

カーネル非線形部分空間法によるパターン認識

前田 英作[†] 村瀬 洋^{††}

Kernel Based Nonlinear Subspace Method for Pattern Recognition

Eisaku MAEDA[†] and Hiroshi MURASE^{††}

あらまし 新しいパターン識別手法、カーネル非線形部分空間法 (Kernel based Nonlinear Subspace method; KNS 法) を提案する。本手法はカーネル関数によって定義された非線形変換を利用して高次元非線形空間上での部分空間法を実現したものである。近年研究が盛んな Support Vector Machine はカーネル関数を利用した非線形識別手法であり高い識別性能を有するが、パターン数、クラス数の増加に伴い学習に要する計算量が爆発的に増えるという問題がある。一方、従来の部分空間法は多クラスの識別に有効でかつ高速な識別手法であるが、パターンの分布が非線形性をもつ場合やクラス数に比較して特徴空間の次元が小さい場合に十分な識別性能が得られない。提案手法は、両者の利点を組み合わせることにより、互いの欠点を補完し高い識別性能を有する多クラスの非線形識別を少ない計算量で実現する。本論文では、非線形部分空間法がカーネル関数で定義される非線形変換を用いて定式化可能なことを示し、非線形分布及び多クラス分布に対する識別性能、パラメータ変動に対する識別性能の安定性、学習及び識別に要する計算コストなどの観点から提案手法を評価し、従来手法に比べて優れていることを検証した。

キーワード パターン認識, 部分空間法, カーネル関数, サポートベクトルマシン, 非線形変換

1. ま え が き

計算機によるパターン認識の実現法には、パターンの確率密度関数の形を既知として与え関数のパラメータを学習パターンから推定することにより識別関数を構成するパラメトリックな方法と、パターンの確率密度関数の形を未知として学習パターンから直接識別関数を設計するノンパラメトリックな方法とがある。現実の多くの問題では特徴空間におけるパターンの分布形を仮定できないことが多く、実用性がより高いのはノンパラメトリックな方法である。その代表例の一つが線形関数によって識別関数を記述する方法であり、種々の基準において最適な線形識別関数を得るための学習アルゴリズムも知られている。そして、線形識別関数は記述が簡単なわりに頑健で実用的価値も高い。

しかし、真の境界面が複雑な形状をしているときには線形識別関数では十分な性能が実現できないので、

非線形関数を用いて識別関数を記述することが必要となる。非線形手法の代表的な例として区分的線形識別関数法 [1]、 Φ 関数法 (一般線形識別関数法) [1]、ポテンシャル関数法 [2]、最近傍決定則 [3]、ニューラルネットワーク [4] などが知られている。しかしながら、区分的線形識別関数法はその収束性や最適性など学習アルゴリズムの特性に関する解析が不十分である、 Φ 関数法は最適な非線形関数の組を決定する方法が明らかでない、ポテンシャル関数法や最近傍決定則は特に特徴空間が高次元のときに識別精度を高めるには多くの学習パターンを必要とする、ニューラルネットワークは一般に局所最適解しか得られない、などの問題点ははらんでおり一長一短がある。

最近、新しい非線形パターン認識手法として Support Vector Machine (SVM) [5], [6] が注目されている。SVM は、マージン最大化基準を用いることによって最適化問題が凸 2 次計画法の問題に帰着され、大域的に最適な識別関数が求まる点、カーネル関数を介して定義される非線形変換によってパターンを高次元空間へ写像する点に特徴がある。しかし、SVM は 2 クラスの識別を行う識別機であること、パターン数の増加に伴って最適解を得るのに必要な計算量が膨大にな

[†] NTT コミュニケーション科学基礎研究所, 京都府
NTT Communication Science Laboratories, Seika-cho,
Soraku-gun, Kyoto-fu, 619-0237 Japan

^{††} NTT コミュニケーション科学基礎研究所, 厚木市
NTT Communication Science Laboratories, Atsugi-shi, 243-
0198 Japan

ることなどの問題点が指摘されている。

一方、従来から文字認識などにおいて多クラスの識別手法として利用されている部分空間法は、各クラスを特徴づける部分空間をクラスごとに定めて識別に利用する方法である。そして、一般にクラス数が多い場合には特徴空間の次元が高いことが望ましい。したがって、カーネル関数を利用した高次元空間への非線形変換と部分空間法とを組み合わせることができれば、それぞれの利点を生かしてより少ない計算量でより高い識別性能を有する多クラス識別のための非線形識別機が設計できると期待される。

本論文では、まず、SVM 及び部分空間法の問題点を述べた後、カーネル関数を介して定義される非線形変換を用いて非線形部分空間法が定式化可能なことを示す。次に、非線形性をもつ分布に対する識別性能、多クラス分布に対する識別性能、パラメータ変動に対する識別性能の安定性、学習及び識別に要する計算コストなどの観点から提案手法を評価し、従来手法に比べて優れていることを検証した結果について報告する。以下、提案手法をカーネル非線形部分空間法と呼び、KNS 法 (Kernel based Nonlinear Subspace method) と略記する。また最近、本研究 [7] ~ [9] とは独立に同様のアイデアが報告された [10]。

2. カーネル関数とパターン認識

2.1 Parzen window とポテンシャル関数法

ノンパラメトリックな確率密度推定の一手法である Parzen window 法は、カーネル関数と呼ばれる関数の線形和によって確率密度関数を記述する方法である [11]。クラスごとの確率密度関数と事前確率を精度良く推定することができれば、ベイズの定理から最適な識別機を設計することができる。こうした考え方を発展させてカーネル関数をパターン認識に直接利用した手法がポテンシャル関数法である [2]。これはカーネル関数を用いて直接識別関数を記述する方法であり、識別関数 $f(x)$ をカーネル関数 k と学習パターン x_i ($i = 1, \dots, n$) を用いて

$$f(x) = \sum_{i=1}^n q_i k(x, x_i) \quad (1)$$

と定義し、学習によってパラメータ q_i の最適値を決定する。ただし、 x は d 次元パターンベクトル、 n は学習パターン数を表す。特に、 \mathcal{R}^d から \mathcal{R}^{d_ϕ} への写像、

$$\phi: x \mapsto \phi(x) = (\phi_1(x), \dots, \phi_{d_\phi}(x))^t \quad (2)$$

を用いて、パターン x, y に対し

$$k(x, y) = \phi(x)^t \phi(y) = \sum_{i=1}^{d_\phi} \phi_i(x) \phi_i(y) \quad (3)$$

と展開可能なカーネル関数のクラスについて、学習アルゴリズムの収束性能などの様々な研究がなされた [12], [13]。

しかし、Parzen window による確率密度関数やポテンシャル関数法による識別関数の記述には推定精度、識別性能を高めるために非常に多くの学習パターンを必要とするという欠点があり、特に特徴空間の次元が高いときには深刻な問題であった。そして、カーネル関数の特徴を利用した新しい非線形識別関数の構成法として Vapnik によって提案されたのが Support Vector Machine (SVM) である [5], [6]。

2.2 カーネル関数と非線形 SVM

非線形 SVM は、非線形変換によって特徴ベクトルを原空間よりもはるかに高次元の空間へ写像した後に線形 SVM を適用し、高次元空間上で最適な超平面を求めることによって得られる。したがって、非線形変換後に線形識別関数の最適化を行うという点において SVM の学習は Φ 関数法 [1] と同じである。しかしながら、非線形 SVM における非線形変換はカーネル関数を介して定義されているという点が SVM の学習と Φ 関数法との大きな違いであり、これにより非線形 SVM の最適化問題は線形 SVM と同じく凸 2 次計画法に帰着できる。そして、非線形関数の具体的な形を知る必要がなく、原空間のパターンベクトル間の演算によって高次元空間での最適化問題を解くことができる。

ここで、パターンベクトル x, y のある関数 $k(x, y)$ に対し式 (3) を満たす ϕ として非線形変換 ϕ を定義する。ただし、 d_ϕ は $\phi(x)$ の次元を表す。非線形 SVM において、カーネル関数 $k(x, y)$ は $\phi(x)$ と $\phi(y)$ の関数としてではなく、以下の例のように x と y の関数として定義される。

$$k(x, y) = x^t y \quad (4)$$

$$k(x, y) = (1 + x^t y)^p \quad (5)$$

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2p^2}\right) \quad (6)$$

ここで p は任意の定数である。 k を式 (4) と定義すれば ϕ は恒等写像であるから線形 SVM と等価であり、式

(5) は p 次多項式関数に、式 (6) は分散 p^2 の広がりをもつ Gaussian Radial Basis Function (RBF) [14] に相当する [5], [6]。また、式 (3) を満たす具体的な ϕ が存在するためのカーネル関数 $k(x, y)$ の条件は、マーサーの条件として知られている [15]。

このようなカーネル関数を用いた非線形変換の定式化には二つの特徴がある。第 1 に ϕ で写像される空間が一般に非常な高次元になる点である。式 (5) の例では ϕ の次元は $d_\phi = d+pC_p - 1$ であり、元の特徴空間の次元を 100、カーネル関数を 3 次の多項式関数で定義したとすれば、 $\phi(x)$ の次元は 176,850 にもなる。この高次元空間の各軸は互いに線形独立であり、原空間と比べて識別により有利な超平面が存在すると期待できる。第 2 に非線形 SVM の最適化、すなわち高次元空間での線形 SVM の最適化のように、問題がカーネル関数を用いて定式化できれば、低次元原空間のパターン間の演算で解くことができる。これは計算量の観点から大きな利点となる。

2.3 SVM の問題点

一方、SVM には次のような問題点が指摘されている。第 1 に SVM の最適化問題、すなわち凸 2 次計画法を解くためには多くの計算時間を必要とする。しかも学習パターン数に比例して 2 次計画法の規模は大きくなり、それを解くための計算量は爆発的に増大する (図 5 参照)。第 2 に、SVM は 2 クラスパターンの識別機であり、多クラスに適用する場合には工夫が必要となる。例えばクラス対ごとに設計した識別関数を組み合わせる多クラスの識別を行う方法は古くから知られており [16], [17]、これを SVM と組み合わせることが可能である。また、多クラスの識別を目的としてクラス対ごとの線形識別関数を一度の最適化によって解く方法が提案されており [18]、これを SVM の 2 次計画法に組み込んで多クラスの識別のための SVM を実現する方法も提案されている [19], [20]。ただし、クラス数の増大とともに最適化問題の規模も増大し、計算量の問題に直面する。

3. 線形部分空間法

3.1 CLAFIC 法と投影距離法

本論文では部分空間法を広義の意味でとらえ、各クラスを特徴づける部分空間を学習パターンからクラスごとに決定し、そのクラス部分空間を利用して未知パターンの識別を行う方法を部分空間法と呼ぶこととする。この広義の部分空間法は大きく二つに分けられる。

第 1 のものは特徴空間の原点を起点として各クラス分布ごとに KL 展開を行ってクラス部分空間を決定し、未知パターンのクラス部分空間への射影ベクトルの長さが最大となる部分空間のクラスに判定する方法である。狭義にはこの手法を部分空間法と呼ぶ。特徴空間の原点はクラス部分空間の原点でもあるので、この判定法は未知パターンとその射影ベクトルの方向余弦の大きさを判定することと等価である。CLAFIC 法 [21] がその例であり、射影ベクトルの長さの代わりに重み付けされた長さで評価する複合類似度法 [22] はその変法とみなすことができる [23]。

第 2 のものは各クラス重心を起点として各クラス分布ごとに KL 展開を行ってクラス部分空間を決定し、未知パターンからクラス部分空間への投影距離が最小となる部分空間のクラスに判定する方法であり、投影距離法 [24] と呼ばれる。

CLAFIC 法はクラス部分空間を作るときの原点を特徴空間の原点に置くが、投影距離法では各クラスの重心に置くという点がこれら二つの手法の本質的な違いである。したがって、CLAFIC 法では特徴空間の原点と各クラスの分布との相対位置に識別結果が依存するが、投影距離法では依存しない。また、分布が特徴空間の原点をまたいでいる場合には明らかに CLAFIC 法はうまく機能しない。しかしながら、一般的には、CLAFIC 法と投影距離法の識別性能の差は一概には結論できない。

3.2 部分空間法の問題点

前節で述べた線形部分空間法は、多クラスの識別が少ない計算量で実現できる一方で、二つの欠点がある。まず第 1 に、分布が非線形な軸に沿って広がる場合、線形主成分分析によって定まる主成分方向は意味をなさない。したがって、得られるクラス部分空間が必ずしもそのクラスを特徴づける部分空間にはならない。第 2 に、クラス数に対する特徴空間の次元の比が小さい場合、各クラス部分空間同士の重なりが増え、一般に識別性能が低下する。特徴空間の次元を大きくするためには特徴量の数を増やせばよいが、これは必ずしも容易ではない。クラス数が多い課題に対して部分空間法を適用する場合には大きな問題となる。

第 1 の欠点を補う方法として、自己想起型ニューラルネットを用いた次元圧縮型の非線形主成分分析と部分空間法とを組み合わせた識別手法が提案されている [25]。この方法では、クラス部分空間を構成する際に分布の非線形性を吸収することが可能であり、線形

部分空間法よりも性能の高い識別が期待できる．しかしながら，クラス部分空間を構成するためにはニューラルネットの学習が必要となり，局所解に陥る危険性を避けることができない．更に，次元圧縮型の自己想起型回路を用いているので前節で述べた線形部分空間法の第 2 の欠点を克服することができない．

一方，SVM で使われているカーネル関数を介して定義された非線形変換は，前節で述べたように非常に高次元への非線形変換である．そこで，このカーネル非線形変換と部分空間法とを組み合わせることができれば上で述べた部分空間法の二つの欠点を克服することが期待できる．更に，部分空間法は SVM のように 2 次計画法を解く必要がなく，また次節で述べるように，カーネル関数によって定式化できるので ϕ を陽に用いることなく識別が可能となり，計算コストの観点からも有利である．

4. カーネル非線形部分空間法 (KNS 法)

本章では，カーネル非線形部分空間法 (KNS 法) の概要を述べる．まず，クラス非線形部分空間がカーネル非線形主成分分析を用いて定義できることを示し，次に，識別のための基準がカーネル関数を用いて記述できることを示す．

4.1 カーネル関数による非線形部分空間の構成

カーネル関数を利用してクラス非線形部分空間を構成するためには Schölkopf によって提案されたカーネル非線形主成分分析 [26] の技法を用いる．カーネル非線形主成分分析では非線形変換 ϕ が陽に定義されずカーネル関数だけが与えられるので，一般に非線形空間上での分布の主成分方向を求めることはできない．しかし，非線形変換後のパターンベクトルと非線形部分空間の各軸の方向ベクトルとの内積は以下のようにカーネル関数を用いて算出することができる．本節では，以上の事実を特異値分解を用いて示す．

任意の 3 組の d 次元パターン $\mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{y}_1, \dots, \mathbf{y}_m, \mathbf{z}_1, \dots, \mathbf{z}_l$ 及び関数 $\phi: \mathcal{R}^d \rightarrow \mathcal{R}^{d_\phi}$ を用いて，行列 $\mathbf{X}_\phi, \mathbf{Y}_\phi, \mathbf{Z}_\phi$ を

$$\mathbf{X}_\phi = (\phi(\mathbf{x}_1), \dots, \phi(\mathbf{x}_n)) \quad (\in \mathcal{R}^{d_\phi \times n}) \quad (7)$$

$$\mathbf{Y}_\phi = (\phi(\mathbf{y}_1), \dots, \phi(\mathbf{y}_m)) \quad (\in \mathcal{R}^{d_\phi \times m}) \quad (8)$$

$$\mathbf{Z}_\phi = (\phi(\mathbf{z}_1), \dots, \phi(\mathbf{z}_l)) \quad (\in \mathcal{R}^{d_\phi \times l}) \quad (9)$$

と定義する． $\phi(\mathbf{x}_i)$ ($i = 1, \dots, n$) の平均ベクトルを ϕ_M とおけば，

$$\phi_M = \frac{1}{n} \sum_{i=1}^n \phi(\mathbf{x}_i) = \frac{1}{n} \mathbf{X}_\phi \mathbf{1}_n \quad (10)$$

と書くことができる．ただし， $\mathbf{1}_n$ は全要素に 1 をもつ n 次元ベクトルである．関数 $\tilde{\phi}$ を

$$\tilde{\phi}: \mathbf{x} \mapsto \phi(\mathbf{x}) - \phi_M \quad (\mathcal{R}^d \rightarrow \mathcal{R}^{d_\phi}) \quad (11)$$

と定義すると $\mathbf{X}_{\tilde{\phi}}$ は式 (10) を用いて

$$\begin{aligned} \mathbf{X}_{\tilde{\phi}} &= (\phi(\mathbf{x}_1) - \phi_M, \dots, \phi(\mathbf{x}_n) - \phi_M) \\ &= \mathbf{X}_\phi - \phi_M \mathbf{1}_n^t \\ &= \mathbf{X}_\phi - \frac{1}{n} \mathbf{X}_\phi \mathbf{1}_n \mathbf{1}_n^t \quad (\in \mathcal{R}^{d_\phi \times n}) \end{aligned} \quad (12)$$

となる．ただし， $\mathbf{1}_{nn'}$ は全要素に 1 をもつ (n, n') 行列を表す．同様に，

$$\mathbf{Y}_{\tilde{\phi}} = \mathbf{Y}_\phi - \frac{1}{n} \mathbf{X}_\phi \mathbf{1}_{nm} \quad (\in \mathcal{R}^{d_\phi \times m}) \quad (13)$$

$$\mathbf{Z}_{\tilde{\phi}} = \mathbf{Z}_\phi - \frac{1}{n} \mathbf{X}_\phi \mathbf{1}_{nl} \quad (\in \mathcal{R}^{d_\phi \times l}) \quad (14)$$

となる．また任意の (d_ϕ, m) 行列 \mathbf{Y} ， (d_ϕ, l) 行列 \mathbf{Z} に対し， $K(\mathbf{Y}, \mathbf{Z})$ をその第 (i, j) 成分が $\phi(\mathbf{y}_i)^t \phi(\mathbf{z}_j)$ ($= k(\mathbf{y}_i, \mathbf{z}_j)$) となる (m, l) 行列，すなわち，

$$K(\mathbf{Y}, \mathbf{Z}) = \mathbf{Y}_\phi^t \mathbf{Z}_\phi \quad (15)$$

と定義する． K をカーネル行列と呼ぶ．更に， $G_X(\mathbf{Y}, \mathbf{Z})$ をその第 (i, j) 成分が $\tilde{\phi}(\mathbf{y}_i)^t \tilde{\phi}(\mathbf{z}_j)$ となる (m, l) 行列と定義すると，式 (13)，(14)，(15) より，

$$\begin{aligned} G_X(\mathbf{Y}, \mathbf{Z}) &= \mathbf{Y}_{\tilde{\phi}}^t \mathbf{Z}_{\tilde{\phi}} \\ &= K(\mathbf{Y}, \mathbf{Z}) - \frac{1}{n} K(\mathbf{Y}, \mathbf{X}) \mathbf{1}_{nl} - \frac{1}{n} \mathbf{1}_{mn} K(\mathbf{X}, \mathbf{Z}) \\ &\quad + \frac{1}{n^2} \mathbf{1}_{mn} K(\mathbf{X}, \mathbf{X}) \mathbf{1}_{nl} \end{aligned} \quad (16)$$

となる．

ここで，行列 $\mathbf{X}_{\tilde{\phi}}$ に対して特異値分解を行う． $\mathbf{X}_{\tilde{\phi}}^t \mathbf{X}_{\tilde{\phi}}$ の固有値を λ_i ($\lambda_1 \geq \dots \geq \lambda_n$) とおくと，その正の固有値は r ($= \text{rank}(\mathbf{X}_{\tilde{\phi}})$) 個存在し， $\mathbf{X}_{\tilde{\phi}} \mathbf{X}_{\tilde{\phi}}^t$ の正の固有値と一致する．固有値 λ_i に対応する $\mathbf{X}_{\tilde{\phi}}^t \mathbf{X}_{\tilde{\phi}}$ ， $\mathbf{X}_{\tilde{\phi}} \mathbf{X}_{\tilde{\phi}}^t$ の正規直交固有ベクトルをそれぞれ $\mathbf{u}_i, \mathbf{v}_i$ とし，行列 $\mathbf{U}, \mathbf{V}, \mathbf{A}$ を

$$\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_r) \quad (\in \mathcal{R}^{n \times r}) \quad (17)$$

$$\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_r) \quad (\in \mathcal{R}^{d_\phi \times r}) \quad (18)$$

$$\mathbf{A} = \begin{bmatrix} \sqrt{\lambda_1} & & & \\ & \ddots & & \\ & & 0 & \\ & & & \sqrt{\lambda_r} \end{bmatrix} \quad (\in \mathcal{R}^{r \times r}) \quad (19)$$

とおけば

$$\mathbf{X}_{\tilde{\phi}} = \mathbf{V}\mathbf{A}\mathbf{U}^t \quad (20)$$

である。 $\mathbf{X}_{\tilde{\phi}}\mathbf{X}_{\tilde{\phi}}^t$ はパターン集合 $(\phi(x_1), \dots, \phi(x_n))$ の共分散行列の n 倍であるから、 \mathbf{v}_i はパターン集合の i 番目の主成分方向を意味する。特異値分解の特性から

$$\mathbf{v}_i = \frac{1}{\sqrt{\lambda_i}}\mathbf{X}_{\tilde{\phi}}\mathbf{u}_i \quad (21)$$

が成り立つので、ベクトル $\tilde{\phi}(\mathbf{z})$ の \mathbf{v}_i への射影は

$$\begin{aligned} \mathbf{v}_i^t\tilde{\phi}(\mathbf{z}) &= \frac{1}{\sqrt{\lambda_i}}\mathbf{u}_i^t\mathbf{X}_{\tilde{\phi}}^t\tilde{\phi}(\mathbf{z}) \\ &= \frac{1}{\sqrt{\lambda_i}}\mathbf{u}_i^tG_X(\mathbf{X}, \mathbf{z}) \end{aligned} \quad (22)$$

となる。 $\tilde{\phi}(\mathbf{z}_j)$ の \mathbf{v}_i への射影は $\mathbf{V}^t\mathbf{Z}_{\tilde{\phi}}$ の (i, j) 成分であるから

$$\mathbf{V}^t\mathbf{Z}_{\tilde{\phi}} = \mathbf{A}^{-1}\mathbf{U}^tG_X(\mathbf{X}, \mathbf{Z}) \quad (23)$$

と表すことができる。一方、 \mathbf{A} と \mathbf{U} は行列 $\mathbf{X}_{\tilde{\phi}}^t\mathbf{X}_{\tilde{\phi}}$ 、すなわち $G_X(\mathbf{X}, \mathbf{X})$ の固有値と固有ベクトルとして求まる。

以上から、任意のパターン \mathbf{z} に対し、非線形部分空間の第 i 主成分方向 \mathbf{v}_i への $\tilde{\phi}(\mathbf{z})$ の射影は式 (15) で定義される K が与えられれば求まる。したがって、カーネル関数 k の定義式と学習パターンが与えられればよく、 ϕ の具体的な形を知る必要はない。

4.2 カーネル非線形部分空間法の識別規則

本論文で提案するカーネル非線形部分空間法 (KNS 法) は、部分空間の構成手段として前節で述べたカーネル非線形主成分分析の技術を使う。非線形クラス部分空間の構成法及び識別規則は、CLAFIC 法、投影距離法のどちらの方式でも採用することができる。それぞれの方式と組み合わせた KNS 法の識別規則が、カーネル関数を用いて記述できることを以下に示す。本節では、 \mathbf{X} をある一つのクラスに属する学習パターンからなるパターン行列、 \mathbf{z} を属するカテゴリーが未知のあるテストパターンとする。

まず、CLAFIC 法をベースとした KNS 法について述べる。 $K(\mathbf{X}, \mathbf{X}) = \mathbf{X}_{\tilde{\phi}}^t\mathbf{X}_{\tilde{\phi}}$ の上位 d' 個の固有値を λ'_i ($i = 1, \dots, d'$)、固有値 λ'_i に対応する $\mathbf{X}_{\tilde{\phi}}^t\mathbf{X}_{\tilde{\phi}}$ 、 $\mathbf{X}_{\tilde{\phi}}\mathbf{X}_{\tilde{\phi}}^t$ の正規直交固有ベクトルをそれぞれ \mathbf{u}'_i 、 \mathbf{v}'_i とし、 d' 個の \mathbf{u}'_i 、 \mathbf{v}'_i 、 λ'_i から式 (17)、(18)、(19) と同様に $\mathbf{U}'_{d'}$ 、 $\mathbf{V}'_{d'}$ 、 $\mathbf{A}'_{d'}$ を定義する。 $\mathbf{V}'_{d'} = (\mathbf{v}'_1, \dots, \mathbf{v}'_{d'})$

で張られる部分空間への $\phi(\mathbf{z})$ の射影ベクトルの長さを $P'(\mathbf{z})$ とすると、式 (15)、(21) より、

$$\begin{aligned} P'^2(\mathbf{z}) &= \sum_{i=1}^{d'} (\mathbf{v}'_i{}^t\tilde{\phi}(\mathbf{z}))^2 = \|\mathbf{V}'_{d'}{}^t\tilde{\phi}(\mathbf{z})\|^2 \\ &= \|(\mathbf{A}'_{d'})^{-1}\mathbf{U}'_{d'}{}^t\mathbf{X}_{\tilde{\phi}}^t\phi(\mathbf{z})\|^2 \\ &= \|(\mathbf{A}'_{d'})^{-1}\mathbf{U}'_{d'}{}^tK(\mathbf{X}, \mathbf{z})\|^2 \end{aligned} \quad (24)$$

となる。 d' はクラス部分空間の次元である。各クラスについて $P'^2(\mathbf{z})$ を求めその値が最小となるクラスに \mathbf{z} を識別すればよい。

次に、投影距離法をベースとした KNS 法について述べる。 $G_X(\mathbf{X}, \mathbf{X}) = \mathbf{X}_{\tilde{\phi}}^t\mathbf{X}_{\tilde{\phi}}$ の上位 d' 個の固有値を λ_i ($i = 1, \dots, d'$)、固有値 λ_i に対応する $\mathbf{X}_{\tilde{\phi}}^t\mathbf{X}_{\tilde{\phi}}$ 、 $\mathbf{X}_{\tilde{\phi}}\mathbf{X}_{\tilde{\phi}}^t$ の正規直交固有ベクトルをそれぞれ \mathbf{u}_i 、 \mathbf{v}_i とし、 d' 個の \mathbf{u}_i 、 \mathbf{v}_i 、 λ_i から上と同様に $\mathbf{U}_{d'}$ 、 $\mathbf{V}_{d'}$ 、 $\mathbf{A}_{d'}$ を定義する。あるパターン \mathbf{z} に対し、 $\phi(\mathbf{z})$ から ϕ_M までの距離、すなわちベクトル $\tilde{\phi}(\mathbf{z})$ の長さを $L(\mathbf{z})$ 、正規直交ベクトルの組 $\mathbf{V}_{d'} = (\mathbf{v}_1, \dots, \mathbf{v}_{d'})$ で張られる部分空間への $\tilde{\phi}(\mathbf{z})$ の射影ベクトルの長さを $P(\mathbf{z})$ とすると、投影距離の 2 乗 $D^2(\mathbf{z})$ は $L^2(\mathbf{z})$ と $P^2(\mathbf{z})$ の差であるから、式 (16)、(23) より、

$$\begin{aligned} D^2(\mathbf{z}) &= \tilde{\phi}^t(\mathbf{z})\tilde{\phi}(\mathbf{z}) - \sum_{i=1}^{d'} (\mathbf{v}_i^t\tilde{\phi}(\mathbf{z}))^2 \\ &= G_X(\mathbf{z}, \mathbf{z}) - \|\mathbf{A}_{d'}^{-1}\mathbf{U}_{d'}{}^tG_X(\mathbf{X}, \mathbf{z})\|^2 \end{aligned} \quad (25)$$

となる。したがって、各クラスについて $D^2(\mathbf{z})$ を求めその値が最小となるクラスに \mathbf{z} を識別すればよい。

前節で述べたように、 $\mathbf{U}'_{d'}$ 、 $\mathbf{A}'_{d'}$ は $K(\mathbf{X}, \mathbf{X})$ から求まり、 $\mathbf{U}_{d'}$ 、 $\mathbf{A}_{d'}$ は $G_X(\mathbf{X}, \mathbf{X})$ から求まる。したがって、上のどちらの方法においてもカーネル関数 $k(\mathbf{x}, \mathbf{y})$ の定義式と学習パターンが与えられれば ϕ の形を知らなくても未知パターンの識別が可能である。

また、線形部分空間法の場合と同様に、KNS 法においてクラス部分空間の次元 d' は識別結果に影響する。通常、 d' をクラスにかかわらず一定とするか、又は、累積寄与率 (cumulative proportion, c.p.) がクラスにかかわらず等しくなるように d' の値を定めることが多い。ただし、累積寄与率は固有値 λ_i ($\lambda_1 \geq \dots \geq \lambda_n$) を用いて

$$\text{c.p.} = \sum_{i=1}^{d'} \lambda_i / \sum_{i=1}^{d_\phi} \lambda_i \quad (26)$$

と定義される．最適な d' または c.p. の値は課題ごと異なる．この点については 5.4 で再度言及する．

5. 認識実験の結果と考察

カーネル非線形部分空間法 (KNS) の特性と効果を調べるため、最近傍決定則 (kNN), Support Vector Machine (SVM), 部分空間法 (LSS) などの従来手法と識別性能の比較を行った．実験対象として、第三者による追試及び比較が可能なデータのみを選び、公開手書き文字データから作成した 10 及び 64 次元の 48 クラスパターン、2 次元非線形分布をなす 2 クラス人工発生パターン、公開データベース Proben1 [27] のデータ、の 3 種類を用いた．これらを用いて、様々な条件における識別率、パラメータの変動に対する識別性能の安定性、学習及び識別における計算コストの観点から KNS 法の性能評価を行った．

各手法を表す略記号において、LSS, KNS では末尾に “-C” または “-P” を付けることにより CLAFIC 法と投影距離法のどちらを用いたかを表した．また、SVM, KNS におけるカーネルの種類 poly と rbf 及びそのパラメータ p は、式 (5), (6) で定義されるカーネル関数とそのパラメータをそれぞれ表す．例えば、KNS-P(rbf-10) は、式 (6) で定義されるカーネル関数を $p = 10$ として用いた投影距離法ベースの KNS 法であることを示す．更に、 d' はクラス部分空間の次元を、c.p. は累積寄与率を、 k は k -最近傍決定則における k の値を、 c は SVM の評価式におけるペナルティ係数を表すものとする．

5.1 多クラス分布に対する識別性能

クラスが十分多くかつ性能評価を行う上で十分なパターン数をもつ公開された特徴量データはほとんどない．そこで、次のような方法で多クラスパターンを生成した．まず、原データとして ETL-5 [28] に納められている手書き文字データのうち、片仮名 48 種 10,608 文字の 72×76 画素の画像パターンを用意した．次に、画像の左下を原点とみなしたときに (19, 14) の座標で与えられる画素を起点として 40×50 画素の範囲を切り出し、画素値の平均と分散による正規化を各画像ごとに行った．更に、これを 2,000 次元のベクトルとみなして全パターンに対し KL 展開を施し、64 次元及び 10 次元に圧縮した．最終的に各クラス当り 208 個の 64 次元パターン及び 10 次元パターンを得た．こうして得られたパターンは必ずしも文字認識に有効な特徴ではないが、第三者が容易に再現可能であること、特

表 1 64 次元多クラス分布に対する識別誤り率 (%)
Table 1 Classification error rates (%) against 64-dimensional multi-class patterns.

d: dimension			64	
# of classes			48	
# of patterns	c.p.	d' (mean \pm s.d.)	train	test
			4,800	4,800
kNN (k = 1)			0.00	13.92
kNN (k = 5)			10.65	15.65
LSS-P	0.96	(58.5 \pm 5.7)	0.25	6.58
LSS-C	0.97	(24.3 \pm 3.1)	0.29	7.23
KNS-P (rbf-100)	0.98	(43.6 \pm 4.4)	0.00	5.87
KNS-C (rbf-100)	0.995	(30.7 \pm 3.9)	0.04	5.90
KNS-P (poly-2)	0.99	(82.4 \pm 4.8)	0.00	6.92
KNS-C (poly-2)	0.998	(94.2 \pm 2.6)	0.00	6.98

徴空間上の有界な領域にパターンが分布していること、多クラスの分布であること、人工データと比較してパターン分布に恣意性が少ないなどの利点があり、識別系の評価用データとして適する．識別実験においては、各クラス当り 100 パターンを学習用に、別の 100 パターンをテスト用に用いた．

まず、64 次元パターンを用いて 48 クラスの識別を各手法で行ったときの識別誤り率 (%) を表 1 に示す．表中の各列の左側 (train) の数値は学習パターンに対する誤り率を表し、右側 (test) の数値はテストパターンに対する誤り率を表す．カーネル関数のパラメータ p 及び累積寄与率 c.p. は最適な結果が得られる場合を選択しており、 d' はその c.p. のときに得られたクラス部分空間の次元数の平均と標準偏差を表す．

KNS の識別率は LSS に比べてやや良いが、カーネル非線形変換の顕著な効果は見られない．一方、kNN の識別率は KNS, LSS と比べて極端に悪い．この結果は、多クラスの識別において LSS が kNN に比べ良い識別率を達成することが多いというこれまでの報告と一致する．これは、特徴空間の次元が高いと学習パターン数に対する識別率の収束が遅くなるという、kNN の欠点に起因すると考えられる．次に述べる 10 次元 48 クラスパターンの識別では LSS と kNN の結果が逆転していることもその傍証となる (表 2 参照)．更に、高次元多クラスパターンに対する kNN, LSS, KNS の識別率の比較に関して、本論文と同様の傾向を示す結果が報告されている [10] ．

次に、10 次元パターンを用いてクラス数を 10, 20, 48 と変化させ各手法について識別誤り率を求めた結果を表 2 に示す．表 1 と同様、パラメータの最適化を行った後の結果を示した．この表 2 の結果の中で表 1

表 2 10次元多クラス分布に対する識別誤り率(%)
Table 2 Classification error rates (%) against 10-dimensional multi-class patterns.

<i>d</i> : dimension		10			10			10		
# of classes		10			20			48		
# of patterns	c.p.	<i>d'</i> (mean ± s.d.)	train	test	<i>d'</i> (mean ± s.d.)	train	test	<i>d'</i> (mean ± s.d.)	train	test
			1,000	1,000		2,000	2,000		4,800	4,800
kNN (k = 1)			0.0	11.7		0.0	22.5		0.0	27.4
kNN (k = 5)			7.9	11.9		13.7	21.2		18.2	27.0
LSS-P	0.81	3.9 ± 0.5	14.0	16.8	3.9 ± 0.7	25.8	29.8	4.0 ± 0.7	36.1	39.5
LSS-C	0.85	3.5 ± 0.7	18.1	19.4	3.3 ± 0.8	31.9	33.9	3.4 ± 0.7	40.8	41.8
KNS-P (rbf-15)	0.96	49.9 ± 8.0	0.0	8.0	50.6 ± 8.9	0.0	16.2	49.9 ± 9.3	0.0	19.2
KNS-C (rbf-15)	0.95	41.8 ± 9.0	0.0	8.4	42.3 ± 9.5	0.1	16.3	41.4 ± 10.4	0.2	19.3
KNS-P (poly-2)	0.96	17.8 ± 2.4	2.1	8.7	18.4 ± 3.2	5.5	18.1	18.3 ± 2.8	7.4	20.2
KNS-C (poly-2)	0.98	19.9 ± 3.7	3.0	8.2	20.4 ± 4.1	6.1	17.6	20.5 ± 3.6	7.5	19.3

の結果と比較して大きく異なる点は、KNS法が他の手法と比較してはるかに良い性能を示している点、LSS法よりkNN法が良い性能を示している点である。

まず、10次元パターンに対するkNNの識別率がクラス数とともに増加しており、kNNにおける識別誤り率とベイズ誤り率との関係から[3]、これはベイズ誤り率の増加、すなわち、分布の重なりが増加しているためであると考えられる。更に、LSSとkNNの識別率を比べてみると、64次元パターンに対してはLSSの方が良いが10次元パターンではkNNの方が良いこと、また、LSSもクラス数の増加とともに誤り率が増加しているがクラス数が10から48へ増加したときの識別誤り率の増加量はkNNで約15~16%、LSSで約22~23%でありLSSの方が多いことがわかる。したがって、クラス数の増加に伴うLSSのこの識別誤り率の増加は、分布の重なりが増加だけでなく、3.2で述べた部分空間法の欠点であるクラス部分空間の重なりが増加に伴う識別率の低下を反映していると考えられる。

一方、KNS法ではクラス数の増加に伴う識別誤り率の増加が約11%にとどまっている。64次元パターンの結果においてLSSとKNSの差が小さいことからこのパターン分布の非線形性は小さいと予想されるので、10次元パターンの識別においてKNS法の性能が非常に優れているのは、カーネル非線形変換による高次元への写像がクラス部分空間の重なりが増加という部分空間法の欠点を補完したためであると推察される。

以上の議論は、テストパターンに対する識別率の比較に関するものであったが、学習パターンに対する識別誤り率の低さもKNS法の大きな特徴である。通常、学習パターンに対する過度の最適化は過学習と呼ばれ、テストパターンに対する汎化能力の低下を招く。とこ

ろがKNS法の特徴は、パラメータ最適化後において学習パターンに対する識別誤り率がほぼ0に近いにもかかわらず、テストパターンに対する識別率は他の手法と比較して最良であるという点である。すなわち、より多くの学習パターンがあれば更に識別能力が上昇することを示唆している。この点については次節においてより詳しく論ずる。

5.2 2クラス非線形分布に対する識別性能

人工パターンを用いた識別系の性能評価は、人工パターンの生成法に強い恣意性が入る点や非現実的なパターン分布を発生させているという点から批判を受けることもある。しかしながら、どんなパターン分布であってもその潜在的な存在可能性は否定できない、という意味において後者の批判は必ずしも妥当なものではないばかりか、以下に述べる理由からむしろ人工パターンを用いた性能評価には大きな意味がある。

第1に、あるパターン集合に対する識別機の性能評価は、母集団に属するパターンすべてを用いた識別機の最適化によって初めて可能となる。しかし、通常は限られたパターンを用いて識別機を最適化しなければならない。そして、その最適化は学習パターン数に依存し、しかもその依存の仕方は識別機によって異なる。したがって、識別機の評価のためには学習パターン数を変化させて、その識別性能の収束傾向を比較しなければならない。こうした評価実験は十分な数のパターンが発生可能な人工パターンにおいて可能である。第2に、学習及びテストに使われるパターンは母集団からサンプリングされたものであるから、そこから推定される識別率はそのパターン数に依存したばらつきをもつ。従って、複数回の独立試行を繰り返してその平均と分散を推定することによって、異なる手法間の識別率の有意差を論ずることが必要となる。このために

は、独立試行の繰返しが容易な人工発生パターンが有用である。第 3 に、人工発生パターンを用いた識別実験は、第三者による各種識別手法の追試、比較が容易である。

前節では、KNS 法が多クラスの識別に対し有効であり、それが高次元空間への写像という KNS 法の特徴によることを示した。クラス数に比べ特徴空間の次元が高い場合には、KNS 法と LSS 法との識別性能の大きな差は観察されず、それは恐らく実験データとして扱ったパターン分布の非線形性が小さいためであると予想された。そこで、KNS 法の期待されるもう一つの効果である、非線形分布に対する識別性能の向上を検証するため、非線形性の非常に強い 2 クラスパターンを用いて識別実験を行った。また、識別境界の記述能力を手法間で比較するために、クラス分布の重なりが存在し、かつ複雑な境界をもつ人工パターンを実験データとした。

x_1, y_1, x_2, y_2 を平均と分散 (μ, σ^2) がそれぞれ $(0, 10), (10, 5), (3, 10), (20, 5)$ の正規分布からランダムにサンプリングされた値とし、クラス 1 (\mathcal{X}_1), クラス 2 (\mathcal{X}_2) に属する 2 次元パターンをそれぞれ

$$(x_1, 0.5x_1^2 + y_1) \in \mathcal{X}_1 \quad (27)$$

$$(x_2, -0.5x_2^2 + y_2) \in \mathcal{X}_2 \quad (28)$$

として各クラス同数生成した。

300 個の学習パターンの例と、その学習パターンから各手法を用いて識別系を構成して得られた識別境界とを図 1 に示す。2 クラスのパターンを赤丸と青丸によって示し、識別境界は緑線で示した。また、SVM において得られたサポートベクトルは誤識別されたパターンすべてと正しく識別されたパターンの内図中で水色の丸で囲んだパターンである。図に示したように、生成されたパターンは分布の形状が非線形であり、かつ最適なクラス境界が複雑な曲線をなしており、識別が非常に難しい例である。LSS では有効な識別境界が形成されていないが、他の手法ではほぼ最適な識別が可能な識別境界が得られている。

次に、テストパターン数を 100 に固定し、学習パターン数を 10 から 300 まで変化させて識別実験を行った。各実験条件に対して独立に 100 回の試行を繰り返し、識別誤り率の平均と分散を算出した。識別誤り率の平均値を各学習パターン数、各手法ごとに示したのが図 2 である。図左上に示したように、同じ色の 2 本の折れ線は各手法の学習パターンに対する識別誤り率

の平均 (実線) とテストパターンに対する識別誤り率の平均 (点線) を表す。A と B という二つの手法の識別性能の比較は、学習パターンが無制限の場合に推定される潜在的な最適識別率とその最適識別率への漸近性という二つの観点からなされなければならない。更に、最適識別率の推定は、学習パターンに対する識別率 (R 法) とテストパターンに対する識別率 (H 法) によって狭み撃ちにすることによって可能となる。

図 1 で示したような複雑な境界を有するパターン分布に対し、SVM は一般に高い識別能力を有する。SVM(poly-1), SVM(poly-2) では識別境界の記述能力が十分でなく学習パターン数が増えても識別率は向上しないが、SVM(poly-3) 及び SVM(rbf-2) では学習パターン数 300 において他の手法と比較して最も高い識別率であった。LSS では学習パターン数にかかわらず低い識別率であったが、KNS は SVM に準ずる識別率を達成した。KNS(rbf-1), SVM(poly-3), kNN(k=3) の学習パターン 300 でのテストパターンに対する識別率は、対応 2 試料の検定により有意差が認められた。

また、KNS, kNN の学習パターンに対する識別率は SVM のそれよりもはるかに良く、したがって、学習パターン無限大でのテストパターンに対する識別率は SVM のそれを上回る可能性がある。更に学習パターンの増加に対する識別誤り率の収束性は SVM が最も早く、kNN が最も遅い。更に、学習パターンの増大に対する識別誤り率の収束は、KNS-P, KNS-C にかかわらず KNS(poly) の方が KNS(rbf) よりが早い。KNS-P と KNS-C の間に識別性能、収束特性について大きな差は見られない [8]。

5.3 公開データを用いた識別性能の比較

現実世界の中で実際に採取された特徴データの中で性能比較用に公開されているものに Proben1 [27] がある。その中の何種類かを用いて KNS 法と従来手法との比較を行った。表 3 に 3 種類の soybean データを用いた識別実験の結果を示す。soybean データは soybean の状況から病気の種類を識別するためのデータセットであり、どのセットもクラス数 19, 特徴ベクトルの次元 82, パターン数 683 である。識別率の評価は [27] に記述されている方法に従った。指定された 170 パターンをテストパターン、残りを学習パターンとし、テストパターンを変えて 4 回の試行を繰り返して学習パターン及びテストパターンに対する識別誤り率の平均と標準偏差を示した。全般的に KNS 法が

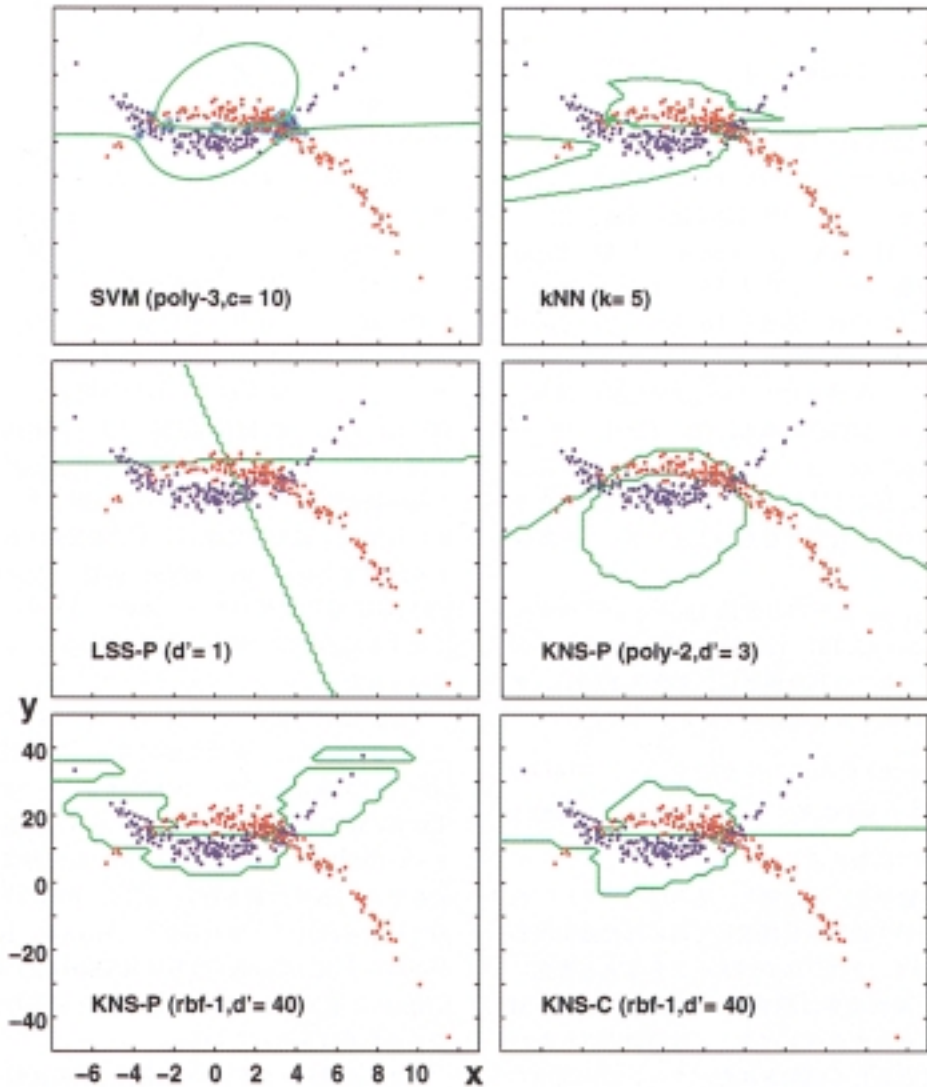


図1 2次元2クラス人工パターンの例と各手法で得られた識別境界
 Fig. 1 An example of 2-dimensional 2-class artificial patterns and classification borders optimized by various classification methods.

やや良いが統計的な有意差はない。これは1セットのデータからただか4回の試行しかできないことにも起因する。Proben1の他のデータセットのほとんどはクラス数がただか2か3であるが、すべてにおいてKNS法はkNN法やSVMとほぼ同等の識別性能を示した。

5.4 パラメータの変動に対する識別性能の安定性

KNS法を適用する上で解決しなければならない問題は、部分空間の次元 d' 、カーネル関数のクラス、カー

ネル関数のパラメータ p の最適化である。そこで、 p 、 d' に対してKNS法がどのような振舞いを示すかについて調べた。5.1で述べた10クラス10次元パターンを用い、学習及びテストパターン数は各50とした。

図3は部分空間の次元 d' が識別に与える影響を調べた結果である。LSS法、KNS(poly)法、KNS(rbf)法の三つを比較した。同じ種類の2本の折れ線のうち下方に位置しているのが学習パターンに対する誤り率、上方に位置するのがテストパターンに対する誤り率で

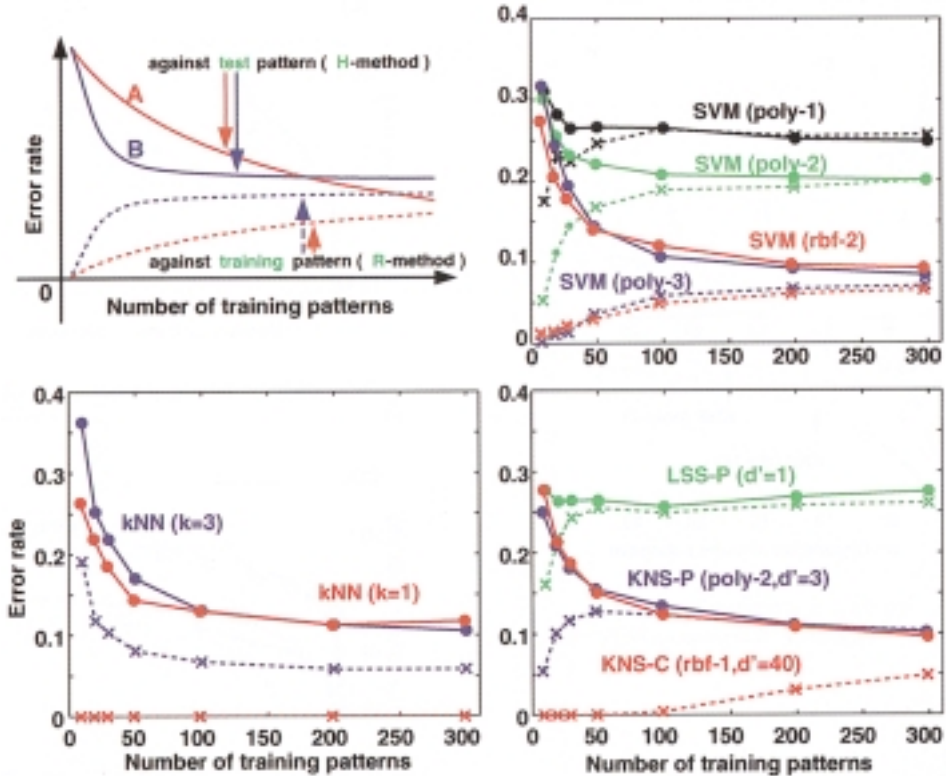


図2 識別性能の評価法を示した模式図(左上)とSVM, 最近傍決定則(kNN), 線形部分空間法(LSS), カーネル非線形部分空間法(KNS)の2次元2クラス非線形パターンに対する識別誤り率の比較

Fig. 2 Scheme for the evaluation of classification performance (upper left) and the comparison of error rates against 2-dimensional 2-class nonlinear patterns by SVM, k-nearest neighbor rule (kNN), linear subspace method (LSS) and kernel based nonlinear subspace method (KNS).

表3 Proben1のsoybeanデータセットに対する識別誤り率(%)
Table 3 Classification error rates (%) against soybean data sets in Proben1.

Classifier	Data set	soybean1		soybean2		soybean3	
		training	test	training	test	training	test
kNN	k = 3	7.38 ± 0.87	8.52 ± 1.48	7.53 ± 1.46	10.29 ± 2.28	7.82 ± 0.97	9.56 ± 1.47
LSS-P	$d' = 5$	7.53 ± 1.46	7.65 ± 3.81	6.94 ± 2.56	9.26 ± 1.47	7.53 ± 1.46	9.26 ± 1.47
KNS-P(rbf-2)	$d' = 5$	6.87 ± 0.84	6.32 ± 3.01	7.01 ± 2.42	9.56 ± 1.47	7.60 ± 1.33	9.41 ± 1.44
KNS-P(poly-2)	$d' = 5$	8.11 ± 0.65	7.64 ± 3.81	7.09 ± 2.28	9.71 ± 1.56	7.31 ± 1.86	8.97 ± 1.69

ある．部分空間法に関するこれまでの研究からよく知られているように識別誤り率は d' に強く依存する．図に示した結果はLSSとKNSで d' に対する識別誤り率の振舞いが大きく異なることを示している．LSS法では最適若しくは準最適な d' の範囲が狭いのに対し，KNS法の識別誤り率は d' に対してLSSのそれより緩やかなこう配をもっている．LSS法，KNS法では d' が特徴空間の次元 d ，変換後の次元 d_ϕ にそれぞれ

近くなるとクラス部分空間の重なりが増えるため識別性能が低下する．しかし，KNS法では高次元空間上でクラス部分空間を構成しているため d_ϕ が非常に大きく d' の変動の影響を受け難い．したがって，KNS法は最適若しくは準最適な d' の範囲が広く d' の選択が容易であり，これはKNS法の実用性の高さを表している．

図4は，カーネル関数のパラメータ p の識別に与え

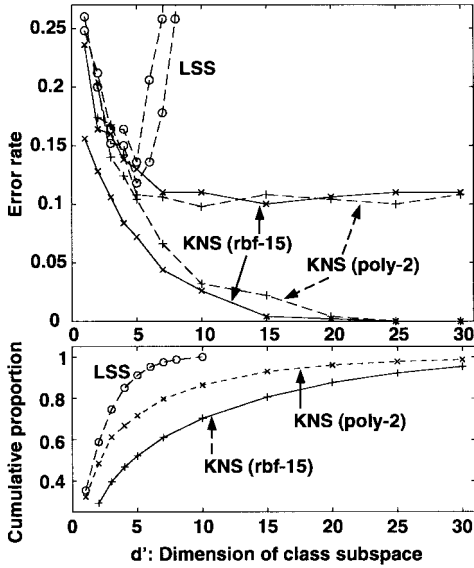


図3 部分空間の次元の識別率に与える効果
Fig.3 Effect of subspace dimension on error rate.

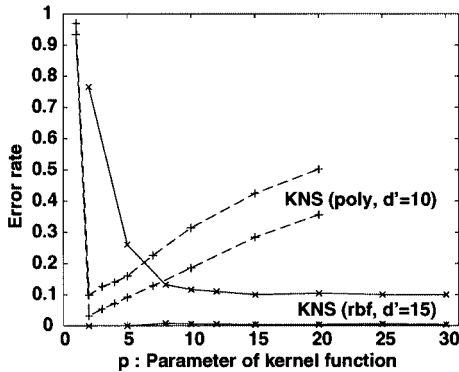


図4 カーネル関数のパラメータ p の識別率に与える効果
Fig.4 Effect of parameter values of kernel function to classification error.

る効果を調べた結果である．同じ種類の2本の折れ線の意味は前図と同じである．KNS(poly)とKNS(rbf)のそれぞれについて様々な p での識別誤り率を求めた．KNS(poly)では p の比較的小さいところに最適値があり、しかも p の変動に敏感である．KNS(poly)の p は整数値しかとることができないので p の選択には慎重を期す必要がある．一方KNS(rbf)ではより大きな p に最適値があり最適値の周辺での変動は緩やかであった．更に p を大きくしていくとKNS(rbf)の誤り

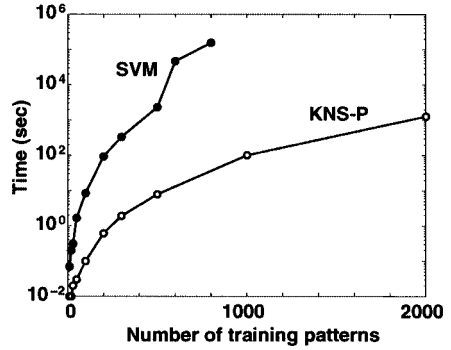


図5 識別系の学習に必要な計算時間の比較
Fig.5 Computation time for classifier training.

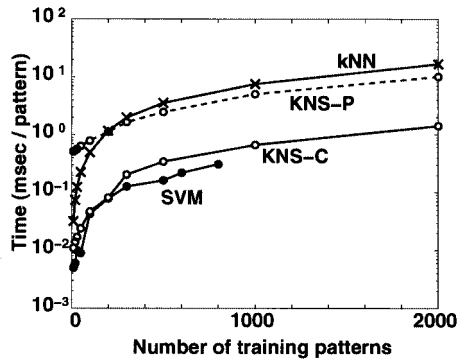


図6 識別に必要な計算時間の比較
Fig.6 Computation time for classification.

率は再び上昇する．こうしたKNS(poly), KNS(rbf)の p に対する振舞いは、今回試みた他のデータに対してもほぼ同様の傾向を示した．

5.5 計算コストの比較

識別手法の学習に要する計算時間及び未知パターンの識別に要する計算時間という観点から各識別手法の比較を行った．測定にはSUN ULTRA 60上のMATLABを用い、パターンは5.2で述べた2クラス2次元人工パターンを利用した．学習パターンを10から2,000(SVMは10から800)まで変化させて学習を行い、要した時間を計測した結果を図5に示す．学習終了後、改めて発生させた10,000個のテストパターンの識別に要した時間を計測し平均した結果を図6に示す．学習パターン及びテストパターンはクラスあたり同数とした．比較した手法及びそのパラメータは、SVM(poly-3), KNS-P(poly-2, $d' = 2$), KNS-C(rbf-15, $d' = 2$), kNN($k = 1$)である．

KNS の学習に要する時間は、SVM と比較して学習パターン数 100 のときで約 1/100, 学習パターン数 1,000 のときで約 1/1,000 であり, 学習パターンが増えるに伴いその比は小さくなる (図 5). この実験条件では, SVM の場合学習パターンが 1,000 を超えると学習に非実用的な時間を要する. また, SVM では学習に要する時間の 99% 以上を 2 次計画法を解くために要していること, パターンの分布によって 2 次計画法の解法に要する時間は大きく変動する可能性があることから, SVM では学習パターン数が同一であっても課題によって学習に要する時間が大きく変動する. 一方, KNS では学習に要する時間はほぼ学習パターン数だけで決まるので SVM と比較して所要時間の推定が容易であり, この点においても実用上使いやすい. KNS-P と KNS-C では学習に要する時間に大きな差はなかった. kNN ではメモリ上に学習パターンを配置するだけであるので, 学習に要する時間は実質上 0 である.

一方, 識別に要する時間は, 四つの方法の中で SVM が最も短く, kNN の数十分の一であった (図 6). SVM の識別時間は学習の結果得られるサポートベクトルの個数に依存する. KNS-P と KNS-C では約 10 倍弱の差があった. これは, 式 (15) の K を計算する手間と, 式 (16) の G_X を計算する手間との違いによる. したがって, 高速な識別を要するアプリケーションに KNS 法を使う場合には, KNS-P と KNS-C のどちらを用いるかの選択には注意が必要である.

ここで示した学習及び識別に要する計算時間の測定は, アルゴリズムやプログラムの最適化の程度, パラメータ等の実験条件によって変化する可能性は残る. しかしながら, 現在, 学習に要する計算時間の破綻^{たん}が大きな問題とされている SVM に比べ KNS 法は数オーダー早い時間で学習が可能であり, また, 識別においても標準的な方法として非常によく使われている kNN と比較して KNS は高速であるから, KNS 法は計算コストの観点からも非常に実用性の高い識別手法であると考えられる.

6. む す び

本論文では, 新しいパターン認識手法としてカーネル非線形部分空間法 (KNS 法) を提案した. 本手法は, 多クラス非線形分布の識別に有効な手法であり, 多クラスの識別問題には従来の部分空間法や最近傍決定則よりも良い性能を示した. また, 2 クラスの識別

問題にも十分有効な方法であり, 最近傍決定則より優れ, Support Vector Machine に近い識別能力を示した. したがって, KNS 法は非線形かつ多クラスという通常の方法では困難な課題に有効だけでなく, 通常の方法でも解決可能な識別課題に対しても有効な方法である. 更に, 識別能力の優秀性だけでなく, 計算コスト, 学習, 識別アルゴリズムのインプリメントの容易さ, パラメータの最適化の容易さなどから KNS 法は極めて使いやすく便利な識別手法といえる.

また, KNS 法では SVM のように非線形最適化の必要がなくなかつカーネル関数を介した定式化が可能であるので, 少ない計算量で多次元性の利点を生かすことができる. カーネル関数を利用した多次元非線形空間での操作は, 求めたい量がカーネル行列 K で記述できれば他のアルゴリズムにも適用可能である. KNS 法は SVM とともにカーネル非線形変換の特徴を生かすことのできた例であるといえよう.

これまでの実験の結果得られた KNS 法の識別能力は, 実は, 筆者にとっても当初の予想を超えるものであった. 現在, KNS 法が不得手とするような典型的なパターン分布というのが把握できておらず, 今後の研究課題として残されている. “KNS 法が本当にそんなに良い方法なのか” という問いに答えるためには, 今後より多くのデータによる評価が必要となる. また, 本論文で報告した高い識別性能が引き出されている裏にはその根拠となる本質的理由があるはずであり, その点に関する理論的なアプローチも今後の課題である. 更に, 部分空間法については複合類似度法や学習部分空間法といった改良手法が提案されており, また SVM についてもカーネル関数の改良などに関する研究がある. これら先行研究の成果を KNS 法に組み入れていくことも今後の検討課題の一つである.

謝辞 本研究を進めるにあたり貴重な御助言をいただいた NTT コミュニケーション科学基礎研究所石井健一郎部長, 白木善尚主任研究員, 上田修功主幹研究員に感謝致します.

文 献

- [1] N.J. Nilsson, “Learning Machines,” McGraw-Hill, 1965. (渡邊 茂訳, “学習機械” コロナ社, 1967)
- [2] O.A. Bashkurov, E.M. Braverman, and I.B. Muchnik, “Potential function algorithms for pattern recognition learning machines,” Automation and Remote Control, vol.25, pp.629–631, 1964.
- [3] T.M. Cover and P.E. Hart, “Nearest neighbor pattern classification,” IEEE Trans. Inf. Theory, vol.IT-13,

- no.1, pp.21-27, 1967.
- [4] D.E. Rumelhart, G.E. Hinton, and R.J. Williams, "Learning representations by back-propagating errors," *Nature*, vol.323, no.9, pp.533-536, 1986.
- [5] V. Vapnik, "Statistical Learning Theory," John Wiley & Sons, Inc., 1998.
- [6] C. Burges, "A tutorial on Support Vector Machines for pattern recognition," *Data Mining and Knowledge Discovery*, Kluwer, pp.1-43, 1998.
- [7] 前田英作, 村瀬 洋, "Support Vector Machine とカーネル非線形部分空間法," *信学技報*, PRMU98-81, 1998.
- [8] E. Maeda and H. Murase, "Kernel based nonlinear subspace method for multi-category classification," *Technical Report of Information Science Research Laboratory, NTT Basic Research Laboratories*, ISRL-98-1, 1998.
- [9] E. Maeda and H. Murase, "Multi-category classification by kernel based nonlinear subspace method," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1999.
- [10] 津田宏治, "ヒルベルト空間における部分空間法," *信学技報*, NC98-36, 1998.
- [11] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Statist.*, vol.33, pp.1065-1076, 1962.
- [12] M.A. Aizerman, E.M. Braverman, and L.I. Rozonoer, "Theoretical foundations of the potential function method in pattern recognition learning," *Automation and Remote Control*, vol.25, pp.821-837, 1964.
- [13] M.A. Aizerman, E.M. Braverman, and L.I. Rozonoer, "Method of Potential Functions in the Theory of Pattern Recognition (in Russian)," *Nauka*, 1970. (北川敏男, 林 順雄訳, *パターン認識と学習制御*, 共立出版, 1978)
- [14] M.J.D. Powell, "Radial basis functions for multivariable interpolation, a review," *Algorithms for Approximation*, eds. J.C. Mason and M.G. Cox, Clarendon Press, Oxford, 1987.
- [15] R. Courant and D. Hilbert, "Methods of Mathematical Physics," Wiley, 1953. (齊藤利弥, 丸山滋弥訳, *数理物理学の方法*, 東京図書, 1959)
- [16] J.T. Tou and R.C. Gonzalez, "Pattern Recognition Principles," Addison-Wesley, 1974.
- [17] R.O. Duda and P.E. Hart, "Pattern Classification and Scene Analysis," John Wiley & Sons, Inc., 1973.
- [18] K. Bennett and O.L. Mangasarian, "Multicategory discrimination via. linear programming," *Optimization Methods and Software*, vol.3, pp.27-39, 1994.
- [19] J. Weston and C. Watkins, "Multi-class support vector machines," *Technical Report CSD-TR-98-04*, Royal Holloway University of London, 1998.
- [20] 下平 博, 佐藤弘一, M. Vlach, "Support Vector Machines による複数カテゴリの識別," *信学技報*, PRMU98-36, 1998.
- [21] S. Watanabe, "Knowing & Guessing—Quantitative Study of Inference and Information," John Wiley & Sons, Inc., 1969. (村上陽一郎, 丹治信春訳, *知識と推測*, 東京図書, 1975)
- [22] 飯島泰蔵, "パターン認識理論," 森北出版, 1989.
- [23] E. Oja, "Subspace Methods of Pattern Recognition," *Research Studies Press Ltd.*, 1983. (小川英光, 佐藤誠訳, *パターン認識と部分空間法*, 産業図書, 1986)
- [24] 池田正幸, 田中英彦, 元岡 達, "手書き文字認識における投影距離法," *情処学論*, vol.24, pp.106-112, 1983.
- [25] 井上 聡, 若林哲史, 鶴岡信治, 木村文隆, 三宅康二, "競合自己想起回路による手書き数字認識," *信学技報*, PRMU96-207, 1997.
- [26] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol.10, pp.1299-1319, 1998.
- [27] L. Prechelt, "Proben1—A set of neural network benchmark problems and benchmarking rules," *Technical Report 21/94*, Fakultät für Informatik, Universität Karlsruhe, 1994.
- [28] The Electrotechnical Laboratory, "ETL character database," *ICDAR'93*, vol.B, ETL-5, 1993.
(平成 10 年 7 月 28 日受付, 11 月 17 日再受付)

前田 英作 (正員)



1984 東大・理・生物卒・1986 同大大学院修士課程了。同年日本電信電話(株)入社。以来,パターン認識及び神経生理学の研究に従事。1995~1996 英国 Cambridge 大学客員研究員。現在,NTTコミュニケーション科学基礎研究所研究総括担当課長・工博。1994 日本神経回路学会奨励賞受賞。IEEE, 日本神経回路学会, 日本神経科学会, Society for Neuroscience 各会員。

村瀬 洋 (正員)



1978 名大・工・電子卒。1980 同大大学院修士課程了。同年日本電信電話公社(現 NTT)入社。以来,文字・図形認識,コンピュータビジョン,マルチメディア認識の研究に従事。1992~1993 米国 Columbia 大客員研究員。現在,NTTコミュニケーション科学基礎研究所メディア情報研究部グループリーダー・工博。1985 本学会学術奨励賞,1992 電気通信普及財団テレコムシステム技術賞,1994 IEEE-CVPR 最優秀論文賞,1995 情報処理学会山下記念研究賞,1996 IEEE-ICRA 最優秀ビデオ賞受賞。IEEE, 情報処理学会各会員。