分類スコアの仮確率を用いた“あいまい度の尺度”への適用

1C−2

佟 国祥　　星 仰
筑波大学　　茨城大学

1 はじめに

多値画像データの分類やパターン認識などの認識過程においては、既知パターンを標準パターンとして、認識しようとするパターンの特徴はいろいろな分類手法または認識手法によって抽出される。しかし、この多値画像データの分類の認識過程で種々な原因によって、多値画像の画素が誤分類あるいは誤認識されることがよくある。その原因の一つは多値画像の画素の場合に１画素が一般的に純粋なパターンからなるとは限らない。つまり、１画素の濃淡レベル値が複数のパターンの平均値または総合値である。このことは１画素データが１物体１現象という概念で取り扱える場合と複合体が１画素のデータを形成していると考えなければならない場合がある（ミクセル問題）。前者の画素データの分類に対して、トレーニングデータで分類を評価するとき、分類スコアが算出される。この分類スコアはＣｒｉｓｐデータ（正の整数値をとる）のため、分類スコア全体を１つの尺度で評価するのに、あいまい度の尺度などは適した尺度ともいえよう。しかし、後者の場合分類スコア中の誤分類に関してはＣｒｉｓｐデータとして取り扱うのでは不十分であり、帰属クラスの確率が複数の分類項目に分散させるのが妥当と思われる。

ところで、ファジイ理論はＬ．Ａ．Ｚａｄｅｈによって、１９６５年に提案され、その後、多くの分野への応用が研究されてきた。多値画像データの分類においても、ファジイ理論に基づくファジイセット分類法が使われてきている。ファジイ理論の適用によって、従来の一画素、一つのクラスに属するという概念を変え、メンバーシップ関数によって、より客観的に画素内の性状を反映できるようになった。本論文ではCrisp Data（分類前のクラスのメンバーシップ関数値）とFuzzy Data（分類後メンバーシップ関数値）間の関数関係を定義し、この関係関数に基づいて、仮頻度（pseudo-frequency）という概念を提案する。この仮頻度が算出されると、自然にこの仮頻度から分類結果の仮確率（pseudo-probability）を求めることができ、定義することもできる。そして、いくつかの分類結果の数学モデルを利用して、ファジイ理論に基づく分類結果の評価の有効性を検証する。

2．あいまい度の尺度の定義

前述のように、分類結果の評価は処理過程が複雑で、かつ分類項目であるクラスの数が多いため、容易ではない。そこで、分類結果の評価についての理論が必要であり、一つの評価尺度で総合評価できることが望まれよう。

あいまい度の尺度Ｔ（Ｇ，$\bar{G}$）はＳｈａｎｎｏｎの情報理論に基づくエントロピーの概念を導入している。これらの従来の評価法の概要を以下で述べる。

表１のＧは分類前のクラスの総称を意味し、g_1，g_2，・・・，g_cは個々のクラスを示す。$\bar{G}$は分類後のクラスの総称を意味し、$\bar{g}_1$，$\bar{g}_2$，・・・，$\bar{g}_c$はその個々のクラスを示す。ｃはクラス数である。また、m_{ij}はテスト地区の各クラスの分類結果を画素数で書き表わしている。表１の対角要素m_{ij}(i=j)は分類前と分類後同じクラスに割り当てられた画素数を表す。つまり、m_{ij}(i=j)は自己のクラスに割り当てられたことを意味する。対角項以外のm_{ij}(i≠j)はすべて他のクラスに分類されたことを意味する。

表１　　パターン分類スコアの例

G＼$\bar{G}$	$\bar{g}_1$	$\bar{g}_2$	・・・	$\bar{g}_c$
g_1	m_{11}	m_{12}	・・・	m_{1c}
g_2	m_{21}	m_{22}	・・・	m_{2c}
・	・			・
・	・			・
・	・			・
g_c	m_{c1}	m_{c2}	・・・	m_{cc}

エントロピーの概念をあいまい度の尺度に導入された理由は、分類前の多値画像データの画素のクラスと分類後の画素のクラスとの処理過程が情報理論における通信文の送信と受信の過程に類似しているからである。そこで、エントロピーの概念を適用した分類結果の評価尺度（あいまい度の尺度）の理論を以下に示す。

ここでも表１の分類スコアから総合的な分類評価尺度を求めることを考える。表１の頻度行列の要素m_{ij}は、次の式（１）を用いることによって確率p_{ij}に変換できる。

$$p_{ij}=m_{ij}\Big/\sum_{i}^{n}\sum_{j}^{m}m_{ij} \qquad (1)$$

上式に示した行おやび列の和をＰ（g_k），Ｐ（$\bar{g}_k$）とすると、式（２）のように書き表すことができる。

$$P(g_k)=\sum_{j=1}^{n}p_{kj} \qquad P(\bar{g}_k)=\sum_{i=1}^{n}p_{ik} \qquad (2)$$

したがって、分類前のクラスＧのエントロピーＨ（Ｇ）と分類後のクラス$\bar{G}$のエントロピーＨ（$\bar{G}$）を求めると、それぞれ式（３）と（４）となる。

$$H(G)=-\sum_{k=1}^{n}P(g_k)\ \log P(g_k) \qquad (3)$$

$$H(\bar{G})=-\sum_{k=1}^{n}P(\bar{g}_k)\ \log P(\bar{g}_k) \qquad (4)$$

同様にＧと$\bar{G}$の結合エントロピーはＨ（Ｇ，$\bar{G}$）とすると、結合エントロピーＨ（Ｇ，$\bar{G}$）は次式のように求めることができる。

$$H(G,\bar{G})=-\sum_{i}^{c}\sum_{j}^{c}p_{ij}\ \log p_{ij} \qquad (5)$$

Application to Equivocation Quantification Using Pseudo-probability of Classification Score

Guoxiang Tong　Takashi Hoshi
University of Tsukuba　Ibaraki University

GとḠの相互情報量は次式となる。

$$I(G,\bar{G})=H(G)+H(\bar{G})-H(G,\bar{G}) \quad (6)$$

ここに相互情報量 $I(G,\bar{G})$ とGのエントロピー $H(G)$ との比を $T(G,\bar{G})$ と書き表し、この $T(G,\bar{G})$ はあいまいの程度に関する尺度であることから、”あいまい度の尺度”として定義することにする。そうすると、$T(G,\bar{G})$ は上記の関係から式（７）となる。

$$T(G,\bar{G})=I(G,\bar{G})/H(G)$$
$$=1-\{H(G,\bar{G})-H(\bar{G})\}/H(G) \quad (7)$$

このあいまい度の尺度のとり得る範囲は $0 \leqq T(G,\bar{G}) \leqq 1$ であって、最も正確に分類された結果を得たとき（すなわち、$m_{ij}=0,\ i \neq j$）、$T(G,\bar{G})$ は最大値１となる。また、すべての m_{ij} が等しくなったとき、$T(G,\bar{G})$ は最小値０となる。

３　メンバーシップ関数による仮確率

前述のように、あいまい度の尺度の計算は確率行列によって計算される。メンバーシップ関数はあいまい度の尺度の計算に適用できない。しかし、分類スコアに類似する分類結果がメンバーシップ関数によって計算できれば、メンバーシップ関数によるあいまい度の尺度を求めることができよう。２節で述べたように、確率行列Pは頻度行列Mによって計算される。ファジイ理論によるとメンバーシップ関数によって頻度行列に類似する仮頻度（pseudo-frequencies）を求めることができる。もし、仮頻度行列を求めることができれば、自然に仮確率（pseudo-probabilities）も算出することができる。

与えられたデータ $X=\{x_1, x_2, ..., x_n\}$ に対し、データXをcクラスに分けるとき、分類前のＣｒｉｓｐ Ｓｅｔは $G=\{g_1, g_2, ..., g_c\}$（表２）とし、分類後のＦｕｚｚｙ Ｓｅｔは $\bar{G}=\{\bar{g}_1, \bar{g}_2, ..., \bar{g}_c\}$（表３）とする。Ｃｒｉｓｐ Ｓｅｔ のメンバーシップ関数値は $u_i(x_j)$ で表し、Ｆｕｚｚｙ Ｓｅｔのメンバーシップ関数は $u'_i(x_j)$ で表す。

表２　Ｃｒｉｓｐデータ

G ＼ X	x_1	x_2	・・・	x_n
g_1	1	0	・・・	0
g_2	0	1		1
・	・	・		・
・	・	・		・
・	・	・		・
g_c	0	0	・・・	0

まず表２のＣｒｉｓｐデータを見ると、与えられるデータXに対して、このデータをcクラス（g_i、$i=1, 2, ..., c$）に分けるとき、各点 $x_j \in X$ の帰属度はメンバーシップ関数 $u_i(x_j)$ で表す。この $u_i(x_j)$ は次の性質をもっている。

$$u_i(x_j) \in \{0, 1\},\quad \sum_{i=1}^{c} u_i(x_j)=1 \quad (8)$$

次に表３のＦｕｚｚｙデータを考えてみる。与えられるデータXに対して、このデータXをcクラス（$\bar{g}_i$、$i=1, 2, ..., c$）に分けるとき、各点 $x_j \in X$ の帰属度はファジィセットのメンバーシップ関数 $u'_i(x_j)$ で表す。$u'_i(x_j)$ は次の性質を持っている。

$$u'_i(x_j) \in [0, 1],\quad \sum_{i=1}^{c} u'_i(x_j)=1 \quad (9)$$

表３　ファジイデータ

Ḡ ＼ X	x_1	x_2	・・・	x_n
$\bar{g}_1$	0.98	0.00	・・・	0.00
$\bar{g}_2$	0.00	0.76		0.85
・	・	・		・
・	・	・		・
・	・	・		・
$\bar{g}_c$	0.02	0.24	・・・	0.15

ＣｒｉｓｐデータとＦｕｚｚｙデータ間の結合確率を求めるために各データ間のメンバーシップ関数の結合関数を定義しなければならない。ＣｒｉｓｐデータとＦｕｚｚｙデータ間の結合関数はファジィ集合の通集合の演算理論によって下記のように定義される。

$$\mu_{ij}(x_k)=u_i(x_k)\cdot u'_j(x_k) \quad x_k \in X \quad (10)$$

ここで、$i, j=1, 2, ..., c$、$k=1, 2, ..., n$ である。

$\mu_{ij}(x_j)$ は式（１０）によって計算されるとき、ＣｒｉｓｐデータとＦｕｚｚｙデータ間の結合頻度は式（１１）によって求めることができる。

$$N(i,j)=\sum_{k} \mu_{i,j}(x_k) \quad (11)$$

この頻度 $N(i,j)$ は２章で定義した頻度 m_{ij} とは概念上異なるから、ここでは $N(i,j)$ を仮頻度（ｐｓｅｕｄｏ－ｆｒｅｑｕｅｎｃｉｅｓ）と呼ぶことにする。

式（１１）の仮頻度によって、式（１２）で変換するとき、仮確率が得られる。

$$p(i,j)=N(i,j)/\sum_{i}^{c}\sum_{j}^{c} N(i,j)$$

この仮確率は導びき出された概念以外、数学的に２節の確率とまったく同じである。したがって、仮頻度が分かるなら、自然にこの仮頻度によって分類スコアの確率に類似する仮確率を求めることもできる。

４　結論

本論文では、従来の分類スコアの計算過程は集合の理論上で分析し、メンバーシップ関数による分類スコアの計算を試みた。そして、集合の演算原理の一つ、共通集合で分類結果の分類スコアを求めることができた。上記の集合間の演算を踏まえて、Ｃｒｉｓｐ集合とＦｕｚｚｙ集合間の結合関数を定義した。この結合関数の定義によって、ファジイデータの分類スコアー仮頻度を定義した。この仮頻度が定義されるから、この仮頻度によって自然に分類結果の仮確率を求めることができる。この仮確率の定義によって、従来あいまい度の尺度に適用できないファジイデータの分類結果にあいまい度の尺度 $T(G,\bar{G})$ が適用できるようになった。具体な適用結果は発表時に譲る。

参考文献

１）星　仰：あいまい度の尺度の理論とその性質ーリモートセンシングデータ分類の新しい評価尺度、日本写真測量とリモートトセンシング　ｖｏｌ．１９、Ｎｏ．２、ｐｐ．４－１０、１９８０。

２）星　仰、佟　国祥：ファジイセット尺度による分類スコアの評価法、写真測量とリモートシンシング、Ｖｏｌ．３１、Ｎｏ．３，ｐｐ．３４－４１、１９９２．