

ファジィ・エントロピーによる手書き類似漢字識別

大倉 充*・大橋一輝**・塩野 充***

*岡山理科大学大学院工学研究科システム科学専攻

**(株)住金制御エンジニアリング

***岡山理科大学工学部電子工学科

(1990年9月30日 受理)

1. まえがき

手書き文字認識へのファジィ理論 (*Fuzzy sets theory*)¹⁾の導入の試みは、国外ではかなり以前から行われてきた²⁾⁻⁶⁾。それらの研究では、文字パターンを構成するストローク或はセグメントの直線 (或は曲線) の程度及び位置、方向 (角度) 等の情報をあいまいなものと見なす方向で行われたものが多く見られる。我が国では最近になって、手書きの漢字、ひらがなを認識対象として、導入の試みが活発化している⁷⁾⁻¹²⁾。我が国におけるこれらの研究では、*On-line* 手書き文字¹³⁾を認識対象としている場合が多く (上述の国外の研究と比較した場合、文字パターンにおけるあいまいさのとらえ方には本質的な違いはないと思われる)、*Off-line* 手書き文字¹³⁾を扱っている研究報告は、わずかに文献(9)に見られる程度である。なお国外の研究ではあるが、文献(6)では、工業技術院電子技術総合研究所 (以下、電総研と記す) 作成の手書き教育漢字データベース *ETL-8 (B)*¹⁴⁾をデータとして用いている。これら2つの研究報告は、提案している認識手法自体全く異なっているにもかかわらず (文献(6)では構造解析的な手法¹³⁾を、文献(9)では重ね合せ的手法¹³⁾を土台とした認識手法の提案を行っている)、入力文字パターンの字種 (カテゴリー) 決定時にあいまいさの評価尺度 (*Measures of fuzziness*)¹⁵⁾を導入している点で一致している。あいまいさの評価尺度とは、ファジィ集合自体の持つあいまいさ (通常の集合はあいまいさを持たない) の程度を測る尺度のことで、文献(15)に詳しい。

本論文では、*Off-line* 手書き漢字を認識対象として、あいまいさの評価尺度の一つであるファジィ・エントロピー¹⁶⁾をカテゴリー決定時に導入した認識手法の提案を行っている。本手法は、特徴整合法¹³⁾の応用手法であり、文字パターンを種々の特徴量から構成されるファジィ集合と見なすものである。

本論文では、電総研の *JIS* 第一水準手書き漢字データベース *ETL-9*¹⁷⁾より類似漢字7組 (4カテゴリー/組) を選出して認識実験を行い、本手法の認識性能を調査した結果を示す。更に本手法に関する課題について言及する。

2. 提案する認識手法

本論文で提案する認識手法は次のようなものである。まず漢字パターンを種々の特徴量から構成される一つの集合（各特徴量を集合の要素と考える）と見なす。またコンピュータ内部に用意する標準パターンは、同一の要素（特徴量）から成る一つの集合とし各カテゴリーを表現すると考える。そして入力された漢字パターンから特徴量の抽出を行い、標準パターンとの間でマッチングを行う。この結果、漢字パターンを表わす集合の各要素について、標準パターンの数だけの類似性の度合を示す値が得られる。本論文では、これらの値を類似性の度合を示すメンバーシップ関数（*Membership function*）¹⁾の値（以下、グレード（*Grade*）と記す）と見なす。従って一つの入力パターンは、標準パターンの数だけのファジィ集合によって表現されることになる。そして各ファジィ集合の各カテゴリーへの所属の度合を定めるために、あいまいさの評価尺度により各ファジィ集合の持つあいまいさの程度（各カテゴリーへの所属に伴うあいまいさと考える）を求める。最終的に入力パターンは、得られたあいまいさが最小のファジィ集合を生成したカテゴリーに属すると決定される。

以下、本論文で用いたあいまいさの評価尺度であるファジィ・エントロピー、3種類の特徴量、グレードの決定法そして具体的な認識手順について説明を行う。

(2-1) ファジィ・エントロピー

ファジィ・エントロピーの説明の前に、まずファジィ集合の定義¹⁾について述べる。ある系 U （全体集合とする）があり、その要素 u_i ($i=1, 2, \dots, m$; m は要素数) とする。このとき U 上で定義されるファジィ集合 A とは、 U の各要素 u_i が A に属する度合を示す特性関数 μ_A （メンバーシップ関数）によって特性づけられた u_i の集まりである。 μ_A は、

$$\mu_A : U \rightarrow [0, 1] \quad (1)$$

と、 U の各要素 u_i に0から1までの実数値を与える。ここで1に近い値を持つ要素ほど A に属する度合が大きく、逆に0に近い値を持つ要素ほど A に属する度合が小さい。

ファジィ集合のメンバーシップ関数の形は多様に設定でき、各集合自体の持つあいまいさは異なると考えられる。そこでこのようなファジィ集合の備えたあいまいさの程度を測る尺度として、De Luca and Termini によりファジィ・エントロピーが提案された¹⁷⁾。まず全体集合 U におけるファジィ集合 A に $d(A)$ という関数であいまいさを測る尺度を与えると、De Luca and Termini によれば、少なくとも $d(A)$ は次に示す3つの条件を満たす必要がある。

- (1) A が通常の部分集合（クリस्प（*Crisp*）集合と記す）の場合、 $d(A)=0$ （あいまいさがない）である。
- (2) $d(A)$ が最大値を取るのは、 $\forall u \in U$ に対して $\mu_A(u)=0.5$ の場合に限る。

(3) AとBの2つのファジィ集合があり、AがBより“尖鋭”ならば、 $d(A) \leq d(B)$ とする。

ここで尖鋭とは、 $\mu_A(u) \leq 0.5$ のuに対し、 $\mu_A(u) \leq \mu_B(u)$ 、 $\mu_A(u) > 0.5$ のuに対し、 $\mu_A(u) > \mu_B(u)$ のことをいう。

条件(1)~(3)を満たす関数d(A)は種々考えられ、ファジィ・エントロピーはその一つで、次式で定義される。

$$d(A) = K \sum_{i=1}^m S(\mu_A(u_i)) \tag{2}$$

ここでKは正定数、Uは有限集合でその要素数をmとしている。また関数S(・)はシャノン関数(Shannon's function)であり、次式で表わされる。

$$S(x) = -x \log_2 x - (1-x) \log_2 (1-x) \tag{3}$$

図1にシャノン関数を示す。この図より $d(A) = d(\bar{A})$ (\bar{A} : Aの補集合で、 $\mu_{\bar{A}} = 1 - \mu_A$ で定義されるファジィ集合¹⁾)であることがわかる。

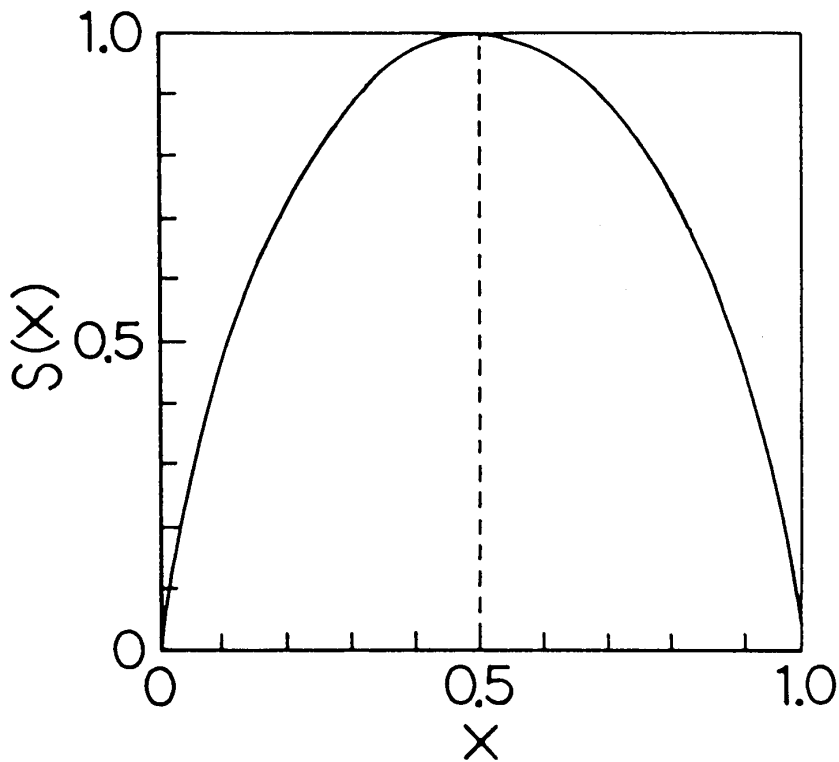


図1 シャノン関数

式(1)は、“要素 u_i が集合に属するか属さないか”という事象の平均情報量を全ての u_i について総和するという意味を持つ。

なお本論文ではファジィ・エントロピーとして、 $K = 1/m$ とした場合のもの、すなわち正規化ファジィ・エントロピー(Normalized fuzzy entropy)を用いた。

(2-2) 使用した手書き文字の特徴量

手書き漢字は筆記者の書き癖による手書き歪みを伴うため、単一の特徴量による識別はかなり困難なことと思われる。そのため、異なった図形情報を持つと考えられる複数の特徴量を併用することによって、認識率の向上を目指す試みが成されてきた¹⁸⁾。本論文では、提案されている代表的な3種類の大分類用の特徴量を選出し、漢字パターンより抽出した。以下、それら3種類の特徴量について説明を行う。

(2-2-1) 粗いメッシュパターン¹⁸⁾

粗いメッシュパターン（以下、M特徴 (M-feature) と記す）は、漢字パターンを構成しているストロークの概形情報（ストロークの大まかな存在形状と位置などの情報）に着目して考案されたものであり、次のようにして作成される。

まず、図2(a)の2値化された漢字パターンをより低次元のベクトル空間上で表現するために、同図(b)に示すように $n \times n$ （本論文では $n=8$ ）の粗いメッシュ領域に分割する。そして各メッシュ領域に含まれる文字部（画素値が1の部分）の全文字部に対する割合を計算することによって、 $n \times n$ 次元の特徴ベクトルを得る。

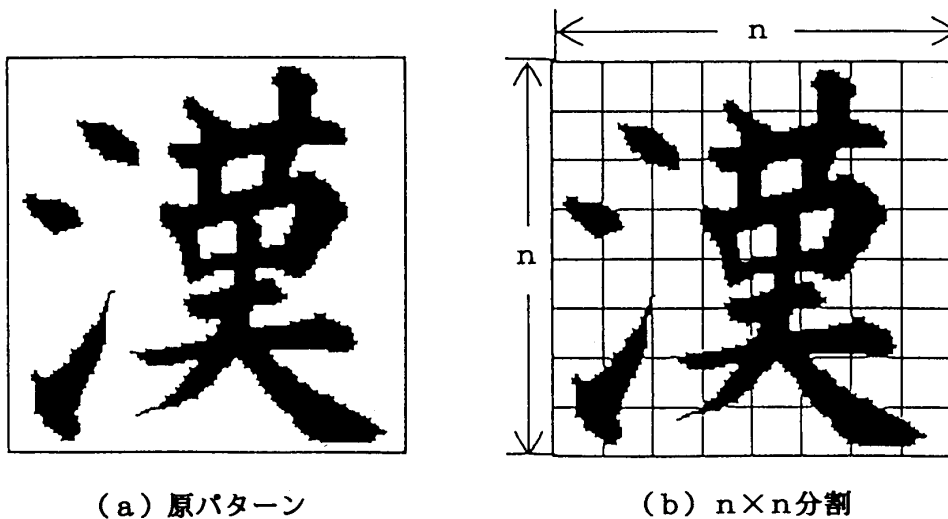


図2 粗いメッシュパターン

(2-2-2) 粗いペリフェラルパターン¹⁸⁾

粗いペリフェラルパターン（以下、P特徴 (P-feature) と記す）は、漢字パターンの持つ輪郭部分の特徴に着目し、構成ストロークの大まかな輪郭形状や位置などの情報を面積におきかえて抽出しようとするもので、次のようにして作成される。

まず、2値化された漢字パターンを図3に示すように n 分割（本論文では、 $n=8$ ）し、一方の外接枠から他方の外接枠に向かって、最初の文字部（画素値が1の部分）に出会うまでの文字部でない領域（画素値が0の部分）の面積を計数する。この操作を4つの外接枠について行い、 $n \times 4$ 次元の粗い1次のペリフェラルパターンとする。同様に、最初の文字部との出会いではそのまま計数を続け、2度目の文字部に出会うまでの面積を計数し、

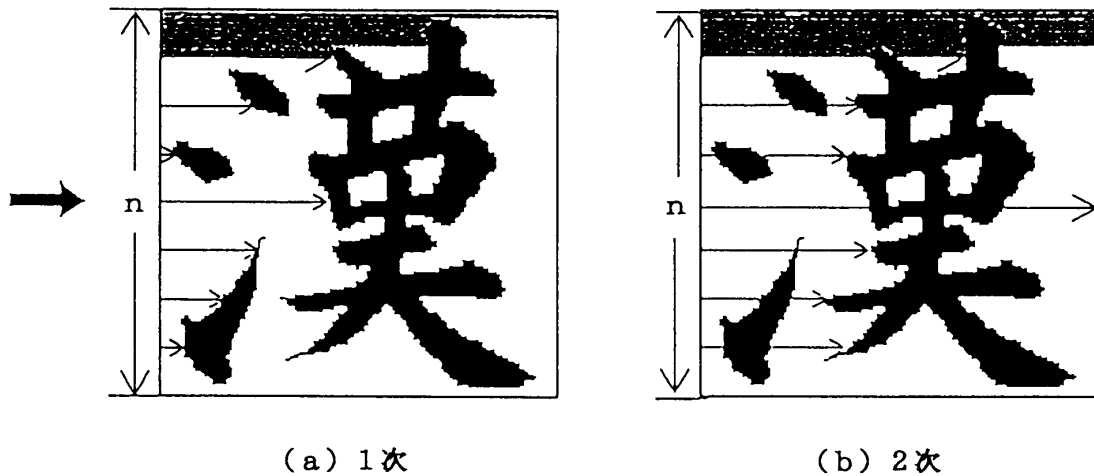


図3 粗いペリフェラルパターン (左側からの例)

これを $n \times 4$ 次元の2次のペリフェラルパターンとする。本論文では、1次と2次の2つのペリフェラルパターンを使用したため、特徴ベクトルの次元数は $n \times 4 \times 2$ となる。(2-2-3) ストローク密度関数 (Stroke density function)¹⁸⁾

ストローク密度とは、文字枠内の位置をパラメータとしてストロークの存在の度合を示すものである。ストローク密度関数 (以下、*SDF* 特徴 (*SDF-feature*) と記す) は、ストロークの本数に着目することによって漢字パターンの複雑さを反映する特徴量であり、次のようにして作成される。

図4に示すように、漢字パターンを水平及び垂直方向に走査し、ストロークを横切る回数を計数する (本論文では、用いた漢字パターンの画面次数が 64×64 であることより、 2×64 個の値を得る)。次に垂直及び水平方向に n 分割 (本論文では、 $n=32$) し、各領域に

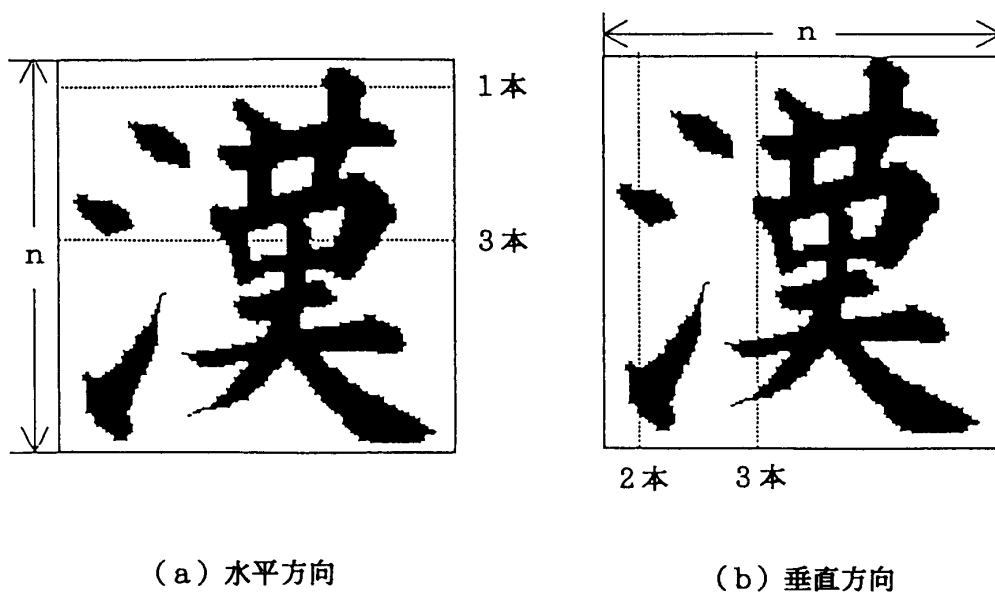


図4 ストローク密度関数

において得られた本数を平均する。従って得られる特徴ベクトルの次元数は、 $2 \times n$ となる。

(2-3) グレードの決定法

(2-3-1) 標準パターンの構成

2-2で述べた3種類の特徴量を漢字パターンより抽出した後、それぞれの特徴量から各カテゴリに対して3種類の標準パターンを作成する。これら3種類の標準パターンを一組の集合と考える。

標準パターンは、標準パターン作成用サンプル（以下、学習サンプル (*Training samples*) と記す）から特徴ベクトルを抽出した後、各ベクトルの対応する要素を学習サンプル分すべて加え合わせ、これを平均して得た。

(2-3-2) グレード

ファジィ集合はメンバーシップ関数により規定されるため、メンバーシップ関数の決定は非常に重要な問題である。本来、メンバーシップ関数の設定は人間の主観によって成されるべきものであるが、本論文では標準パターンと入力パターンから抽出された特徴ベクトルとの間で単純類似度（以下、類似度と記す）を計算し、この値をメンバーシップ関数の値、すなわちグレードとして用いた。類似度 γ は、二つのベクトル X 、 Y の類似性の尺度であり、次式で与えられる。

$$\gamma = \frac{(X, Y)}{\|X\| \cdot \|Y\|} \quad (0 \leq \gamma \leq 1) \quad (4)$$

ここで (X, Y) はベクトル X 、 Y の内積を表わし、 $\|\cdot\|$ はベクトルのノルムである。

(2-4) 認識手順

提案する認識手法の認識手順について述べる。まず入力パターンより2-2で述べた3種類の特徴量を抽出する。M特徴では、入力パターンを 8×8 の粗いメッシュパターンに分割、P特徴（1次+2次）では、8つの領域に分割、SDF特徴では、32の領域に分割することにより、それぞれ64次元の特徴ベクトルを得る。次に標準パターンと特徴ベクトルとの間で類似度を求める。その結果、各標準パターン（カテゴリ）に対して3つの類似度値が得られる。本手法では、これら3つの類似度値をファジィ集合の各要素に割り当てられたグレードと見なす。そして各ファジィ集合の各カテゴリへの所属の度合を定めるために、ファジィ・エントロピーにより各ファジィ集合の持つあいまいさの程度（各カテゴリへの所属に伴うあいまいさ）を求める。最終的に入力パターンは、得られたあいまいさが最小のファジィ集合を生成したカテゴリに属すると決定される。

ここで重要な点はしきい値の設定である。ファジィ・エントロピーは、ファジィ集合 A とその補集合 \bar{A} に対して等しい値を与えるため、本論文ではしきい値 T_v (≥ 0.5)を定め、それ以上の値を示すグレードをもとにしてあいまいさを求めた。そのため、入力パターンに対して割り当てられたすべてのグレードがしきい値 T_v を下回った場合には、認識不能として棄却処理 (*Rejection*) を行った。

3. 認識実験

(3-1) 実験に使用したデータ

実験に使用したデータは、電総研の JIS 第一水準手書き漢字データベース ETL-9¹⁷⁾ である。このデータベースは、2 値化及び、位置と大きさの正規化済みの文字パターンから構成されているため、本論文では前処理を行わず原パターンをそのまま使用した。全 200 データセットから成り、1 データセットにひらがなを含め JIS 第一水準漢字 3036 カテゴリーが収納されている。実験に際し、1 カテゴリー-200 サンプルがデータセット番号の順番に並ぶように編集し直して用いた。なお、画面次数は 63×64 である（特徴量抽出時において、1 ライン（64 個の 0）をデータの末尾に追加し、画面次数 64×64 とした）。認識実験においては、200 サンプルの内前半の 100 サンプルを学習サンプルとして、後半の 100 サンプルを未知サンプル (*Unknown samples*) として用いた。

表 1 認識実験に用いた類似カテゴリーの組

組番号	類似カテゴリー
①	論 輪 輪 論
②	詰 結 紹 詔
③	渴 掲 湯 揚
④	丑 五 互 互
⑤	閏 開 閑 関
⑥	狙 祖 租 粗
⑦	鏡 鎖 鐘 鎮

類似カテゴリーとして、表 1 に示す 7 組（4 カテゴリー/組）を選出した¹⁹⁾。組番号①～③は、偏と傍のそれぞれ 2 種類の組み合わせで各カテゴリーが表現できるもの²⁰⁾を、④は、形状自体が似ていると筆者らが判断したものを、⑤～⑦は、構え、傍、偏がそれぞれ同じものを選んでいく。

実験には本学情報処理センターの FACOM/M380 を用い、言語は FORTRAN77 を使用した。

(3-2) 実験結果

(3-2-1) しきい値固定 ($T_v=0.5$) の場合の識別率

提案する認識手法による結果を示す前に、本論文で用いた 3 種類の特徴量を個別に用いた場合の結果を表 2, 3 に示す。なおこの場合、入力パターンの所属カテゴリーは、類似度値の最大値を示した標準パターンの所属するカテゴリーとした。また同表には、単純類似度法 (*Simple similarity method*) による結果も併せて示してある¹⁹⁾。単純類似度法では、学習サンプルを全て加え合わせることによって濃度レベル (*Gray level*) を持った

標準パターンを作成した。

表2 各特徴量による学習サンプルに対する識別率 (%)

組番号	M 特徴	P 特徴	SDF特徴	単純類似度法
①	77.00	70.75	78.00	79.25
②	75.50	83.00	83.50	80.50
③	84.00	76.75	73.00	85.75
④	70.75	78.00	88.00	70.00
⑤	62.50	60.75	71.00	69.25
⑥	63.75	69.25	65.75	73.75
⑦	74.25	71.75	73.00	77.50
平均	72.5	72.9	76.0	76.6

表3 各特徴量による未知サンプルに対する識別率 (%)

組番号	M 特徴	P 特徴	SDF特徴	単純類似度法
①	72.50	67.25	70.00	67.00
②	73.25	75.00	79.75	75.25
③	78.75	71.25	69.50	77.75
④	68.25	78.25	88.75	66.50
⑤	60.50	54.50	63.00	58.50
⑥	57.75	68.00	66.50	59.00
⑦	69.75	68.75	69.50	72.75
平均	68.7	69.0	72.4	68.1

学習サンプルに対しては単純類似度法を用いた場合に、組番号④を除いて比較的良好な結果が得られている。組番号④については、SDF 特徴を用いた場合に最も良好な結果が得られており、他の特徴量を用いた場合の結果との差が大きい。未知サンプルに対しては、類似カテゴリーの組によって若干のばらつきはあるものの、SDF 特徴を用いた場合に比較的良好な結果が得られている。7組の平均値に注目すると、学習サンプルでは最も良好であった単純類似度法による結果が、未知サンプルでは最も劣っていることがわかる。このことは、単なる重ね合せ的手法では、本論文で用いたような非常に類似したカテゴリーの識別が不可能であることを示唆している。

表4 提案する認識手法による識別率 (%)

組番号	特徴量併用		本手法	
	学習	未知	学習	未知
①	83.25	76.25	83.25	76.00
②	86.50	83.25	88.00	85.25
③	87.00	84.00	86.75	82.75
④	79.00	78.50	81.75	78.50
⑤	72.00	68.00	73.75	69.25
⑥	72.50	66.50	73.00	69.00
⑦	79.25	75.75	80.00	76.25
平均	79.9	76.0	80.9	76.7

表4に本論文で提案する認識手法(本手法と記す)による識別実験結果を示す。また比較のために、3種類の特徴量を併用した場合の結果も示している。この場合3種類の特徴量により得られた類似度値をそれぞれ $\gamma_M, \gamma_P, \gamma_S$ とし、3次元のベクトル $V (= (\gamma_M, \gamma_P, \gamma_S))$ を考え、 V のノルムの最大値を示すカテゴリーを入力パターンの決定カテゴリーとした。特徴量を併用すると、特徴量を単独で用いた場合に比べて学習、未知両サンプル共に識別率の上昇の傾向が見られる。ただし組番号④に関しては、やはりSDF特徴を単独で用いた場合の方が良好な結果が得られている。本手法による結果を特徴量併用による結果と比較すると、平均すると学習、未知両サンプル共に若干ではあるが識別率上昇の傾向が見られる。学習サンプルでは、7組中5組、未知サンプルでは、7組中4組において識別率の向上が得られた(両サンプル共に1組は変化なし)。

本論文では、わずかに3種類の特徴量(ファジィ集合の要素数が3)しか用いていないため断定はできないが、ファジィ・エントロピーを用いることにより、わずかではあるが識別率の改善が成される傾向があると思われる。この傾向を明確にするために、本論文で用いた特徴量とは別の他の構造情報を担った特徴量を追加しての識別実験を行う必要があると考えている。また本論文ではグレードとして便宜的に類似度値を用いた。しかしファジィ理論の適用においては、メンバーシップ関数の設定が最も重要なことであるため、この点に関しても検討する必要がある。

(3-2-2) しきい値の変化に伴う平均識別率の変化

図5にしきい値 T_V を変化させた場合の平均識別率 R_A を示す。 R_A は、各組の平均識別率を7組で平均した値を表わす。図中、●が学習サンプルに対する値を、○が未知サンプルに対する値を示している。しきい値の増大に伴って、学習、未知両サンプルに対して同様の R_A の減少の傾向が見られ、しきい値を固定して見た場合、 $T_V=0.5$ の場合に、学習、

未知両サンプルに対して最も良好な R_A が得られている。 $T_v=0.9$ から0.95となった場合に R_A の上昇が生じている。これは、ある特徴量に関して誤識別を起こしている場合の類似度値（グレード）がかなり大きく、 $T_v=0.95$ でファジィ・エントロピーを求めるときに、そのグレードを除くことになるためである（残りの特徴量で得られたグレードを用いてファジィ・エントロピーを求めることになる）。

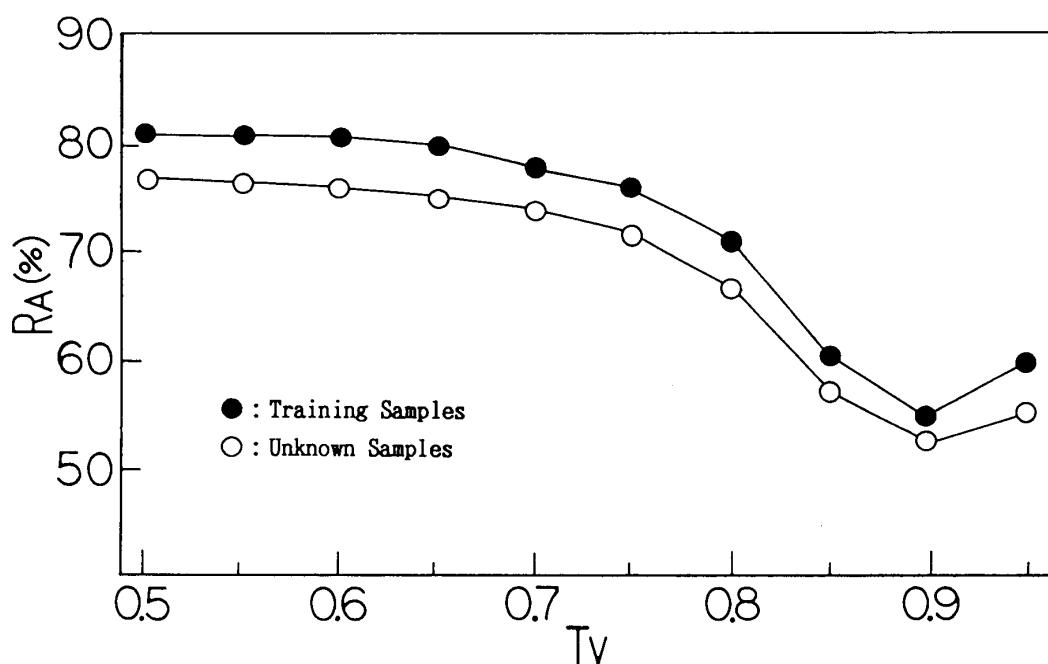


図5 しきい値と平均識別率との関係

4. むすび

本論文では、あいまいさの評価尺度の一つであるファジィ・エントロピーによりカテゴリーの決定を行う手書き漢字認識手法の提案を行った。本手法は、入力パターンとカテゴリーを代表する標準パターンとの間でグレードを定め、その値をもとに入力パターンの各カテゴリーへの所属に伴うあいまいさを求め、あいまいさが最小となるカテゴリーに決定するというものである。本手法を文字認識手法として位置付けるならば、ファジィ理論の考えを基礎とした特徴整合法ということになる。

本文中で示したように、従来の特徴整合法に比べて提案した認識手法によって、非常にわずかではあるが、学習、未知両サンプルに対して識別率の向上が得られた。またファジィ・エントロピーを求めるにあたっての最適なしきい値であるが、本論文で用いた特徴量に対しては、しきい値0.5で最も良好な識別率が得られた。

今後の課題としては、本文中で述べた次の2点が考えられる。

- (1) 新たな文字パターンの構造情報を担った特徴量の追加。
- (2) メンバーシップ関数（入力パターンと標準パターン間のグレード）の検討。

謝 辞

認識実験において使用させていただいた電総研手書き漢字データベース ETL-9 を作成された関係各位に感謝いたします。また本研究への有益なご助言を頂いた、本学藤田志郎教授、橋本礼治教授に深謝いたします。

参考文献

- 1) 水本雅晴：“ファジィ理論とその応用”，サイエンス社（1988）。
- 2) P. Siy and C.S. Chen: “Fuzzy logic for handwritten numeral character recognition”, IEEE Trans. Syst., Man, Cybern., Vol. SMC-4, pp. 570-575 (1974-11).
- 3) W.J. M. Kickert and H. Koppelaar: “Application of fuzzy set theory to syntactic pattern recognition of handwritten capitals”, IEEE Trans. Syst., Man, Cybern., Vol. SMC-6, pp. 148-151 (1976-02).
- 4) G. C. Oden: “A fuzzy logical model of letter identification”, J. Exp. Psychol. Human Percept. Perform., Vol. 5, 2, pp. 336-352 (1979).
- 5) P. Biswas and A. K. Majumdar: “A multistage fuzzy classifier for recognition of handprinted characters”, IEEE Trans. Syst., Man, Cybern., Vol. SMC-11, pp. 834-838 (1981-12).
- 6) F-H Cheng, W-H Hsu and C-A Chen: “Fuzzy approach to solve the recognition problem of handwritten chinese characters”, Pattern Recognition, Vol. 22, No.2, pp. 133-141 (1989).
- 7) 太田俊一, 秋元 学, 町田富夫, 向殿政男: “ファジィ理論を応用したオンライン手書き漢字認識”, 情処学第36回全大, 6 V-5, pp. 1791-1792 (1988-03).
- 8) 大森健児: “ファジー推論による実時間手書き漢字認識”, 信学論 (D-II), J72-D-II, 3, pp. 369-379 (1989-03).
- 9) 大倉 充, 島田英之, 塩野 充: “「あいまいさ」の評価尺度を用いた手書き類似漢字の識別実験”, 信学技報, PRU 89-13, pp. 95-102 (1989-05).
- 10) 天野光康, 菊池浩明, 向殿政男: “ファジィ Prolog 'PROFIL' によるオンライン手書き文字認識”, 情処学第39回全大, 5 E-3, pp. 540-541 (1989-10).
- 11) 櫻庭祐一, 山口博史, 馬籠良英: “ファジー集合論を応用したオンライン手書き文字認識”, 信学論 (D-II), J72-D-II, 12, pp. 2032-2040 (1989-12).
- 12) 大森健児: “続け字と崩し字に対応したヒューリスティックなストローク合わせ法によるオンライン手書き漢字認識”, 情処学論, Vol. 31, No.5, pp. 710-720 (1990-05).
- 13) 橋本新一郎(編著): “文字認識概論”, 電気通信協会(昭57).
- 14) 斎藤泰一, 山田博三, 山本和彦, 森 俊二: “手書文字データベースの解析(V)”, 電子技術総合研究所彙報, 45, Nos. 1, 2, pp. 49-77 (1981-02).
- 15) 本多中二: “ファジィ・エントロピーとはなにか”, 数理科学, No.294, pp. 54-61 (1987-12).
- 16) A. De Luca and S. Termini: “A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory”, Information and Control, 20, pp. 301-312 (1972).
- 17) 斎藤泰一, 山田博三, 山本和彦: “手書文字データベースの解析(VIII)”, 電子技術総合研究所彙報, 49, 7, pp. 487-525 (1985).
- 18) 萩田紀博, 梅田三千雄, 増田 功: “三つの概形特徴を用いた手書き漢字の分類”, 信学論 (D), J63-D, 12, pp. 1096-1102 (1980-12).
- 19) 大倉 充, 塩野 充: “多重辞書類似度法における類似文字識別能力について”, 情処学第37回

- 全大, 5 W-2, p. 1633 (1988-09).
- 20) 森 吉弘, 横澤一彦, 梅田三千雄: “PDP モデルによる手書き漢字認識”, 信学技報, MBE 87-156, pp. 407-414 (1988).

On Discrimination of Similar Handprinted KANJI Characters by Fuzzy Entropy

Mitsuru OHKURA*, Kazuki OH-HASHI** and Mitsuru SHIONO***

**Graduate School of Engineering, Okayama University of Science,
1-1 Ridaicho, Okayama, 700 Japan*

***SMI Control Engineering Co., Ltd.*

****Faculty of Engineering, Okayama University of Science*

A method based on the concept of fuzzy sets for handprinted KANJI character recognition is proposed in this paper. KANJI characters can be viewed as a collection of several feature patterns. Since the feature patterns under consideration here are fuzzy in nature, the concept of fuzzy sets is utilized. Three kinds of feature patterns are used and the pattern similarity is used as a grade of membership function between an input pattern and a templet. Category determination is made by the function of fuzzy entropy which is one of the measures of fuzziness of a fuzzy set. 7 sets (4 categories/set) of similar KANJI characters in *ETL-9* (200 variations/category) are used as the test patterns.