マージンの最大化による 楕円領域を持つファジィクラシファイアの 汎化能力の向上方式*

阿部 重夫[†]・坂口 慶太[†]

Improvement of Generalization Ability of a Fuzzy Classifier with Ellipsoidal Regions by Maximizing Margins*

Shigeo ABE[†] and Keita SAKAGUCHI[†]

Since a fuzzy classifier with ellipsoidal regions is based on the Mahalanobis distance, the generalization ability is degraded when the associated covariance matrices become singular. In this paper, we discuss two methods for improving the generalization ability: 1) during the symmetric Cholesky factorization of the covariance matrix we replace the input of the root with a prescribed positive value when it is smaller than the prescribed value, and 2) we tune the slopes of the membership functions so that the margins are maximized. We demonstrate the validity of our methods by computer simulations.

1. はじめに

多層ニューラルネットと同じく数値データから直接にファジィルールを抽出してパターン認識を行う学習機能付きのファジィクラシファイアの開発が進められている [1]. ファジィクラシファイアは入力領域の近似の仕方により, 超直方体領域を持つファジィクラシファイア, 多面体領域を持つファジィクラシファイア, 楕円領域を持つファジィクラシファイアとに分類される. このうち楕円領域を持つファジィクラシファイアでは, メンバシップ関数の傾きを解析的にチューニングすることにより認識率を向上する方法がとられており, 入力連続のときに多層ニューラルネットの最大認識率と同等の認識率が得られることが示されている [2,3]. しかしながら教師データ数が少ないときに重み付き距離を計算するときに用いる共分散行列が特異になるために汎化能力が低下する. このため参考文献 [4] では, 共分散行列を特異値分解して値の小さな特異値を制御することにより, 汎化能力を向上する方法を用いている.

対称 Cholesky 分解は行列が対称な正定行列のときに高速に三角行列の積に分解できるが, 特異になったとき

には一般には適用できず, 特異値分解を行う必要が生じ, 計算時間がかかるという問題がある. そこで本論文では, 共分散行列の特異値分解の代わりに対称 Cholesky 分解を用いて, 分解中に平方根の中が指定された正の値より小さくなったときに, 入力をその指定値で置き換えることにより汎化能力を向上する. さらに, 教師データに誤認識を生じないメンバーシップ関数の傾きの区間を解析的に求め, この区間の中心に傾きを設定することにより傾きのマージンを最大化する. これは, 最近注目を集めているサポートベクトルマシン (SVM) の汎化能力を最大化する学習方式 [5] の考え方に基づいたものである.

以下では, 2. で参考文献 [2,3] に基づいて楕円領域を持つファジィクラシファイアについて述べたあと, 3., 4. でおおの対称 Cholesky 分解による汎化能力の向上方式, および傾きのマージンを最大化する方法を述べる. 最後に 5. で四つのデータを用いて, 共分散行列を特異値分解する方法および SVM と比較して, 本論文で提案する方法の有効性を示す.

2. 楕円領域を持つファジィクラシファイア

m 次元入力ベクトル $\mathbf{x}$ を n 個のクラスに分類することを考える. 最初は一つのクラスに一つのクラスターを割り当てるが, 動的にクラスターを追加する方法によりクラス $i (i=1, \dots, n)$ に対する教師データを, いくつかのクラスター ij (クラス i の j 番目のクラスター) に分割する. そして各クラスター ij に対してつぎのファジィルールを定義

* 原稿受付 2001 年 2 月 27 日

[†] 神戸大学 大学院 自然科学研究科 Graduate School of Science and Technology, Kobe University; Rokkodai, Kobe, Hyogo 657-8501, JAPAN

Key Words: pattern recognition, fuzzy classifiers, membership tuning, symmetric Cholesky factorization.

する.

$$R_{ij}: \text{if } \mathbf{x} \text{ is } \mathbf{c}_{ij} \text{ then } \mathbf{x} \text{ belongs to class } i \quad (1)$$

ここで $\mathbf{c}_{ij}$ はクラス ij の中心で, クラス ij に属する教師データを用いてつぎのように計算する.

$$\mathbf{c}_{ij} = \frac{1}{|X_{ij}|} \sum_{\mathbf{x} \in X_{ij}} \mathbf{x} \quad (2)$$

ここで X_{ij} はクラス ij に属するデータの集合を示し, $|X_{ij}|$ は X_{ij} に含まれる要素の数である. 入力 $\mathbf{x}$ が与えられたときの (1) 式のメンバーシップ関数 $m_{ij}(\mathbf{x})$ をつぎのように定義する.

$$m_{ij}(\mathbf{x}) = \exp(-h_{ij}^2(\mathbf{x})) \quad (3)$$

$$h_{ij}^2(\mathbf{x}) = \frac{d_{ij}^2(\mathbf{x})}{\alpha_{ij}} \quad (4)$$

$$d_{ij}^2(\mathbf{x}) = (\mathbf{x} - \mathbf{c}_{ij})^t Q_{ij}^{-1} (\mathbf{x} - \mathbf{c}_{ij}) \quad (5)$$

ここで $d_{ij}(\mathbf{x})$ は Mahalanobis 距離 [6] とよばれる $\mathbf{x}$ と $\mathbf{c}_{ij}$ との間の重みつき距離, $h_{ij}(\mathbf{x})$ はチューニングされた距離, α_{ij} はクラス ij のチューニングパラメータ, Q_{ij} はクラス ij の $m \times m$ 次の共分散行列を表す. また Q_{ij}^{-1} は共分散行列の逆行列を示し $(\mathbf{x} - \mathbf{c}_{ij})^t$ は転置行列を示す. ここで共分散行列 Q_{ij} は

$$Q_{ij} = \frac{1}{|X_{ij}|} \sum_{\mathbf{x} \in X_{ij}} (\mathbf{x} - \mathbf{c}_{ij})(\mathbf{x} - \mathbf{c}_{ij})^t \quad (6)$$

で求める. ある入力 $\mathbf{x}$ に対して, メンバーシップ値 $m_{kl}(\mathbf{x})$ が最大なら $\mathbf{x}$ をクラス k に分類する. (3) 式は $[0,1]$ の範囲で変化し, (3) 式の入力を用いて入力 $\mathbf{x}$ を分類する場合は, 最小の $h_{ij}(\mathbf{x})$ を見つければよい.

楕円領域を持つファジィクラシファイアの学習では, まず各クラスが一つのクラスから構成されるとして, 中心と共分散行列をおおの (2), (6) 式で求めた後, 直接に教師データに対する認識率を最大化するようにメンバーシップ関数のパラメータをチューニングする方法を用いる.

ここでパラメータのチューニングは一つのパラメータを増加あるいは減少したときに誤認識を生じる点と誤認識していたデータが正しく認識する点を解析的に求め, 認識率が最大になるようにそのパラメータをチューニングし, 残りのパラメータについても順次チューニングすることにより行う.

チューニングの結果, 教師データの認識率が十分でないときは, 誤認識したデータにより, 新たにファジィルールを定義して, これらのファジィルールをチューニングすることにより認識率の向上を図る [2].

3. 対称 Cholesky 分解による汎化能力の向上

共分散行列 Q_{ij} は半正定であることが保証されているが, 特に正定となるためにはクラス ij に属するデータの数が少なくとも入力数に 1 を加えたものより大きい必要がある [6]. すなわち

$$|X_{ij}| \geq m + 1 \quad (7)$$

である. ここで Q_{ij} が特異のときは, Q_{ij} は通常特異値分解される [8]. 参考文献 [4] では小さな特異値を制限することにより, 楕円領域を持つファジィクラシファイアの汎化能力を向上している. しかしながら特異値分解を用いると学習時間がかかるという問題がある. このために以下では Q_{ij} の対称 Cholesky 分解 [8] 中に平方根の中の値が指定された正の値よりも小さくなった時点でその値を指定された値に置きかえることにより汎化能力を向上する方法を提案する.

Q_{ij} が正定の場合は Q_{ij} の対角要素は正で列要素の中で絶対値が最大となる. すなわち

$$q_{ii} > |q_{ij}| \quad i, j = 1, 2, \dots, m, \quad i \neq j \quad (8)$$

である. したがって Q_{ij} は対称 Cholesky 分解によりつぎのように二つの三角行列に分解できる [8].

$$Q_{ij} = L_{ij} L_{ij}^t \quad (9)$$

ここで L_{ij} は下三角行列で L_{ij} のおおのの要素は次式で与えられる.

$$l_{op} = \frac{q_{op} - \sum_{n=1}^{p-1} l_{pn} l_{on}}{l_{pp}} \quad \text{for } o = 1, \dots, m, \quad p = 1, \dots, o-1 \quad (10)$$

$$l_{aa} = \sqrt{q_{aa} - \sum_{n=1}^{a-1} l_{an}^2} \quad \text{for } a = 1, 2, \dots, m \quad (11)$$

L_{ij} を用いると (5) 式は次のように書きかえられる.

$$d_{ij}^2(\mathbf{x}) = (L_{ij}^{-1}(\mathbf{x} - \mathbf{c}_{ij}))^t L_{ij}^{-1}(\mathbf{x} - \mathbf{c}_{ij}) \quad (12)$$

ここでベクトル $\mathbf{y}_{ij} (= (y_{ij1}, \dots, y_{ijm})^t)$ を

$$\mathbf{y}_{ij} = L_{ij}^{-1}(\mathbf{x} - \mathbf{c}_{ij}) \quad (13)$$

と定義すると, 次式を

$$L_{ij} \mathbf{y}_{ij} = \mathbf{x} - \mathbf{c}_{ij} \quad (14)$$

$\mathbf{y}_{ij}$ について解くことにより, $d_{ij}^2(\mathbf{x})$ を Q_{ij} の逆行列を求めることなく

$$d_{ij}^2(\mathbf{x}) = \mathbf{y}_{ij}^t \mathbf{y}_{ij} \quad (15)$$

で求めることができる.

教師データの数が少ないときに (11) 式の平方根の中の値が非正となり、分解ができなくなる。これを避けるために、もし

$$q_{aa} - \sum_{n=1}^{a-1} l_{an}^2 \leq \eta \quad (16)$$

のときは平方根の中を次式のように正の η で置きかえる。

$$l_{aa} = \sqrt{\eta} \quad (17)$$

三角行列に分解途中に平方根の中の値を正の値で置き換えることは、教師データが少ないために分散の値が小さくなっているのを、大きくすることに対応し、あいまい性を増すはたらきがある。

4. 傾きのマージンの最大化による汎化能力の向上

4.1 考え方

楕円領域をもつファジィクラシファイアではクラス間に重なりがあると、メンバーシップ関数を調整することにより重なりを調整する。しかしながら、クラス間に重なりがないとメンバーシップ関数の調整は行われない。教師データの数が少ないときには通常の場合、重なりはあまりなく、汎化能力が低下することになる。教師データ数が少ないときにもメンバーシップ関数をチューニングして汎化能力を向上するために、サポートベクトルマシンの学習方式の考え方を利用する。サポートベクトルマシンでは二つのクラスを分離する超平面を決定する場合に、二つのクラスの教師データの中で超平面に最も近いデータからの距離が最大になるように決定する [5]。このとき、これらのデータをサポートベクトルとよび、決定された超平面を最適超平面、サポートベクトルから最適超平面までの距離をマージンとよぶ。楕円領域を持つファジィクラシファイアにおいては、メンバーシップ関数の傾きのマージンが最大になるように傾きをチューニングする。

メンバーシップ関数をチューニングする際に α_{ij} の初期値を 1 にするが、もし各クラスが一つのクラスタからなるときは、楕円領域をもつファジィクラシファイアは Mahalanobis 距離をベースにした識別器と等価となる。つぎに正しく認識されていた教師データが誤認識しない範囲でマージンが最大になるように傾きをチューニングする。チューニングした後で、教師データの認識率が 100% でないときは、参考文献 [2] に示された手順で誤認識を解消するように α_{ij} をチューニングする。以下ではまずマージンを最大化する方法について述べる。

サポートベクトルマシンの学習では三つ以上のクラスの分離問題では三つの二クラス問題に分離して最適超平面を決定するが、メンバーシップ関数のチューニングを行うときはこのような制約は不要である。しかし、以下

ではチューニングの考え方を各クラスが一つのクラスタからなる Fig. 1 の二クラスの問題についてチューニング法を説明する。図において黒塗りの四角および円はおののクラス i , o に属し、クラス o , i に近い教師データとする。このときクラス間の境界は二つの曲線の間に存在するが、これらの曲線の中央に境界を設定することにより汎化能力の最大化を図る。

Fig. 2 において、データがクラス i に属するとすると、クラス i の成立度が大きいために正しく認識されている。このデータはクラス i のメンバーシップ関数が図に示す点線の位置まで減少されるまで正しく認識される。同様に Fig. 3 においてデータがクラス o に属するとすると、クラス o の成立度が大きいために正しく認識されている。このデータはクラス i のメンバーシップ関数が図の点線の位置まで増加されるまで正しく分類される。したがって、このようにして決まる誤認識を起こさない区間の中央に α_{ij} を設定すると傾きのマージンが最大化される。以下では α_{ij} をチューニングする方法をさらに詳細に議論する。

4.2 α_{ij} の上下限

X を初期の α_{ij} で正しく認識された教師データの集合とする。ここで $\mathbf{x}(\in X)$ をクラス i に属するとする。も

Fig. 1 Concept of maximizing margins

Fig. 2 Upper bound of α_{ij}

Fig. 3 Lower bound of α_{ij}

し、クラス i に対するメンバーシップ値 $m_{ij}(\mathbf{x})$ が最大でない場合は、 α_{ij} を減少することにより $m_{ij}(\mathbf{x})$ が減少しても誤認識は生じない。同様に、 $m_{ij}(\mathbf{x})$ が最大でかつ $m_{ik}(\mathbf{x}) (k \neq j)$ が二番目に大きいときは、 α_{ij} を減少することにより $m_{ij}(\mathbf{x})$ が減少しても $m_{ij}(\mathbf{x})$ に代わって $m_{ik}(\mathbf{x})$ が最大値をとるために $\mathbf{x}$ は誤認識しない。したがって、これらのデータはチューニングから除外する。

もし $m_{ij}(\mathbf{x})$ が最大で $m_{ik}(\mathbf{x}) (k \neq j)$ の中で二番目に大きいものがないときは α_{ij} を減少することにより $m_{ij}(\mathbf{x})$ が減少して、二番目に大きい $m_{op}(\mathbf{x}) (o \neq i)$ と等しくなった時点で誤認識を生じる。すなわち、データ $\mathbf{x}$ が誤認識を起こさない α_{ij} の下限 $L_{ij}(\mathbf{x})$ が次式で与えられる。

$$L_{ij}(\mathbf{x}) = \frac{d_{ij}^2(\mathbf{x})}{\min_{o \neq i} h_{op}^2(\mathbf{x})} \quad (18)$$

これより (18) 式であたえられる下限の最大値をすべての正しく認識されたデータの中で求めることにより、正しく認識されたデータに誤認識が生じない下限 $L_{ij}(1)$ が求まる。すなわち

$$L_{ij}(1) = \max_{\mathbf{x} \in X} L_{ij}(\mathbf{x}) \quad (19)$$

で下限が与えられる。

同様に、クラス i 以外のクラスに属するデータ $\mathbf{x} (\in X)$ に対して α_{ij} の上限を計算することができる。 $m_{jk}(\mathbf{x}) (o \neq i)$ が最大であるとする、 $\mathbf{x}$ の誤認識を生じない α_{ij} の上限 $U_{ij}(\mathbf{x})$ は

$$U_{ij}(\mathbf{x}) = \frac{d_{ij}^2(\mathbf{x})}{\min_q h_{oq}^2(\mathbf{x})} \quad (20)$$

で与えられる。

これより誤認識を生じない上限 $U_{ij}(1)$ は

$$U_{ij}(1) = \min_{\mathbf{x} \in X} U_{ij}(\mathbf{x}) \quad (21)$$

で与えられる。

4.3 α_{ij} のチューニングの手順

正しく認識されたデータは α_{ij} を $(L_{ij}(1), U_{ij}(1))$ の区間の任意の値に設定しても正しく認識されたままである。これより α_{ij} のマージンを最大化するチューニングをつぎのように行う。 α_{ij} に対して $L_{ij}(1)$ と $U_{ij}(1)$ を計算して、 α_{ij} を $L_{ij}(1)$ と $U_{ij}(1)$ の中央の値に設定する。すなわち

$$\alpha_{ij} = \frac{1}{2}(L_{ij}(1) + U_{ij}(1)) \quad (22)$$

とする。

上記のチューニングを各 α_{ij} について一度だけ行う。チューニング結果はチューニングの順序により異なるが、いくつかのデータで評価した結果ではそれほど差がないために以下のシミュレーションではクラス 1 から順次行うことにする。

以上の手順では、正しく認識されたデータが誤認識を起こさない範囲でマージンが最大となるようにチューニングを行ったために、 α_{ij} の初期値で教師データに誤認識がある場合は、チューニング後も誤認識が残る可能性がある。誤認識が残ったときは誤認識を解消するように α_{ij} のチューニングを行う [2]。その大まかな手順を以下に示す。

$l-1$ 個まで正しく認識されていたデータが誤認識してもよい α_{ij} の上限と下限を $U_{ij}(l), L_{ij}(l)$ とする。たとえば α_{ij} が $[U_{ij}(1), U_{ij}(2))$ の間にセットされていれば、一つのデータだけクラス i に認識されていたのが誤認識されるのを許す。

α_{ij} の増減で誤認識されていたデータが正しく認識される可能性がある。 $\beta_{ij}(l)$ を $U_{ij}(l)$ より小さくて、以前誤認識されていたデータが正しく認識される α_{ij} の上限を表すとする。また $\gamma_{ij}(l)$ を $L_{ij}(l)$ より大きくて、以前誤認識されていたデータが正しく認識する下限を表すとする。

つぎに $(L_{ij}(l), \gamma_{ij}(l))$ と $(\beta_{ij}(l), U_{ij}(l)) (l=1, \dots)$ の区間の中でどの区間に α_{ij} を設定すれば認識率が最大になるかを調べる。探索区間を限定するために l の最大値、すなわち l_M をパラメータとして設定する。 $(L_{ij}(l), \gamma_{ij}(l))$ が教師データの最大認識率を得る区間とすると、 α_{ij} をその区間内のどこに設定しても教師データの認識率は変化しない。しかし、テストデータに対する認識率は変化するので、汎化能力を調整するために α_{ij} をつぎのようにセットする。

$(\beta_{ij}(l), U_{ij}(l))$ の区間で最大認識率を得るとき、

$$\alpha_{ij} = \beta_{ij}(l) + \delta(U_{ij}(l) - \beta_{ij}(l)) \quad (23)$$

$(L_{ij}(l), \gamma_{ij}(l))$ の区間で最大認識率を得るとき、

$$\alpha_{ij} = \gamma_{ij}(l) - \delta(\gamma_{ij}(l) - L_{ij}(l)) \quad (24)$$

ただし $0 < \delta < 1$ である。

5. シミュレーションによる評価

共分散行列が特異になるクラスがある数字データ [9] およびひらがなデータを用いて本論文で提案した手法と共分散行列を特異値分解したとき [4] および SVM で評価したときの最良値と比較した。Table 1 に各データの入力数、クラス数、教師データ数、テストデータ数を示す。表において hiragana-50 および hiragana-105 はナンバープレートから切り出したひらがなデータの画像をおおの 10×5 および 15×7 の小領域に分割して、おおのの小領域の濃淡を平均値を入力値としたもので、hiragana-13 は hiragana-105 の二次元のモーメントを計算して最初の 13 個の中央モーメントに変換したものである。

Table 1 Feature of benchmark data

Data	Inputs	Classes	Train.	Test
Numeral	12	10	810	820
Hiragana-50	50	39	4610	4610
Hiragana-105	105	38	8375	8356
Hiragana-13	13	38	8375	8356

なお楕円領域を持つファジクラシファイアのパラメータは $\alpha_{ij}=1, \delta=0.1$ とし、おおののクラスは一つのクラスタからなるとした。また、数字、ひらがなデータでは、正しく認識していたデータが誤認識を許す最大の数を表す l_M を $l_M=10$ とした。特異値分解を用いたときは、テストデータの認識率が最大となるように値の小さな特異値を 0 とした [4]。サポートベクトルマシンでは参考文献 [10] を用いて、テストデータの認識率が最大となるようにカーネル関数等を調整した。

各クラシファイアの比較はテストデータの認識率のよさで判断した。計算には Sun UltraSPARC IIi 333MHz を用い、以下に述べる時間は学習と教師データ、テストデータの認識をあわせた時間である。

Table 2 に Cholesky 分解において (16) 式の η を変えたときの数字データの認識率と学習時間を示す。認識率はテストデータと括弧内に教師データの認識率を示す。“Initial” の列は $\alpha_{ij}=1$ のときの認識率を示し、“Final” は誤認識を解消するために α_{ij} をチューニングした後の認識率を示す。 η が増加するにつれ、テストデータの最終の認識率は向上し、 $\eta=10^{-2}$ のときに最もよくなった。

Table 3 に η を変えて α_{ij} のマージンが最大となるようにチューニングした数字データの結果を示す。“Initial” の列は α_{ij} のマージンを最大化したときの結果を示し、“Final” は誤認識を解消するように α_{ij} をチューニングした結果を示す。Tables 2, 3 を比較するとおおのの η に対して認識率はほとんど変わりが無い。

Table 4 に提案手法とサポートベクトルマシンを用いたときの数字データの最大の認識率を示す。表において

Table 2 Recognition rates for numeral data

η	Initial [%]	Final [%]	Time [s]
10^{-5}	99.63 (99.75)	99.39 (99.88)	0.5
10^{-4}	99.51 (99.75)	99.51 (99.88)	0.5
10^{-3}	99.51 (99.75)	99.51 (99.88)	0.5
10^{-2}	98.90 (99.63)	99.76 (100)	0.5
10^{-1}	98.29 (98.77)	99.02 (99.51)	0.6

Table 3 Recognition rates for numeral data by maximizing margins

η	Initial [%]	Final [%]	Time [s]
10^{-5}	99.15 (99.75)	99.27 (99.88)	1.2
10^{-4}	99.27 (99.75)	99.51 (99.88)	1.2
10^{-3}	99.76 (99.75)	99.76 (99.88)	1.2
10^{-2}	99.39 (99.63)	99.88 (100)	1.3
10^{-1}	98.29 (98.89)	98.66 (99.38)	1.3

Table 4 Best performance for numeral data

Method	Rates [%]	Time [s]
CF	99.76 (100)	0.5
CF + MM	99.88 (100)	1.2
SVM	99.76 (100)	27

CF は Cholesky 分解を示し、CF + MM は Cholesky 分解にマージンの最大化が行われたことを示す。方式上の差異はほとんどないといえる。特異値分解の最良値は Table 2 の第一行の結果と同じであった。

Table 5 に Cholesky 分解において (16) 式の η を変えたときのひらがな 50 データの認識率と学習時間を示す。テストデータの最終の認識率は η が増加するにつれ増加し $\eta=10^{-2}$ のときに最大となった。

Table 5 Recognition rates for hiragana-50 data

η	Initial [%]	Final [%]	Time [s]
10^{-5}	77.51 (100)	—	35
10^{-4}	81.39 (100)	—	35
10^{-3}	95.53 (99.98)	95.64 (100)	43
10^{-2}	98.79 (99.65)	98.85 (100)	101
10^{-1}	94.43 (96.14)	94.69 (97.57)	310

Table 6 に η を変えて α_{ij} のマージンを最大化したときのひらがな 50 データの結果を示す。Tables 5, 6 を比較すると、テストデータの最大値は同じであるが、 $\eta=10^{-5}$, 10^{-4} のときのマージンを最大化したときの認識率は最大化しないときよりも大幅に向上している。

Table 7 に提案手法を用いたとき、特異値分解 (SVD) を用いたときおよびサポートベクトルマシンを用いたときの認識率の最大値を示す。SVD を除いてはほぼ同程度

Table 6 Recognition rates for hiragana-50 data by maximizing margins

η	Initial [%]	Final [%]	Time [s]
10^{-5}	90.50 (100)	—	431
10^{-4}	92.17 (100)	—	458
10^{-3}	96.25 (100)	—	575
10^{-2}	98.83 (99.87)	98.85 (100)	488
10^{-1}	94.64 (96.59)	94.84 (97.51)	996

Table 7 Best performance for hiragana-50 data

Method	Rates [%]	Time [s]
CF	98.85 (100)	101
CF + MM	98.85 (100)	488
SVD	94.51 (99.89)	946
SVM	99.07 (100)	7144

Table 8 Recognition rates for hiragana-105 data

η	Initial [%]	Final [%]	Time [s]
10^{-5}	93.59 (100)	—	304
10^{-4}	97.93 (100)	—	305
10^{-3}	100 (100)	—	305
10^{-2}	99.99 (99.99)	100 (100)	380

Table 9 Recognition rates for hiragana-105 data by maximizing margins

η	Initial [%]	Final [%]	Time [s]
10^{-5}	99.23 (100)	—	7459
10^{-4}	99.90 (100)	—	4284
10^{-3}	100 (100)	—	3284
10^{-2}	99.99 (99.99)	100 (100)	3312

の認識率になっている。

Table 8 に Cholesky 分解において (16) 式の η を変えたときのひらがな 105 データの認識率と学習時間を示す。 η が増えるにつれ、最終のテストデータの認識率が上昇し、 $\eta=10^{-3}$ のときに教師データとテストデータの認識率が 100% となった。

Table 9 に η を変えて α_{ij} のマージンを最大化したときのひらがな 105 データの結果を示す。Tables 8, 9 を比較すると、テストデータの最大値は同じであるが、 $\eta=10^{-5}$, 10^{-4} のときのマージンを最大化したときの認識率は最大化しないときよりも大幅に向上している。

Table 10 に提案手法を用いたとき、特異値分解 (SVD) を用いたときおよびサポートベクトルマシンを用いたときの認識率の最大値を示す。四つの方法はほぼ同程度の認識率になっている。

Table 10 Best performance for hiragana-105 data

Method	Rates [%]	Time [s]
CF	100 (100)	305
CF+MM	100 (100)	3284
SVD	99.99 (100)	2467
SVM	100 (100)	10443

Table 11 Recognition rates for hiragana-13 data

η	Initial [%]	Final [%]	Time [s]
10^{-6}	98.36 (99.84)	98.79 (99.99)	26
10^{-5}	98.59 (99.86)	99.02 (99.99)	26
10^{-4}	99.41 (99.86)	99.59 (99.96)	31
10^{-3}	99.49 (99.62)	99.44 (99.77)	42
10^{-2}	95.22 (95.52)	96.48 (97.00)	110

Table 12 Recognition rates for hiragana-13 data by maximizing margins

η	Initial [%]	Final [%]	Time [s]
10^{-6}	99.34 (99.94)	99.34 (99.99)	154
10^{-5}	99.46 (99.94)	99.45 (99.99)	155
10^{-4}	99.41 (99.90)	99.43 (99.96)	156
10^{-3}	99.56 (99.67)	99.49 (99.79)	185
10^{-2}	95.64 (96.08)	96.27 (96.93)	240

Table 13 Best performance for hiragana-13 data

Method	Rates [%]	Time [s]
CF	99.59 (99.96)	31
CF + MM	99.49 (99.79)	185
SVD	98.79 (99.99)	36
SVM	99.77 (100)	23216

Table 11 に Cholesky 分解において (16) 式の η を変えたときのひらがな 13 データの認識率と学習時間を示す。 η が増えるにつれ、最終のテストデータの認識率が上昇し、 $\eta=10^{-4}$ のときにテストデータの認識率が最大となった。

Table 12 に η を変えて α_{ij} のマージンを最大化したときのひらがな 13 データの結果を示す。Tables 11, 12 を比較すると、テストデータの最大値は同じであるが、 $\eta=10^{-6}$, 10^{-5} のときのマージンを最大化したときの認識率は最大化しないときよりも向上している。

Table 13 に提案手法を用いたとき、特異値分解 (SVD) を用いたときおよびサポートベクトルマシンを用いたときの認識率の最大値を示す。SVD を除いてはほぼ同程度の認識率になっている。

6. おわりに

教師データが少ないときの楕円領域を持つファジィクラシファアの汎化能力を向上する以下の二つの方法を提案した。

- (1) 共分散行列を Cholesky 分解中に、平方根の中の値が指定された正の値より小さくなったときに、その値を指定値で置き換える。
- (2) 正しく認識された教師データが誤認識を起こさないメンバーシップ関数の傾きの区間を求め、その区間の中央に傾きを設定し直す。

以上の方式を共分散行列が特異になる車のナンバープレートの数字およびひらがなデータを用いて評価したところ、(1)の方法で大幅にテストデータの認識率が向上することが示された。また(1)の方法に加えて(2)の方法により、指定値が適切でないときにテストデータの認識率が向上することが示された。

また従来の特異値分解の小さな特異値を制御する方法よりもテストデータの認識率が向上し、サポートベクトルマシンと同等のテストデータの認識率となることが確かめられた。

今回扱ったデータは最大で100入力で8000データ程度であるが、入力数、およびデータ数が大きいときに、ファジィクラシファアがサポートベクトルマシンより高速でかつ同等の汎化能力を実現できるかは今後の課題とした。

なお、本研究は、文部省科学研究補助金 基盤研究(C) 12650409「ファジィパターン認識システムの統一的学習パラダイムの開発」の補助を得て行ったものである。

謝 辞

本研究を進めるにあたり、ご討論いただいた本学小澤誠一助教授、米森秀登助手、数字およびひらがなデータを提供していただいた(株)日立製作所日立研究所小林芳樹技術主幹に感謝の意を表す。

参 考 文 献

- [1] 阿部：ニューラルネットとファジィシステム：理論と応用，近代科学社（1995）
- [2] S. Abe and R. Thawonmas: A fuzzy classifier with ellipsoidal regions; *IEEE Trans. Fuzzy Systems*, Vol. 5,

No. 3, pp. 358–368 (1998)

- [3] S. Abe: Dynamic cluster generation for a fuzzy classifier with ellipsoidal regions; *IEEE Trans. Systems, Man and Cybernetics-Part B*, Vol. 28, No. 6, pp. 869–876 (1998)
- [4] 小林, 木股, 阿部：楕円領域を持つファジィクラシファアの最適データ分割による汎化能力の向上方式；システム制御情報学会論文誌, Vol. 13, No. 2, pp. 87–94 (2000)
- [5] V. N. Vapnik: *Statistical Learning Theory*, John Wiley & Sons (1998)
- [6] R. O. Duda and P. E. Hart: *Pattern Classification and Scene Analysis*, John Wiley & Sons, pp. 67–68 (1973)
- [7] W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling (丹慶, 奥村, 佐藤, 小林)：特異値分解；*Numerical recipes in c* [日本語版], pp. 73–84, 技術評論社 (1993)
- [8] G. H. Golub and C. F. Van Loan: *Matrix Computations*, Third Edition, John Hopkins University Press (1996)
- [9] 武長, 阿部, 高藤, 鹿山, 北村, 奥山：感度解析を用いたニューラルネットの入力層の最適化とその数字認識への適用；電学論 D, Vol. 111-D, No. 1, pp. 36–44 (1991)
- [10] C. Saunders, M. O. Stitson, J. Weston, L. Bottou, B. Schölkopf and A. Smola: Support Vector Machine - Reference Manual; Royal Holloway University of London CSD-TR-98-03 (1998)