

情報付加による認識率 100%の実現

—人にも機械にも理解可能な情報伝達のために—

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

[†] 大阪府立大学大学院工学研究科 〒599-8531 大阪府堺市学園町 1-1

^{††} 九州大学大学院システム情報科学研究院 〒812-8581 福岡市東区箱崎 6-10-1

^{†††} 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

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

あらまし パターン認識の究極の目標は認識率 100%の達成であるが、誤認識が全く起こらない識別器の設計は容易ではない。そこで、本報告では認識誤りを補うために正解クラスに関する情報(付加情報)をパターンとは別に用意し、パターンと付加情報を識別器に同時に与えることで認識率 100%かつリジェクト率 0%を達成する方式を提案する。そして、提案する方式で達成可能な認識率やリジェクト率とそれに必要な情報量の関係を求める。実験の結果、印刷されたかな・漢字 3036 字種ではオープン実験で 4 ビット、クローズ実験で 1 ビットの情報の付加で認識率 100%かつリジェクト率 0%を達成できることが示された。また、認識を全く行わずに文字コードの情報をそのまま使用する場合に比べると、提案する方式で必要な情報量は約 1/12, QR コードの面積換算では約 1/9 である。従って、提案する方式は従来のパターン認識のみを用いる場合と認識を行わずコードのみの場合を折衷する新しい枠組みといえる。

キーワード 識別器, 情報量, 混同行列, 付加情報

Realization of 100% Recognition Rate with Supplementary Information

—For Seamless Man-Machine Communication—

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

[†] Graduate School of Engineering, Osaka Prefecture University

1-1 Gakuencho, Sakai-shi, Osaka, 599-8531 Japan

^{††} Faculty of Information Science and Electrical Engineering, Kyushu University

Hakozaki 6-10-1, Higashi-ku, Fukuoka-shi, 812-8581 Japan

^{†††} Graduate School of Engineering, Tohoku University

6-6-05 Aoba, Aramaki, Aoba-ku, Sendai-shi, 980-8579 Japan

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

Abstract The ultimate dream in pattern recognition is to achieve a 100% recognition rate. However, it is not so easy. In this report, for achieving a 100% recognition rate with a 0% rejection rate, we propose a new framework that the discriminator receives not only a pattern itself but also supplementary information that assists recognition. We clarify the relationships among a recognition rate, a rejection rate and quantity of supplementary information. For printed characters, experiments showed that 4 bits were required in the leave-one-out (L) method and 1 bit was in the resubstitution (R) method to achieve a 100% recognition rate with a 0% rejection rate. Furthermore, compared with identification codes such as a character code, the proposed approach compresses about 1/12 in quantity of supplementary information, and about 1/9 in size of area of QR code. Thus, the proposed framework combines advantages of the conventional approach of pattern recognition and an identification code.

Key words discriminator, quantity of information, confusion matrix, supplementary information

1. まえがき

1.1 人にも機械にも理解可能な情報伝達手段

人と機械が同じ情報を理解できることは重要である。例えば将来、人とロボットが共生していくことを考えてみよう。

今、人がロボットを呼び止め、ある文書を見ながら、その文書について議論しようとしている...

このとき両者が議論するためには同じ情報を共有していることが前提となる。情報の共有を実現する方法には、バーコードやRFIDを文書に貼り付けておき、テキスト等の文書情報を記憶させておく方法や、URIを指定して保存した文書情報をダウンロードさせる方法も考えられる。しかし、人が文書の一部を指差し、その部分に関する議論を始めるためには、いずれの方法もロボットのカメラから得られる文書画像と記録している文書情報を対応付ける作業が必要になり^(注1)、実現は容易ではない。

一方、パターン認識は人が自然に理解できる文字、音声等の情報を機械にも理解させようとする試みである。もし前述の例でロボットが文書画像解析や文字認識等を行うことができれば、指差された部分に関する議論を行うことが可能である。

ところが、人と機械の情報共有を可能にするパターン認識が常にうまくいくわけではない。パターン認識は条件さえ整えば高い認識性能を発揮できるが、常に条件が良いとは限らない。条件が悪い場合には認識誤りに対処する必要がある。そこで本稿では、想定し得る悪条件の下でパターン認識技術を利用するとき、十分高い精度の認識結果を得る方法について考える。

1.2 付加情報を利用する認識モデル

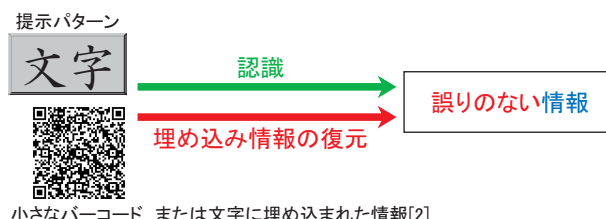
一般にパターン認識の識別器は、文字画像等のパターンから抽出された特徴量を基に、そのパターンが属する可能性が高いクラスを出力する(図1(a))。パターン認識では正しい認識結果にのみ価値が認められることが多いため、識別器の評価には認識率を用いるのが一般的である。その意味では認識率100%を実現できる識別器は理想的であるが、その実現は容易ではない。

それに対して、本報告では通常のパターン認識とは異なる図1(b)のような方式を検討する。これはパターンが属する(正解)クラスの情報(以後、付加情報または符号と呼ぶ。これは後述の記号の集合である。)をパターンと同時に認識系に入力し、パターンと付加情報から矛盾のない答えを導く方式である。本報告では付加情報は誤りなく与えられるとする。この方式は通常のパターン認識(図1(a))と、後述するコードをそのまま伝達する場合(図1(c))との折衷であると考えられる。この方式での識別器の善し悪しは、識別器を単独で用いた場合の認識率ではなく、認識系全体の性能によって評価される。すなわち、識別器に求められる性能が従来とは異なることが考えられ、従来高性能とされた識別器が付加情報を用いる方式で必ずしも高性能であるとは限らない。

(注1): ここでの対応付け作業とは、カメラから得られた文書画像が、ロボットが記録している文書のうちのどの文書で、何ページ目の何行目のどの部分であるか等を特定することを指す。



(a) 通常認識モデル



(b) 本報告で検討する認識モデル



(c) コードを伝達

図1 認識モデル

では、認識に有効な付加情報(符号)はどのようなものであろうか? 本報告で考える付加情報は、クラス数を N としたとき、 K 種類の記号 ($K \leq N$) を用意し、そのうち一つずつを各クラスに割り当てる。このときの付加情報の情報量はビット換算で $\log_2 K$ ビットである。もし N 種類の記号(ビット換算で $\log_2 N$ ビットの情報)を用いるならば、これは図1(c)のようにコードを伝達することに等しく、認識系は確実に正解クラスを言い当てることができる。従って、認識率100%が達成可能である。しかし、認識率100%を達成するためには常に N 種類の記号が必要なわけではなく、識別器の性能が良ければ必要な情報量は減少する。このように、認識率100%を達成するために必要な情報量は識別器の性能に依存するものと考えられる。

付加情報をパターンとは独立に識別器に与えることで認識性能の改善を図る研究には、最近ではアクティブ帳票[1]があり、複比を用いた付加情報の埋め込みも検討されている[2],[3]。しかし、与えられた識別器に対してどのような情報を付加すれば認識性能がどの程度向上するのか等、付加情報と認識性能の関係についてはほとんど解明されていない。そこで本報告では、識別器の誤り傾向を基に識別器に情報を付加した場合の理論を展開する。

まず、リジェクトを行わずに認識率100%を達成するために付加情報(符号)が満たすべき条件を求める。識別器の誤り傾向を表す指標として、本報告では真のクラス(正解)と認識結果の対応を表す Confusion Matrix(混同行列; 以下 CM; 定義については2.1を参照のこと)を用いる。CMを用いて認識誤りに対

		認識結果				
		A	B	C	D	E
真のクラス	A	120		80		
	B		160		20	20
	C	20		180		
	D		20		160	20
	E	40	20			140

図 2 Confusion Matrix の例 (数字のない要素は値が 0)

		認識結果				
		A	B	C	D	E
真のクラス	A	0.6		0.4		
	B		0.8		0.1	0.1
	C	0.1		0.9		
	D		0.1		0.8	0.1
	E	0.2	0.1			0.7

図 3 確率化した Confusion Matrix の例 (数字のない要素は値が 0)

処する研究としては、文書画像から OCR を用いてテキスト情報を抽出し、全文データベースを構築する際に CM を用いる方法 [4] がある。しかし、これは認識後に行う処理であり、本報告のように CM の情報を認識に反映するものではない。

次に、付加する情報量と認識性能の関係を求める。前述の通り、付加情報を用いるパターン認識 (図 1(b)) は、通常のパターン認識 (図 1(a)) とコードをそのまま伝達する場合 (図 1(c)) の折衷である。付加情報の情報量を 0 ビットから $\log_2 N$ ビットまで変化させると、達成可能な認識率は徐々に変化すると考えられる。そこで、(1) 付加する情報量と認識率の関係を識別器の誤り傾向から導く。また、誤認識が起こり得る状況が CM からわかるので、誤認識され得る認識対象をリジェクトし、誤認識率を 0% に保つことができる。そこで、付加情報の情報量を変化させ、(2) 誤認識率を 0% に保ったときに付加する情報量とリジェクト率の関係も導く。これらの情報は付加情報を利用する識別器を設計する際に非常に有用である。

実験では手書き文字と印刷文字の CM を用いる。印刷文字はフォントを限定すれば CM がほぼ正しく求まると考えられるため、この実験は現実に即したものである。一方、手書き文字は CM にない誤認識が起こる可能性が高く、また、文字パターンと同時に付加情報を与えることが困難なため、提案手法を手書き文字に用いることは現実的ではないが、参考のために掲載する。

最後に、本報告で提案するような付加情報を用いる文字認識と文字コードをそのまま送る場合に必要の情報量を比較する。QR コードを始め、バーコードは紙に情報を記録する一形態で、テキストの内容を全て表すことができる。しかし、テキストが多量になるとバーコードが非常に大きくなってしまふ。一方で、提案する方式では文字認識で不足する情報だけを補えばよく、バーコードを小さくできる。

2. Confusion Matrix とその確率的表現

2.1 Confusion Matrix

CM は識別器に入力されるパターンが属するクラス (真のクラス) と識別器が出力するクラス (認識結果) の対応を表す行列である (図 2 参照)。通常、 $\omega_1, \omega_2, \dots, \omega_c$ をクラス、 $n \times n$ 行列 C を CM とすると、 C の (i, j) 要素 c_{ij} はクラス ω_i に属するパターンを認識によってクラス ω_j に属すると判断した回数を表す。

2.2 Confusion Matrix の確率表現

クラス ω_i に属するパターンをクラス ω_j であると認識する確率 $P(\omega_j|\omega_i)$ を (i, j) 要素に持つ $n \times n$ 行列 W を定義する。 W は C の各行の要素の和が 1 になるように正規化して得られる。すなわち、

$$C_i = \sum_{j=1}^N c_{ij} \quad (1)$$

とおいたとき、 W の (i, j) 要素 w_{ij} は

$$w_{ij} = \frac{c_{ij}}{C_i} \quad (2)$$

で与えられる (図 3 参照)。

3. 損失計算のための準備

ここでは、確率化した CM である W を用いて損失を計算するための準備を行う。

3.1 行列 W の分割

提案手法では CM の各行に付加情報となる記号を割り当てる。 k 番目の記号を割り当てた行の集合を考えて、

$$\mathcal{H}_k = \{l | l = l_1, \dots, l_{|\mathcal{H}_k} \} \quad (3)$$

とおくことで W は $\mathcal{H}_1, \dots, \mathcal{H}_K$ に分割される。ここで $|\mathcal{H}_k|$ は同じ記号を割り当てられた行数である。例えば、図 5(a) の場合、 $\mathcal{H}_1 = \{1, 2\}$ 、 $\mathcal{H}_2 = \{3, 4\}$ 、 $\mathcal{H}_3 = \{5\}$ である。ただし、 \mathcal{H}_k の要素は連続している必要はない。

次に、 k 番目の記号が割り当てられた行の要素のうち j 列目の要素を

$$\mathcal{B}_{kj} = \{(l, j) | l = l_1, \dots, l_{|\mathcal{H}_k} \} \quad (4)$$

とおく。例えば、図 5(a) の場合、 $\mathcal{B}_{11} = \{(1, 1), (2, 1)\}$ である。

3.2 \mathcal{B}_{kj} 内の 0 でない要素数

最初に、値が 0 であれば 0、さもなければ 1 を返す関数

$$z(x) = \begin{cases} 0, & \text{for } x = 0 \\ 1, & \text{otherwise} \end{cases} \quad (5)$$

を定める。

\mathcal{B}_{kj} に含まれる要素のうち、 $w_{lj} \neq 0$ を満たす要素の数 q_{kj} を関数 $z(x)$ を用いて定義する。

$$q_{kj} = \sum_{l \in \mathcal{H}_k} z(w_{lj}) \quad (6)$$

4. 認識率 100%, リジェクト率 0%を実現する符号

図 1(b) のような認識系にパターンと付加情報が与えられるモデルを考えたとき, 認識率 100%, リジェクト率 0%を実現する付加情報が満たすべき条件を導き, グラフを用いた定式化を行う. 以後, パターンの生成器と認識系は符号のコードブックを共有しているものとし, 認識系は CM(行列 W) を持つとする. また, 簡単のために各クラスのサンプルは等確率で生成されたとする. すなわち, クラス数を N としたとき, 事前確率は $P(\omega_i) = \frac{1}{N}$ である.

4.1 符号が満たすべき条件

認識率 100%, リジェクト率 0%を実現するために付加情報が満たすべき条件を図 3 を例に考える. 図 3 によると認識結果が A になるのは真のクラスが A, C, E のいずれかのときである. これらは当然, 識別器では区別できない. しかし, もし識別器がそのまま A と出力してしまうと, 真のクラスが C や E だった場合には誤認識が生じてしまう. そのため, 真のクラスが A, C, E のいずれであったのかを区別する情報が必要となる. そこで A, C, E にそれぞれ別々の記号を割り当てると, 誤認識が生じない代わりに少なくとも 3 種類の記号が必要となる.

同様に, 認識結果が B になるのは真のクラスが B, D, E のいずれかのときであるので, B, D, E にも別々の記号を割り当てる. このように考えていくと, 図 5(a) のように, A と B, C と D, E に別々の記号を付加することでリジェクトを行うことなく認識率 100%が達成できる. つまり, 認識率 100%, リジェクト率 0%を実現するために付加情報が満たすべき条件は, 全ての k, j について B_{kj} 内に値が 0 でない要素が 2 つ以上存在しないことである.

4.2 情報量が最小の符号を発見する問題

グラフを用いて問題を定式化する. 各クラスに対応する N 個のノードからなる集合 V と, V の要素間を結ぶエッジの集合 E を考え, 無向グラフ $G = (V, E)$ を定義する.

同じ記号を割り当てても誤認識が起こらないクラスを, ノード間にエッジを張ることで表すことにし, 該当する全てのエッジを張る. ここでクラス i とクラス j に同じ符号を割り当てても誤認識が起こらないのは $w_{ij} = w_{ji} = 0$ のときである.

最後にグラフ G をできるだけ少ない完全グラフに分割する(図 4 参照). このときの完全グラフの数を N_{symmin} とおく. 完全グラフ毎に記号を割り当てるので, N_{symmin} 種類の記号があれば認識率 100%, リジェクト率 0%を実現できる. このとき, 必要な情報量は $\log_2 N_{\text{symmin}}$ で与えられる. 図 3 の場合, $\log_2 3 \sim 1.58$ となる.

ただし, この問題は NP 完全問題である PARTITION INTO CLIQUES [5] を最小化問題にしたものであり, NP 困難問題となる. 従って, 現在のところ多項式時間で実行可能なアルゴリズムは知られていない.

5. 付加する情報量と認識性能

本節では, 付加する情報量と認識性能の関係を求める. 具体

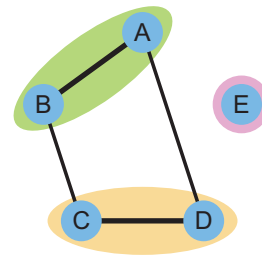


図 4 図 5(a) を表すグラフ

的には, (1) リジェクトを行わないときに付加する情報量と認識率の関係, (2) 誤認識率を 0%に保ったときに付加する情報量とリジェクト率の関係, について考える.

5.1 付加する情報量と認識率の関係 (リジェクトなし)

付加する情報量と認識率の関係を求める. リジェクトは行わない. なお, この問題は 4.2 の問題を含んでいる.

5.1.1 損失関数

誤認識が起こる確率を損失で表す. 4.1 で述べたように, 値が 0 でない要素が B_{kj} 内に 1 つ以下しか存在しない場合 ($q_{kj} \leq 1$) は誤認識が生じず, 2 つ以上の 0 でない要素が存在した場合 ($q_{kj} \geq 2$) は誤認識が生じる. このとき, 認識結果に成り得る真のクラスは q_{kj} 通りの可能性があるが, 認識結果として出力できるのはそのうち 1 つのクラスだけである. 取り得る損失が最も少ない方法は, 事後確率が最大になる要素を出力として採用することである. 例えば, 図 5(b) の場合, 記号が 2 で認識結果が A であれば, C と出力するよりも E と出力したほうが損失が少ないので, E と出力する.

以上より, 損失関数 L_1 を次式のように定式化し, 損失を最小にする問題を考える. ここで L_1 は真のクラス毎の誤認識率の和であり, 言い換えると, 全クラスの平均誤認識率にクラス数を乗じたものである.

$$L_1 = \sum_j \sum_k \left\{ \sum_{l \in \mathcal{H}_k} w_{lj} - \max_{l \in \mathcal{H}_k} w_{lj} \right\} \quad (7)$$

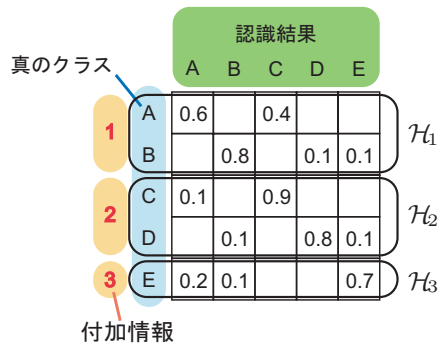
式 (7) の括弧内は第 1 項が B_{kj} 内の要素の和, 第 2 項が損失にならない要素 (出力するクラスに対応する要素) を表す. 損失関数を最小にする具体的なアルゴリズムについては 7.1 で述べる.

5.1.2 認識率

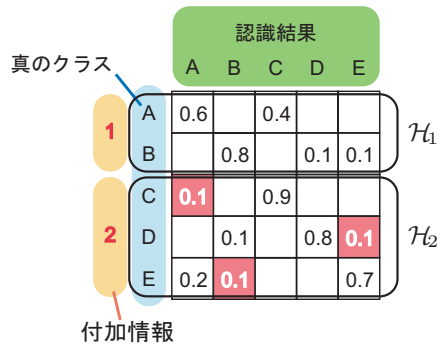
全クラスの平均認識率を求めるために, まず全クラスの平均誤認識率を求める. 損失関数 L_1 は文字毎の誤認識率の和であるので, L_1 を文字数で割れば平均誤認識率が得られる. 1 から平均誤認識率を引けば平均認識率であるので, 平均認識率は $R_{\text{recog}} = 1 - \frac{L_1}{N}$ で与えられる.

5.2 誤認識率を 0%に保ったときに付加する情報量とリジェクト率の関係

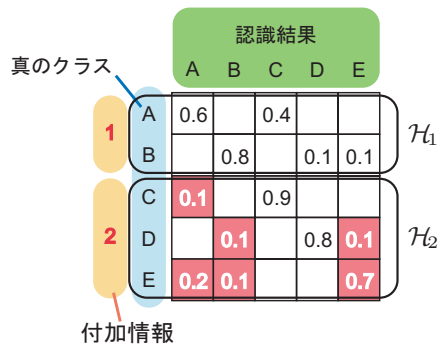
誤認識され得る認識対象をリジェクトすることで誤認識率を 0%に保つ場合を考え, 付加する情報量とリジェクト率の関係を求める. なお, この問題も 5.1.1 同様, 4.2 の問題を含んでいる.



(a) 認識率 100%, リジェクト率 0%を実現する符号
 (符号の種類: 3, 損失: 0, 認識率: 100%, リジェクト率: 0%)



(b) 誤認識を許容する符号 (リジェクトなし)
 (符号の種類: 2, 損失: 0.3, 認識率: 94%, リジェクト率: 0%)



(c) リジェクトによって誤認識を避ける符号
 (符号の種類: 2, 損失: 1.3, 認識率: 100%, リジェクト率: 26%)

図 5 符号の例. (b) と (c) は情報量が $\log_2 N_{\text{symmin}}$ より少ない場合である. 色が反転している要素が損失を表す.

5.2.1 損失関数

リジェクトする確率を損失で表す. 5.1.1 同様, 値が 0 でない要素が B_{kj} 内に 1 つ以下しか存在しない場合 ($q_{kj} \leq 1$) は誤認識が生じず, 2 つ以上の 0 でない要素が存在した場合 ($q_{kj} \geq 2$) は誤認識が生じる可能性がある. このとき, 認識結果に成り得る真のクラスは q_{kj} 通りの可能性があるが, 認識率 100%を達成するためには, 誤る可能性のある場合は全てリジェクトするしかない. 例えば, 図 5(c) の場合, 記号が 2 で認識

結果が A であれば, C と出力しても E と出力しても誤認識の可能性が残るため, この場合はいずれもリジェクトし, 全て損失となる.

以上より, 損失関数 L_2 を次式のように定式化し, 損失を最小にする問題を考える. ここで L_2 は真のクラス毎のリジェクト率の和であり, 言い換えると, 全クラスの平均リジェクト率にクラス数を乗じたものである.

$$L_2 = \sum_j \sum_k s_{kj} \quad (8)$$

ただし, ここで s_{kj} は

$$s_{kj} = \begin{cases} 0, & \text{for } q_{kj} \leq 1 \\ \sum_{l \in \mathcal{H}_k} w_{lj}, & \text{otherwise} \end{cases} \quad (9)$$

である. 損失関数を最小にする具体的なアルゴリズムについては 7.1 で述べる.

5.2.2 リジェクト率

全クラスの平均リジェクト率を求める. 損失関数 L_2 は文字毎のリジェクト率の和であるので, L_2 を文字数で割れば平均リジェクト率が得られる. 従って, 平均リジェクト率は $R_{\text{reject}} = \frac{L_2}{N}$ で与えられる.

6. 認識性能の評価基準

識別器の性能の評価基準である認識率, 認識後に残る不確かさ (認識が与える情報量), 認識率 100%を達成するために必要な情報量について概説する.

6.1 認識率

本研究では, 正しく認識されれば 0, 認識誤りには 1 という損失を考える 0-1 損失基準 [6] を用いたとき, 損失が発生する割合を誤認識率とし, それを 1 から引き算して得られる値を認識率としている. これは CM の対角成分の割合に等しい. 図 3 の場合について計算すると, 認識率は 76%である.

6.2 認識後に残っている不確かさ

X を真のクラス, Y を認識結果とする. 各クラスに属するパターンは等確率で生成されるので, パターンを認識する前に我々が持つ N クラス問題の不確かさは $H(X) = \log_2 N$ である. 一方, CM が与えられれば, 認識した後に残っている不確かさも条件付きエントロピーとして次式で得られる.

$$H(X|Y) = -\frac{1}{N} \sum_i \sum_j w_{ij} \log_2 w_{ij} \quad (10)$$

認識前の不確かさから認識後の不確かさを引き算して得られる相互情報量 $I(X; Y) = H(X) - H(X|Y)$ は, 認識によって打ち消される不確かさ, すなわち認識が与える情報量を表す.

図 3 の場合について計算すると, 認識後に残る不確かさ $H(X|Y)$ は 0.90 ビット, 認識が与える情報量は $\log_2 5 - 0.90 \sim 1.42$ ビットである.

6.3 認識率を 100%にするために必要な情報量

認識率を 100%にするために必要な情報量は 4.2 の $\log_2 N_{\text{symmin}}$ である. この値は最も誤認識され易いクラスの性

質によって決定する値である。

一般にパターン認識問題では認識が容易なクラスと困難なクラスが存在しているが、同様に誤認識され易いクラスとされ難いクラスも存在する。誤認識され易いクラスとは、CM を列毎に見たとき、0 でない要素が多い列に対応するクラスである。認識率 100% を達成するためには、識別器が最も誤認識され易いクラスを認識結果としたときでも誤認識が起きないだけの情報量を与える必要がある。

図 3 の場合について計算すると、認識率 100%、リジェクト率 0% の達成に必要な情報量は 4.2 で述べた通り、 $\log_2 3 \sim 1.58$ ビットである。

7. 実験

7.1 付加する情報量と認識率、リジェクト率の関係

実際の文字データを認識して得た 8 種類の CM を用い、認識後に残っている不確かさ、認識率 100% かつリジェクト率 0% に必要な記号の種類とその情報量を求めた。

8 種類の CM は 2 種類の文字データ、2 種類の識別器、さらにオープン実験とクローズ実験の組み合わせにより作成した。文字データは手書き文字と印刷文字を用いた。手書き文字に情報を埋め込むことはあまり現実的ではないが、情報量と認識性能の関係を調べるために用いた。手書き文字は ETL9B(3036 字種、200 セット) を 64×64 の大きさに非線形正規化 [7] した後、196 次元の方向線素特徴量 [8] を抽出して用いた。印刷文字は 25 種類のフォントから ETL9B に含まれる 3036 字種のひらがなと漢字を抜き出し、ETL9B と同様の処理で特徴量を作成した。ユークリッド距離と SQDF [9] を識別器として用いた。オープン実験は leave-one-out 法で行った。これらの実験条件とそのときの認識率、認識後に残る不確かさを表 1 に示す。なお、3036 クラス問題が持つ不確かさは $\log_2 3036 \sim 11.57$ ビットである。

8 種類の CM を対象として 5.1 の付加する情報量と認識率の関係と、5.2 の誤認識率を 0% に保ったときに付加する情報量とリジェクト率の関係を求めた。その際、欲張り法のアルゴリズムを用いた (Algorithm 1 参照)。結果を図 6、図 7 に示す。また、認識率 100%、リジェクト率 0% の達成に必要な記号の種類とそのときの情報量を表 1 に示す。欲張り法では近似解しか得られないが、認識率 100%、リジェクト率 0% の達成に必要な記号の種類は最適値と同じであった。図 6、図 7 は所望の認識率、リジェクト率を得るために必要な情報量を表している。これは提案手法のようにパターンとは別に付加情報を与える認識系を設計する上で有用である。

SQDF とユークリッド距離を比較すると、認識率と認識後に残っている不確かさでは SQDF が上回っているが、認識率 100%、リジェクト率 0% に必要な記号の種類では必ずしも SQDF は優位ではない。これは 6.3 で述べた誤認識され易いクラスの影響であり、極少数のそのようなクラスに支配された結果である。そのため、実用上は少数のクラスに支配されてしまう認識率 100%、リジェクト率 0% に必要な記号の種類 (情報量) よりも、多少のリジェクトを許容して認識率 100% を達成す

Algorithm 1 損失最小を目指す欲張り法のアルゴリズム

- 1: 全てのクラスに別々の記号を割り当てる。すなわち、 $k = 1, \dots, N$ について、 $\mathcal{H}_k = \{k\}$ とおく。
- 2: 損失を表す変数 $L(\cdot)$ を定義する。明らかに $L(N) = 0$ である。
- 3: for $K = N - 1$ to 1 do
- 4: 同じ記号を割り当てたときの損失が最小になる行の集合を 1 組選び、 $\mathcal{H}_s, \mathcal{H}_t$ とする。ただし、 $s \neq t$ かつ $\mathcal{H}_s, \mathcal{H}_t \neq \emptyset$ である。そして、 $\mathcal{H}_s, \mathcal{H}_t$ に同じ記号を割り当てる。すなわち、
$$\mathcal{H}_s \leftarrow \mathcal{H}_s \cup \mathcal{H}_t \quad (11)$$
$$\mathcal{H}_t = \emptyset \quad (12)$$
とする。このときの損失が $L(K)$ である。
- 5: end for

る状況を考えるほうが有益であると考えられる。一方で、付加情報の情報量を最小にする識別器、すなわち誤認識され易いクラスを生み出しにくい識別器の導出も求められる。これは従来行われていた認識率等の指標を最適にする識別器とは異なる性質を持つと考えられる。

7.2 文字コードを伝達する場合との比較

本節では書籍などの文書に書かれている情報をコンピュータが読み込む状況を想定し、文字コードをそのまま埋め込む場合 (図 1(c)) と、認識率 100% に必要な情報のみを埋め込む場合 (図 1(b)) を比較する。1 ページ当たり日本語のひらがなと漢字が 1000 文字含まれるとして計算した。

QR コードは埋め込むことができるデータ量によって 1 型から 40 型に分かれており、それぞれ大きさが異なる。文字コードをそのまま埋め込む場合、前節の実験で対象とした 3036 字種を表すためには 12 ビット必要なので、必要な情報量は

$$12 \text{ ビット} \times 1000 \text{ 文字} / 8 = 1500 \text{ バイト} \quad (13)$$

である。一方、文字認識を行う場合には、前節の実験結果から印刷文字のクローズ実験の場合には 1 文字当たり 1 ビットの付加情報で済むので、必要な情報量は

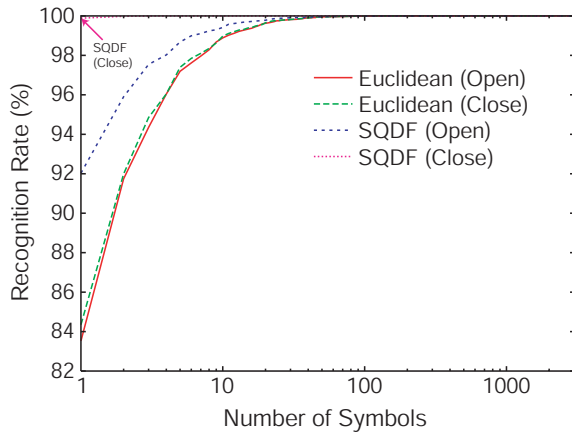
$$1 \text{ ビット} \times 1000 \text{ 文字} / 8 = 125 \text{ バイト} \quad (14)$$

である。

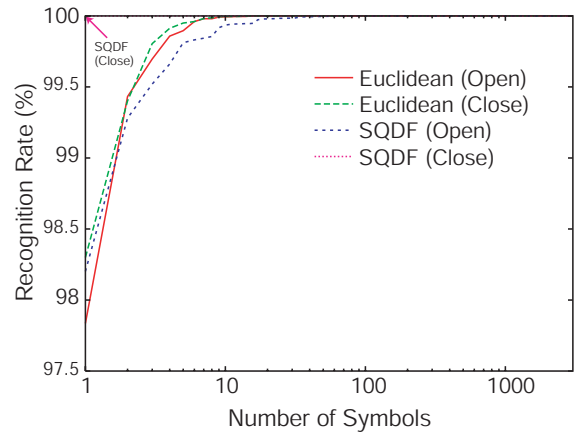
QR コードの基本仕様 (JIS X0510) によると、レベル M の誤り訂正能力を持つ QR コードの「8 ビットバイトモード」を使用した場合、125 バイトのデータを埋め込むことができる最小の型は 8 型であり、1500 バイトのデータを埋め込むことができる最小の型は 32 型である。そこで、8 型と 32 型の QR コードを Web 上の QR コード作成サイトで作成し、図 8 に示す。両者を比較すると、8 型は縦、横の大きさが 32 型の約 1/3、面積が約 1/9 である。このように、提案した認識と付加情報を用いる方式は文字コードのような情報のみを用いる場合に比べて情報量が大幅に削減されており、文字が既に書かれている文書などのメディアの場合、文字コードをそのまま埋め込むよりも提案手法のほうが効率的であることが示された。

表 1 実験条件と認識率, 情報量等

文字データ	識別器	オープン/ クローズ	認識率 (%)	認識後に残る 不確かさ (ビット)	認識率 100%, リジェクト 率 0%に必要な記号の種類	認識率 100%, リジェクト率 0%に必要な情報量 (ビット)
手書き文字	Euclidean	オープン	83.53	1.25	206	7.69
		クローズ	84.35	1.20	206	7.69
	SQDF	オープン	92.03	0.63	459	8.84
		クローズ	99.89	0.01	4	2.00
印刷文字	Euclidean	オープン	97.84	0.12	16	4.00
		クローズ	98.30	0.10	15	3.91
	SQDF	オープン	98.20	0.11	71	6.15
		クローズ	99.99	8.1×10^{-5}	2	1.00

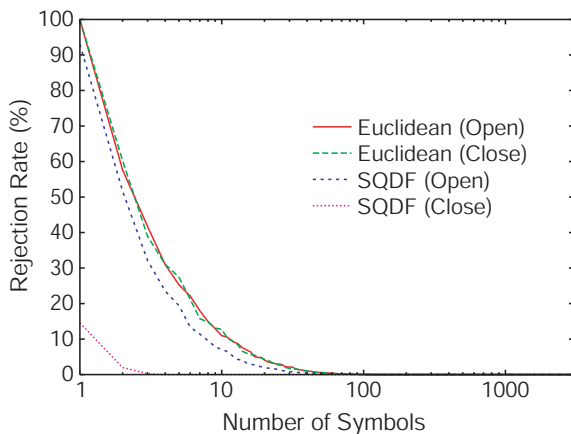


(a) 手書き文字

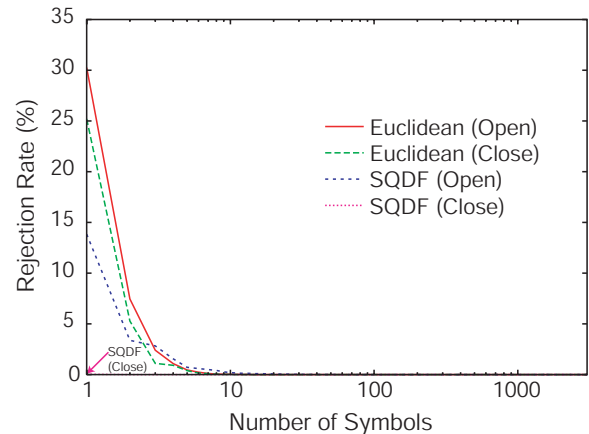


(b) 印刷文字

図 6 記号の種類と認識率の関係 (リジェクトなし)



(a) 手書き文字



(b) 印刷文字

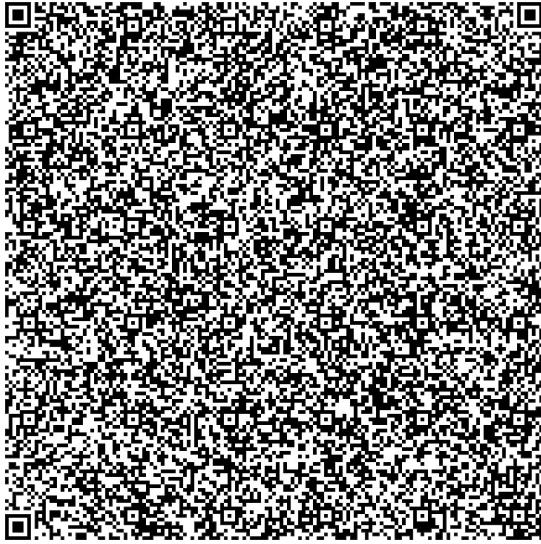
図 7 記号の種類とリジェクト率の関係 (誤認識率 0%のとき)

8. ま と め

本報告では, パターンが属するクラスの情報 (付加情報) をパターンと同時に識別器に入力し, パターンと付加情報から矛盾のない答えを導くことで認識率 100%, リジェクト率 0%を達成する方式について検討した. 文字認識結果を反映した CM を用いて実験したところ, 印刷文字ではオープン実験で 4 ビツ

ト, クローズ実験で 1 ビットの情報の付加で達成できることが示された.

さらに, 文字認識結果を反映した 8 種類の CM を用いて, (1) 付加する情報量と認識率の関係と (2) 誤認識率を 0%に保ったときに付加する情報量とリジェクト率の関係を観察した. この結果は所望の認識率, リジェクト率を得るためにはどの程度の情報量が必要なのかを知る上で有用である.



(a) 32 型 (文字コードに相当)



(b) 8 型 (認識率 100%に必要な付加情報に相当)

図 8 1 ページ (1000 文字) の日本語の文章を表すために必要な QR コード

また、付加情報の情報量は、従来識別器の性能評価によく用いられている認識率や認識後に残る不確かさとは別の性質を持つことを示し、付加情報の情報量を最小にする識別器の必要性を指摘した。今後、付加情報の情報量を最小にする識別器を導出することにより、認識率 100%、リジェクト率 0%の達成に必要な情報量は本報告で示した値よりも減少する可能性がある。

文 献

- [1] 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の設計と開発,” 信学論 (D-II), vol.J87-D-II, no.12, pp.2091–2103, Dec., 2004.
- [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付加情報の埋め込みに関する検討,” 画像の認識・理解シンポジウム (MIRU2005) 論文集, OS7A-29, July, 2005.
- [3] 黄瀬浩一, 大町真一郎, 内田誠一, 岩村雅一, “カメラを用いた文字認識・文書画像解析の現状と課題,” 信学技報 PRMU2004-246, Mar., 2005.
- [4] 太田学, 高須淳宏, 安達淳, “認識誤りを含む和文テキストにおける全文検索手法,” 情報処理学会論文誌, vol.39, no.3, pp.625–635, Mar., 1998.
- [5] M. R. Garey and D. S. Johnson, Computers and Intractability, W. H. Freeman and Company, New York, 1979.
- [6] 石井健一郎, 上田修功, 前田英作, 村瀬洋, わかりやすいパターン認識, オーム社, 東京, 1998.
- [7] 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.

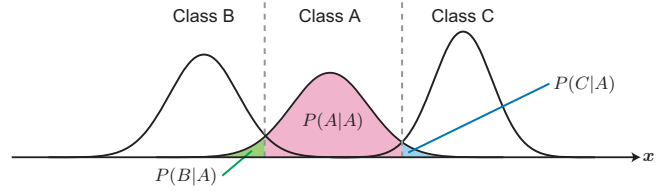


図 A.1 各クラスの確率分布, 識別境界, 確率化した Confusion Matrix の要素の関係 (着色部分の面積の大きさが確率の大きさを表す)

- [8] 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, Mar., 1991.
- [9] S. Omachi, F. Sun and H. Aso, “A new approximation method of the quadratic discriminant function,” Lecture Notes in Computer Science, vol.1876, pp.601–610, Sept., 2000.

付 録

1. Confusion Matrix の推定問題

本報告での議論は CM が絶対に信頼できることを前提としている。しかし、実際に用いることができる CM は有限のサンプルを用いて計算されたものであり、より多くのサンプルを用いれば、サンプルが少ない場合には起こらないとされた認識誤りが起こることも考えられる。そこで、本節では少数のサンプルで推定した CM を用いたときの提案手法への影響を考察するために CM の作成に用いるサンプル数と CM の推定精度の関係を導く。具体的には、CM の推定問題が各クラスの確率分布の推定問題に帰着することを示す。

図 A.1 に各クラスの確率分布と CM の要素の関係を示す。この図はクラス A の確率分布を識別境界で分割したとき、正しくクラス A と認識される確率 $P(A|A)$, B に誤認識される確率 $P(B|A)$, C に誤認識される確率 $P(C|A)$ をそれぞれ面積で表している。これを数式で表すと、

$$P(\omega_j|\omega_i) = \int_{\{x|P(\omega_j|x) = \max_l P(\omega_l|x)\}} P(\omega_i|x) dx \quad (A.1)$$

となる。ここで $P(\omega_i|x)$ は x が ω_i に属している確率 (事後確率) で、 $\{x|P(\omega_j|x) = \max_l P(\omega_l|x)\}$ は、クラス ω_j に認識される観測パターン x という意味である。従って、CM の推定精度は $P(x|\omega_i)$ の推定精度に依存する。従って、より正しい CM を推定するためにはより多くのサンプルを用いてより正しく確率分布を推定する必要があるといえる。