

## 付加情報の誤りと誤認識率の関係

古谷 嘉男<sup>†</sup> 岩村 雅一<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院 〒819-0395 福岡市西区元岡 744

E-mail: †yoshio@m.cs.osakafu-u.ac.jp, †{masa,kise}@cs.osakafu-u.ac.jp,

††machi@ecei.tohoku.ac.jp, †††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新しい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．これまでに，付加情報の復元を誤らないという仮定の下で，誤認識率を理論的に最小にする方法が提案されている．しかし，現実世界で付加情報の復元を行うと，観測誤差があるため付加情報の復元を誤らないということは稀である．そこで，本稿では，付加情報の復元を誤る際に，どのようにして付加情報を与えれば誤認識率を減少させることが可能であるかについて検討を行い，それに基づく実験を行うことで，誤認識率が減少する可能性があることを示す．

キーワード 付加情報，混同行列，ベイズ識別境界，マハラノビス距離，観測誤差

## Relationship Between Errors of Supplementary Information and Misrecognition Rates

Yoshio FURUYA<sup>†</sup>, Masakazu IWAMURA<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 Motooka, Nishi, Fukuoka, 819-0395 Japan

E-mail: †yoshio@m.cs.osakafu-u.ac.jp, †{masa,kise}@cs.osakafu-u.ac.jp,

††machi@ecei.tohoku.ac.jp, †††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 classifier's output and *supplementary information* suggesting the true class. Under the condition that supplementary information does not contain error, we have proposed a theory that reduces misrecognition rates. However, in the real world, we cannot observe any measure without error. Thus, in this paper, we discuss how to reduce misrecognition rates using the erroneous supplementary information, and show the possibility to reduce misrecognition rates experimentally.

**Key words** supplementary information, confusion matrix, Bayesian decision boundary, Mahalanobis distance, observation error

### 1. ま え が き

人間にも機械にも理解可能な情報伝達手段を目指して，付加情報を用いるパターン認識という方式が検討されている [1] ~

[4]．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頑健なパターン認識を実現する．このような正解クラスに関する情報を付加情報と呼ぶ．

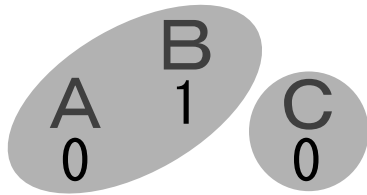


図 1 従来の誤りのない付加情報でうまく認識できる認識問題

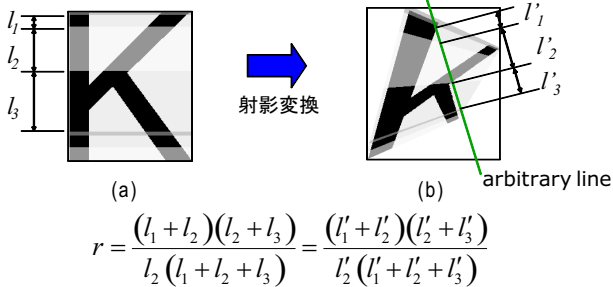


図 2 付加情報の例

例えば、図 1 の例を考えてみる．図 1 は 3 クラス問題を模式的に表したものである．図中の楕円は、識別しにくいクラスを表している．すなわち、通常のパターン認識によって、クラス C に属しているパターンを識別するとき、そのパターンがクラス C に属していることは容易に判断でき、クラス A やクラス B に属しているパターンを識別するとき、そのパターンがクラス C に属していると誤認識することはないがクラス A かクラス B のどちらに属しているのか判断しにくいことを表している．このような誤認識を減少させるために、識別するパターンに 0 や 1 という付加情報を与えることを考える．付加情報がもし誤りなく伝達可能であるならば、クラス A に属しているパターンとクラス B に属しているパターンに 0 と 1 という異なる付加情報を与えることで、これらを誤りなく認識することが可能となる．

しかし、現実世界では埋め込まれた付加情報を常に正しく得られるという状況は稀である．例えば、カメラを用いた文字認識の認識率を向上させるために文献 [2], [3] では、図 2 のような文字の背景に縞模様をつけ、その縞模様から得られる複比を付加情報とする方法を提案している．複比は射影不変量であるため、理論的には射影変換を受けても常に埋め込んだ付加情報が復元可能である．しかし、現実はこの付加情報を復元するためにカメラなどで撮影した場合、観測誤差が生じる．この誤差が大きくなるにつれて、付加情報の復元を誤る可能性は高くなる．文献 [4] にも面積比を用いた付加情報の別の埋め込み方が提案されているが、やはり付加情報が常に正しく得られることは稀である．そこで、本研究では観測値が誤差を含む現実に則した形に理論を発展させ、どのように付加情報を割り当てれば誤認識率が減少するのかを検討する．

## 2. 付加情報を用いた認識の理論

議論に入る前に、付加情報理論における本研究の位置づけを確認しておく．

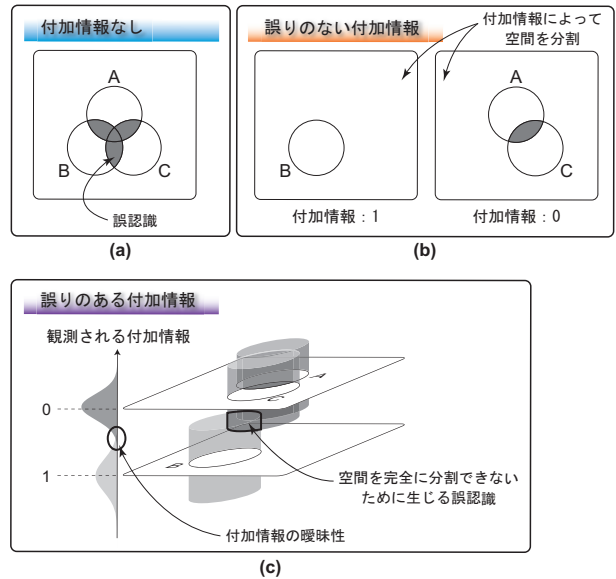


図 3 付加情報を用いた認識の理論

図 3 はある 2 次元の特徴量を持った 3 クラス問題の例である．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を用いる．従って、図 3(a) のように、重なった領域があると、誤認識を無くすることはできない．

付加情報を用いるパターン認識ではこのような誤認識を回避するために、あらかじめ認識対象にクラスを表す情報である付加情報を埋め込んでおき、識別器の認識結果と付加情報の両方を用いて、誤認識率を減少させる．図 3(b) では、クラス A とクラス C に 0 の付加情報を、クラス B に 1 の付加情報をあらかじめ埋め込んでいる．この付加情報を誤りなく復元できるとすれば、付加情報が図 3(b) のように識別空間を分割する．従って、付加情報を復元することにより、0 の付加情報を得ることができれば、クラス A かクラス C のどちらかに属しているかを識別すればよく、1 の付加情報を得ることができれば、クラス B に属していることが付加情報だけからわかる．このようにして、付加情報を用いることにより、通常のパターン認識では回避することのできなかった誤認識を減少させることが可能である．

しかし、付加情報の復元を誤らないという仮定は現実世界では起こり得ない．これは、付加情報を復元する際に、観測誤差が生じるからである．このように、付加情報が誤差を含むと、観測される付加情報は 0 や 1 のような離散値ではなく、図 3(b) のように連続値となる．つまり、付加情報の誤差がある場合、図 3(b) の例に示すように、識別空間を完全に分割することができない．本稿ではこのように付加情報が誤差を含む場合に、どのようにして誤認識率を減少させるかについての検討を行う．

## 3. 常に正しい付加情報が得られる場合の付加情報の割り当て方

誤差を含む付加情報を用いるパターン認識を説明する前に、理解のしやすさの観点から、付加情報を用いるパターン認識に

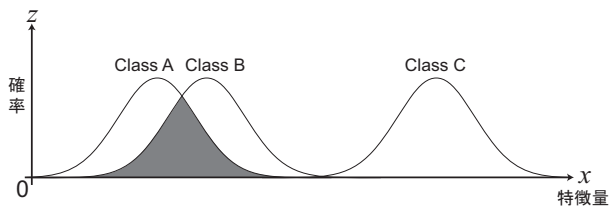


図 4 特徴量とサンプルの発生確率

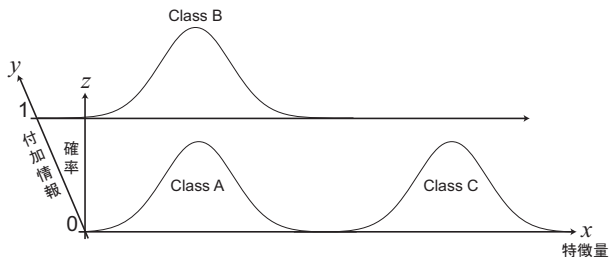


図 5 付加情報を与えた 3 クラス

において、誤差を含まない付加情報における特徴空間について考え、次節で誤差を含む付加情報における特徴空間について考える。

図 3(b) の付加情報を与えていない状態の特徴量とその特徴量におけるサンプルの発生確率を表すグラフは図 4 で表される。図 4 は、例として 1 次元の正規分布に従う特徴量を持つサンプルの発生確率を表しており、横軸が特徴量、縦軸がサンプルの発生確率を表している。図ではサンプルの分布に正規分布を用いているが、説明のために用いているだけであり、正規分布以外の分布でも適用可能であることをここで断っておく。図 4 中の灰色の面積は、誤認識率が最小となるときの誤認識率を表す。これは、ベイズ決定理論から導かれる [5]。

ベイズ決定理論により導かれる誤認識を減少させるように文献 [1] では付加情報を割り当てる。付加情報は多数割り当てればそれだけ誤認識率が減少する。本節では、図 4 に与える付加情報を例として 2 種類に決めるとき、誤認識率が最小になるように付加情報を割り当てるかを考える。図 4 では、クラス A とクラス B の誤認識が多いことがわかるので、クラス A とクラス B の誤認識を無くすように付加情報を与える。クラス A、クラス C に 0 の付加情報、クラス B に 1 の付加情報を与えると、誤認識率が最小になり、特徴量のグラフは図 5 となる。図 5 は  $x$  軸が特徴量を表し、 $y$  軸が付加情報の種類を表し、 $z$  軸がサンプルの発生確率を表す。このように、誤差を含まない付加情報を用いる場合は、 $x$  軸方向にのみ分散を持ち、 $y$  軸方向には分散を持たないグラフとなる。つまり、ある付加情報の  $x$  軸方向にのみ分散を持つグラフとなるため、少しでも付加情報が違えば付加情報を完全に区別でき、識別器にとって強力な情報となる。

本稿ではこの手法を従来手法と呼ぶ。しかし、実際には付加情報の復元は誤るため、このような理想的な誤認識率の減少は見込めない。従って、誤差を含む付加情報を用いる場合にはどのように考えるのかについて次節で述べる。

#### 4. 誤差を含む付加情報を用いる場合の付加情報の割り当て方

前節では、誤差を含まない付加情報を用いる場合の特徴空間について触れた。しかし、誤差を含む付加情報を用いる場合には、図 5 のように、 $x$  軸方向にのみ分散を持つような分布にはならない。この際に、どのようにして誤認識率を減少させるのかについて本節では検討する。

##### 4.1 問題設定

本節では、誤差を含む付加情報を用いる場合にどのようにして誤認識率を減少させるのかについて考える。最初にどのような問題を考えるのかについて述べる。

まず、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ため、誤差を含む付加情報を用いる際には、本稿では付加情報を特徴量として扱う。以下では設計者ができるところでできないことを明らかにする。

- 割り当てる付加情報について考える。図 2 の例では文字の背景に縞模様をつけ、その縞模様から得られる複比を付加情報としている。従って、どの複比の値を付加情報として与えるかは設計者が任意に与えることができる。

- 付加情報の分布について考える。付加情報の分布とは、前節でも少し述べたが、付加情報を復元する際に生じる環境誤差である。従って、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。

- 特徴量の分布について考える。特徴量は、サンプルを観測することによって得られる。従って、認識の際に用いる特徴量が決定されれば設計者が変更することができない。

以上をまとめると、既に決まっており、設計者が任意に設定できない値は、特徴量の分布、付加情報の分布であり、設計者が任意に設定できる値は割り当てる付加情報である。

付加情報は任意に与えることができるため、サンプルの特徴量とは独立である。また、付加情報は任意に与えることができるが、 $-\infty$  から  $+\infty$  の範囲で任意に与えることができるわけではない。なぜなら、付加情報を埋め込む際の複比  $r$  を  $-\infty$  から  $+\infty$  の範囲で与えることは物理的に不可能である。そのため、付加情報は有限の範囲で与える。有限の範囲であれば、付加情報を与える範囲を 0 から 1 の間に限定したとしても、付加情報の軸方向に定数倍することにより範囲を有限とした場合と等価であると考えられるため、一般性を失わない。従って本稿では付加情報は 0 から 1 の範囲で与える。

##### 4.2 付加情報の与え方

誤差を含む付加情報を用いる際には、図 5 のように  $x$  軸方向にのみ分散を持つのではなく、付加情報を表す  $y$  軸方向にも広がりを持つ。 $x$  軸と  $y$  軸方向に広がりを持つ分布がある際に、どのように付加情報を与えれば、誤認識率が減少するかについて本節では検討する。

図 6 はあるサンプルの発生確率の等高線を表した図である。 $\mu_i$  は分布の平均である。図 6 は付加情報がすべて同じであるため、付加情報を与えていない場合と同じである。この場合、分布の交わる面積が大きいため、誤認識率は最も大きくなる。

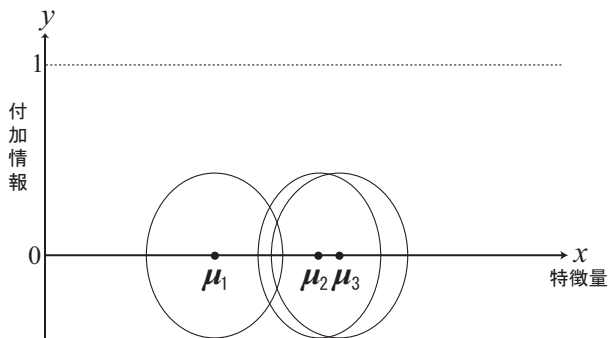


図 6 2次元の特徴量を持った3クラスの等高線

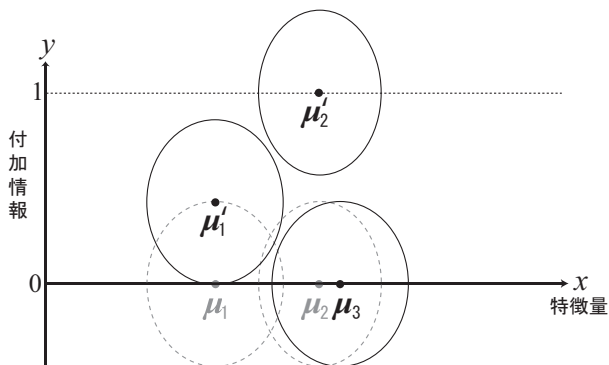


図 7 付加情報の平均を変更した等高線

しかし、図 7 のように付加情報を変更すると、円の交わる面積が減少するため、誤認識率も減少すると考えられる。このとき、どのようにして付加情報を決定するのが問題となる。本研究では付加情報の平均の決定は、誤認識率を推定することで、その誤認識率が最小となるように決定する。

## 5. 実験

付加情報の誤りと認識性能に関する 2 種類の実験を行った。(1) 従来手法 (文献 [1]) と提案手法を比較する実験と、(2) 付加情報の効果に関する実験である。

### 5.1 従来手法との比較実験

付加情報の誤りを考慮することによって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 から 1 の実数を付加情報として割り当てる。そのため、このままでは 2 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 から 1 の範囲に収まるように定めた。すなわち、 $n$  クラスに  $n$  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、付加情報の分散が一定量以下であれば、まずまずの認識性能が見込まれる。

それに対して、提案手法では特徴量の分布と付加情報の分布を考慮して、最も誤認識率が小さくなると考えられる付加情報

を割り当てる。本来であれば、分布を考慮して自動的に付加情報を割り当てる方法を示すべきであるが、初期的検討として、本稿では付加情報の値を手手で与えた。単純な方法であるが、あらかじめ予備実験を繰り返し、認識性能が高い付加情報の割り当てを採用した。つまり、本実験では特徴量と付加情報の分散を考慮して上手く割り当てた場合の認識性能の上限付近を示している。

実験のために、10 次元の正規分布に従う特徴量を持つ人工サンプルを 6 クラスに分作成した。各クラスは共通の共分散行列を持ち、平均のみが異なる。識別関数にマハラノビス距離を用いて各クラス 1,000,000 サンプルを認識した結果を表 1 に示す。表 1 は混同行列 (confusion matrix) であり、行方向のクラスが実際属しているクラスを表しており、列方向のクラスが認識されたクラスを表している。この認識結果を見ることにより、クラス 2 のサンプルをクラス 6 に多く誤認識していることや、クラス 6 のサンプルをクラス 2 に多く誤認識していることから、クラス 2 とクラス 6 に与える付加情報の値を遠くなるようにすれば誤認識率が減少することなどが予想される。前述した提案手法の付加情報の割り当ての際にはこの認識結果も考慮しながら、表 2 のように付加情報の値を 2 種類割り当てた。従来手法では、認識結果を考慮せずに、表 2 のようにクラス 1 からクラス 6 まで順に等間隔で付加情報の値を決定した

付加情報の誤りの分布は正規分布とした。すなわち、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分散とする 1 次元の正規分布である。したがって、認識に用いる人工サンプルは 10 次元の特徴量と 1 次元の付加情報からなる合計 11 次元の特徴ベクトルを持つ。

共分散行列と平均を算出するために各クラス 3,500 サンプルを用い、算出された共分散行列と平均を用いて各クラスさらに別の 1,000 サンプルを認識し、誤認識率を求めた。このような実験を付加情報の標準偏差の値を様々に変えて、同一の付加情報について 50 回ずつ行った。識別器としてマハラノビス距離を用いた。

以上の条件の下で行った実験結果が図 8 である。図より、従来手法と提案手法 1 を比べると、付加情報の標準偏差が小さい場合は提案手法 1 の方が従来手法より誤認識率が減少していることがわかる。これは表 1 からわかるように、付加情報に与える値を従来手法では順番に与えていったが、提案手法では表 1 を基に誤認識を起こしやすいと思われるクラスの付加情報の値を意図的に遠ざけたためであると考えられる。また、付加情報の標準偏差が 2 より大きくなると、従来手法と提案手法の誤認識率がほぼ同じになる。これは、付加情報の標準偏差が大きくなり過ぎて、付加情報の効果がほとんどなくなったためであると考えられる。この点については次節の実験も参照していただきたい。

提案手法 2 より、付加情報の値を等間隔に与えていない場合は、付加情報の標準偏差が 0.02 とかなり小さい値の場合から誤認識率が増加する。これは、クラス 5 の付加情報の値をクラス 3 に近づけてしまったために、付加情報の標準偏差が 0.02 という早い段階からクラス 3 のサンプルをクラス 5 に、クラス 5 の

表 1 各クラス 1000000 サンプルを用いて認識実験を行った結果 (単位は %)

	Class1	Class2	Class3	Class4	Class5	Class6
Class1	78.7	1.2	6.7	11.1	1.0	1.4
Class2	1.5	69.0	6.9	0.4	5.0	17.1
Class3	7.4	6.7	65.7	2.4	5.5	12.3
Class4	10.9	0.2	2.0	84.7	0.8	1.8
Class5	0.9	4.1	4.5	0.7	85.7	4.0
Class6	2.0	19.2	13.7	2.1	5.6	57.4

表 2 与えた付加情報の値

	従来手法	提案手法 1	提案手法 2
Class1	0.0	0.8	0.8
Class2	0.2	0.0	0.0
Class3	0.4	0.4	0.4
Class4	0.6	0.2	0.2
Class5	0.8	0.6	0.5
Class6	1.0	1.0	1.0

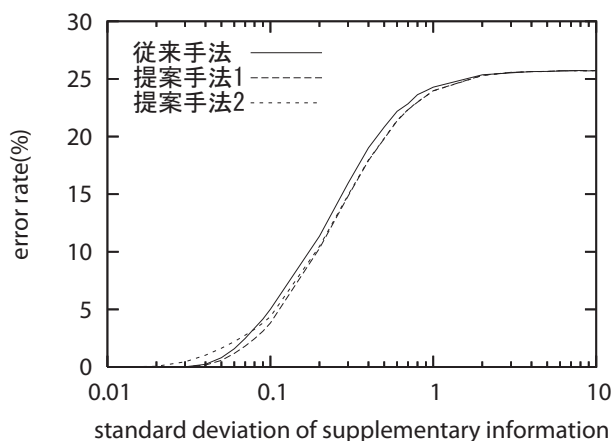


図 8 実験結果

サンプルをクラス 3 に誤認識してしまったからである。また、付加情報の標準偏差が 0.9, 1, 2 の場合は、提案手法 2 の方が提案手法 1 よりわずかであるが誤認識率が減少した。

これらの結果から、付加情報の標準偏差が誤認識率と大きく関わっていることがわかる。付加情報の標準偏差がある程度より大きくなると、誤認識率に及ぼす影響は少なくなり、付加情報の標準偏差がある程度より小さくなると、誤認識率を減少させる可能性があることがわかった。また、付加情報を与える範囲で等間隔に付加情報を与える方が付加情報の標準偏差が小さい場合については、誤認識率が減少する可能性があることがわかった。しかし、付加情報の標準偏差が大きい場合については、等間隔に付加情報を与える方が誤認識率が減少する可能性があることがわかった。

### 5.2 付加情報の効果に関する実験

提案手法は付加情報として実数値を割り当て、サンプルを観測して得られる特徴量と同等に扱って認識に用いる。このような事実から浮んでくる疑問は、サンプルを観測して得られる特徴量と付加情報の決定的な違いは何であろうか、ということである。その答えは、割り当てる値を恣意的に変更できる点であ

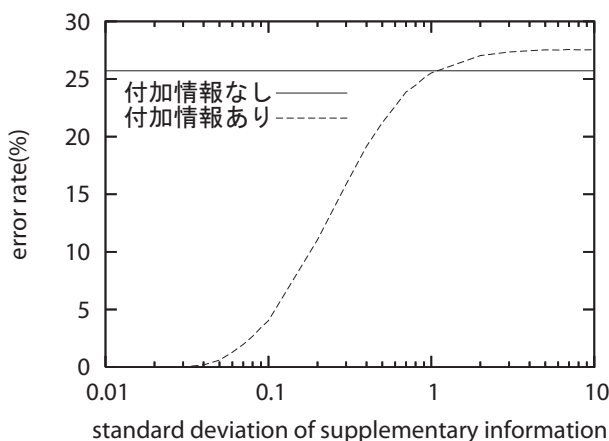


図 9 付加情報の影響を調べた実験結果

ると考える。このような付加情報の意味を考える上で、特徴量の 1 つを付加情報で置き換えた場合の認識実験を行う。

5.1 節で用いた 10 次元の正規分布に従う人工サンプルの特徴ベクトルを 6 クラス分用いて付加情報の効果に関する実験を行った。この実験では、10 次元の特徴ベクトルをそのまま用いて認識を行った「付加情報なし」と、10 次元の特徴ベクトルのうち 1 次元を付加情報の特徴で置き換えた「付加情報あり」の 2 つの認識実験を行った。付加情報の値は 5.1 節の提案手法 1 と同じ値を与えた。共分散行列と平均の算出、認識のために用いるサンプル数、実験回数、識別関数はすべて 5.1 節と同様である。これらの条件の下、実験を行った結果が図 9 である。

図から、付加情報の標準偏差が十分小さければ、付加情報が誤認識率の減少に大きく貢献していることがわかる。反面、本稿の実験では、付加情報の標準偏差が 1 を超えると付加情報が悪影響を及ぼし、誤認識率が従来手法より増加することがわかる。従って、付加情報は与えればよいというものではなく、ある程度復元した際の誤差が少ないものを用いなければ誤認識率を増加させてしまう恐れがあることがわかる。これは、本来恣意的に設定したはずの付加情報が恣意性を失った結果である。

## 6. まとめ

本稿では、付加情報を用いるパターン認識について述べた。これまで我々は、誤差を含まない付加情報を用いるという仮定の下で、付加情報を用いると誤認識率が減少する手法を提案してきた。しかし、現実世界において付加情報の復元を誤らないということは稀である。従って、本稿では誤差を含む付加情報を用いる場合に、サンプルにどのような付加情報を与えることで、誤認識率が減少するかについて検証し、実験を行うことによって誤認識率が減少する可能性があることを示した。この実験結果から、付加情報の標準偏差がある程度小さければ付加情報の与え方によって誤認識率を減少させる可能性があることがわかった。また、もう 1 つの実験から付加情報は与えればよいというものではなく、ある程度復元した際の誤差が少ないものを用いなければ誤認識率を増加させてしまう恐れがあるということがわかった。

今後は、付加情報の標準偏差がどの程度であれば誤認識率を

減少させる可能性があるのかについて検証し、人手で付加情報を与えるのではなく、最適な付加情報の与え方の理論を確立することが課題である。

また、本稿では誤りのある付加情報における付加情報の情報量と認識性能の関係について触れられなかった。誤りのない付加情報を最初に提案した文献 [6] 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た。誤りのない付加情報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。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,” 電子情報通信学会論文誌 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ー情報の埋込に関する検討,” 子情報通信学会論文誌 D, vol.J89-D, no.2, pp.344–352, Feb. 2006.
- [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,” 電子情報通信学会論文誌 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,” 電子情報通信学会論文誌 D, vol.J90-D, , Dec. 2007. 掲載予定.
- [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] 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?—,” 電子情報通信学会技術研究報告, vol.104, no.PRMU-742, pp.35–40, March 2005.