提案手法の誤認識率が従来手法に比べて 0.1% 以上減少し、付加情報の誤り

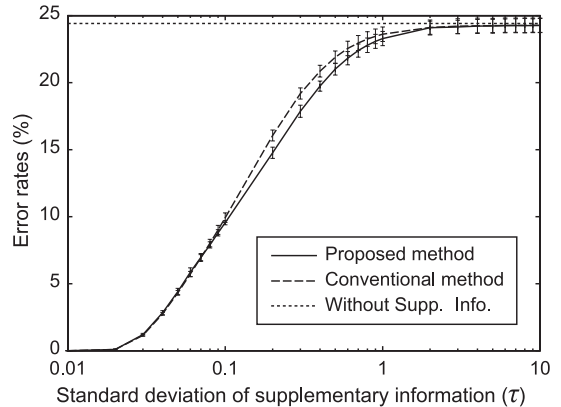


図 7 実験結果
Fig. 7 Experimental result.

を考慮する効果が確認できた。誤認識率が最も減少したのは標準偏差が 0.3 のときで、減少幅は 1.3% であった。次に、付加情報の標準偏差が小さいとき (0.01 ~ 0.07 の範囲) と大きいとき (2 ~ 10 の範囲) では提案手法と従来手法の誤認識率は同程度であった。これらの理由として下記のことが考えられる。

- (1) 標準偏差が小さいときは付加情報の誤りがほとんどないため、従来手法で十分である。
- (2) 誤りの程度がある程度大きくなると、従来手法では対応できず、提案手法のように誤りを考慮した付加情報の割当が必要になる。
- (3) 標準偏差が大きくなりすぎると、付加情報から得られる情報量が減少し、割当方法の違いが意味をもたなくなる。ただし、付加情報を用いない場合より誤認識率が高くなることはない。付加情報を用いない場合は標準偏差が無限大の場合に相当するためである。

7. む す び

特徴量のみでは本質的に避けることができない誤認識 (バイズエラー) 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すぎるため、本論文では誤差を考慮した付加情報の割当方法を導出し、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、付加情報を割り当てる際に付加情報の誤りを考慮することによって誤認識率を減らせることが確認できた。

本論文では 5. 以降, 特徴ベクトルと付加情報が多次元正規分布に従うと仮定し, マハラノビス距離で認識した. それ以外の分布に対する実験やマハラノビス距離以外の識別器を用いた実験は今後の課題である. また, 本論文では文献 [1] で議論されているリジェクトのための付加情報について議論していない. この点についても今後の課題である.

謝辞 非常に丁寧な査読と有益な御指摘を頂いた査読者に感謝致します. 本研究の一部は, マツダ財団研究助成 (No.05KK-276) による.

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] E.H. Kuo, Assist channel coding for improving optical character recognition, Master thesis, MIT, 2000.
- [3] D.H. Greene, L.T. Niles, and E.H. Kuo, “Assist channel coding with character classifications,” United States Patent 6,862,113, March 2005.
- [4] D.H. Greene and E.H. Kuo, “Assist channel coding with vertical block error correction,” United States Patent 6,768,560, May 2000.
- [5] D.H. Greene and A.C. Popat, “Assist channel coding with convolution coding,” United States Patent 6,628,837, May 2000.
- [6] 島村太郎, 朱 碧蘭, 櫻田武嗣, 中川正樹, “アクティブ帳票システム的设计と開発,” 信学論 (D-II), vol.J87-D-II, no.12, pp.2091-2103, Dec. 2004.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] R.O. Duda, P.E. Hart, and D.G. Stork, Pattern classification, 2nd edition, John Wiley & Sons, 2000.
- [11] K. Fukunaga, Introduction to statistical pattern recognition, 2nd ed., Academic Press, 1990.
- [12] P.J. Grother, “NIST special database 19 — Handprinted forms and characters database,” Technical Report, National Institute of Standards and Technology, March 1995.
- [13] 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.

付 録

5.3 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 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分かる. 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, マハラノビス距離において最も遠い 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よいということになる. この条件を図 A.2 に示す確率密度の等高線を用いて考える. まず, ある確率密度の等高線について考えたとき, クラス A と C がマハラノビス距離において最も遠い 2 クラスであるためには, 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. 更に, 分布が単調減少であることを考慮すれば, 近似誤差が起こらないための必要条件は, クラス B の等高線がクラス A と C の重なる領域を完全に囲む位置にあることとなる. これはクラス B の平均が図中の斜線の領域にあることと等しい. 以上のこ

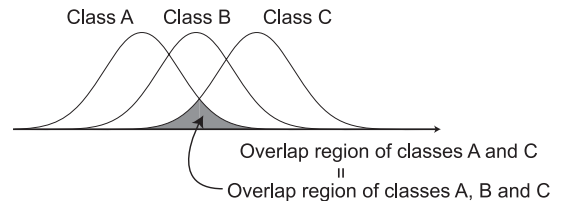


図 A.1 1次元の分布の重なり
Fig. A.1 Overlap of one-dimensional distributions.

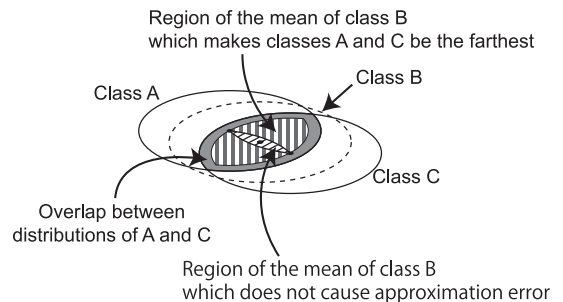


図 A.2 多次元の分布の重なり. 図中の楕円は確率密度の等高線
Fig. A.2 Overlap of multi-dimensional distributions.

とを踏まえて改めてクラス B の分布がクラス A と C の分布を覆う条件を考えると、確率密度 0 以上のすべての等高線において前述の斜線部の条件を満たせばよいことが分かる。前述の斜線領域は等高線の半径が大きくなる（確率密度が小さくなる）につれて狭くなり、確率密度 0 の極限では A と C を結ぶ線分になるため、5.3 の近似が近似ではなく厳密な値となるのは、クラス B の平均がクラス A と C を結ぶ線分上にあるときとなる。

（平成 21 年 11 月 24 日受付，22 年 1 月 12 日再受付）



岩村 雅一（正員）

平 10 東北大・工・通信卒。平 15 同大学院博士課程了。同年同大学院工学研究科助手。平 16 阪府大大学院工学研究科助手。現在助教。博士（工学）。パターン認識，文書画像解析，物体認識，情報検索などの研究に従事。平 18 本会論文賞，平 19 IAPR/ICDAR The Best Paper Award 各受賞。IEEE，情報処理学会各会員。



古谷 嘉男

平 18 阪府大・工・情報卒。平 20 同大学院修士課程了。在学中は付加情報を用いるパターン認識の研究に従事。



黄瀬 浩一（正員）

昭 61 阪大・工・通信卒。昭 63 同大学院博士前期課程了。同年同大学院博士後期課程入学。平 2 阪府大・工・電気助手。平 5 同情報助手，平 8 講師，平 11 助教授を経て，平 17 同大学院工学研究科教授，現在に至る。博士（工学）。その間，平 12～13 ドイツ人工知能研究センター客員教授。文書画像解析，情報検索などの研究に従事。平 18 本会論文賞，平 19 IAPR/ICDAR The Best Paper Award 各受賞。情報処理学会，人工知能学会，電気学会，IEEE，ACM 等各会員。



大町真一郎（正員）

昭 63 東北大・工・情報卒。平 5 同大学院博士課程了。同年同大情報処理教育センター助手。平 8 同大工学部助手。平 11 同大学院工学研究科助教授。平 21 同教授。博士（工学）。その間，平 12～13 米国ブラウン大学客員准教授。パターン認識，コンピュータビジョン，並列処理，文字認識システムの開発などの研究に従事。平 19 MIRU 長尾賞，平 19 IAPR/ICDAR The Best Paper Award 各受賞。IEEE，情報処理学会，人工知能学会等各会員。



内田 誠一（正員）

平 2 九大・工・電子卒。平 4 同大学院修士課程（情報）了。セコム（株）勤務を経て，現在，同大システム情報科学研究院情報知能工学部門教授。博士（工学）。画像パターン・時系列パターンの解析・認識に関する研究に従事。平 15 本会 PRMU 研究奨励賞，平 18 MIRU 長尾賞，平 19 IAPR/ICDAR The Best Paper Award，平 20 本会論文賞各受賞。IEEE，情報処理学会各会員。