

カーネル標本空間射影法によるパターン認識

Kernel Sample Space Projection Method for Pattern Recognition

鷲沢 嘉一*

Yoshikazu WASHIZAWA

山下 幸彦†

Yukihiko YAMASHITA

Abstract: We propose a new kernel based method for pattern recognition. Support vector machine, PCA, and Fisher discriminant are extended to kernel based methods and they achieve better performance. Sample vectors for each category mapped to a high dimensional feature space span a non-dense subspace. Thus, an unknown input vector mapped to the feature space can be discriminated by comparing the norms of vectors projected onto the subspaces.

Different from the kernel PCA which also uses subspaces in the feature space, our method does not need the calculation of an eigenvalue problem. Then, the computational complexity for learning of our method is lower than those of other methods.

In this paper, we provide a closed form of our method and show experimental results of the handwritten digit database “MNIST” and compare it with other methods from several points of view.

キーワード: パターン認識, カーネル法, カーネル非線型部分空間 (KNS) 法, サポートベクトルマシン, カーネル標本空間射影法

1 はじめに

パターン認識において、入力を高次元空間へ写像してから識別を行う“カーネル法”とよばれる手法が、Support Vector Machine (SVM) [1] や Kernel PCA (KPCA) [2] を用いたカーネル部分空間法 [3],[4](以降、KPCA 法と表記する) などにおいて広く使われており、高い識別精度を示している。

しかしながら、これらの手法は学習サンプル数の増加にとともない、識別器を設計するための計算量が非常に多くなってしまいう問題がある。SVM では 2 次計画問題、KPCA 法では固有値問題の大きさが学習サンプル数となる。

一般に学習サンプルの数は、入力次元に比べ大きいほど識別器の性能が上がる。したがって、学習サンプルを得るためのコストが小さいアプリケーションや入力次元が大きいアプリケーションの場合は、なるべく多くのサンプルを用いて識別器を学習させた方がよい。

しかし、前述の通りカーネル法は計算コストが学習サ

ンプル数に大きく関係する。したがって、学習サンプルが多すぎる場合、限られた資源で識別器を設計するためには、その中から、学習サンプルを選びだすという作業が必要となる。したがって、カーネル法による識別器を少ない計算コストで実現することは重要な課題であり、SVM では様々な 2 次計画問題を解く高速アルゴリズムが提案されている [5],[6]。

一方、KPCA 法は SVM に比べ、多クラス識別に適している、閉じた形で解が求まる、学習サンプルのエラーに対して頑健である、などの利点がある。

本論文では設計時の計算コストが少ないカーネル法による識別手法であるカーネル標本空間射影 (KSP: Kernel Sample Space Projection) 法を提案する。この手法は、KPCA 法と同様、多クラス識別に適しており、閉じた形で解が求まる。また、追加学習が簡単に行えるといった特徴がある。さらに、手書き数字認識実験によりその性能を評価し、従来法との比較を行う。

2 カーネル法

本節で、本手法に用いるカーネル法について詳しく述べる。

カーネル法とは、入力ベクトルを特徴空間とよばれる高次元の空間へ非線形写像し、識別を行う手法である。特徴空間での内積の演算は Mercer カーネルとよばれる

*東京工業大学 理工学研究科, 〒 152-8552 東京都目黒区大岡山 2-12-1, tel. 03-5734-3906, e-mail washizawa@ide.titech.ac.jp, Tokyo Institute of Technology, 2-12-1, Ookayama, Meguro-ku, Tokyo, 152-8552, Japan.

†東京工業大学 理工学研究科, 〒 152-8552 東京都目黒区大岡山 2-12-1, tel. 03-5734-3497, e-mail yamasita@ide.titech.ac.jp, Tokyo Institute of Technology, 2-12-1, Ookayama, Meguro-ku, Tokyo, 152-8552, Japan.

関数を使って簡単に計算することができる。したがって、写像 Φ の形を陽に知ることなく高次元での計算が可能になる。

定理 1. Mercer. [7]

入力空間を X とし、 (X, μ) を測度有限な測度空間とする。 $L_\infty(X)$ を X 上で定義される有界な可測関数全体の集合とし、 $L_2(X)$ を X 上で定義される 2 乗可積分な可測関数全体の集合とする。 $k \in L_\infty(X^2)$ を対称なカーネル関数とし、以下の線形変換 $T_k : L_2(X) \rightarrow L_2(X)$ が半正定値作用素であるとする。

$$T_k f(\cdot) = \int_X k(\cdot, y) f(y) d\mu(y) \quad (1)$$

このとき、 $\psi_j \in L_2(X)$ を T_k の固有値 $\lambda_j \neq 0$ に対応する正規化固有関数とすると、

1. $(\lambda_1, \lambda_2, \dots) \in l_1$
2. $\psi_j \in L_\infty(X)$ かつ $\sup_j \|\psi_j\|_{L_\infty} < \infty$
3. $k(x, y) = \sum_{j \in \mathbb{N}} \lambda_j \psi_j(x) \psi_j(y)$ がほとんどすべての (x, y) について成り立つ。ここでこの和はほとんどすべての (x, y) について絶対一様収束する。

が成り立つ。ここで、 l_1 は絶対和が有界な数列全体の集合である。□

式 (1) を満たす 2 変数関数 k を Mercer カーネルと呼ぶ。任意の Mercer カーネルは、それ自身を再生核とする再生核ヒルベルト空間を持つことが示されている。

定義 1. 再生核ヒルベルト空間 Reproducing kernel hilbert space [8], [9]

集合 D 上の関数ヒルベルト空間 \mathcal{H} において

1. 任意の $x \in D$ に対して $k(x, \cdot) \in \mathcal{H}$ となる。
2. \mathcal{H} での内積を $\langle \cdot, \cdot \rangle$ で表す。任意の $f \in \mathcal{H}$, $x' \in D$ に対して

$$\langle f(\cdot), k(\cdot, x') \rangle = f(x') \quad (2)$$

を満たす $D \times D$ 関数 $k(x, x')$ が存在するとき、関数ヒルベルト空間 \mathcal{H} を再生核ヒルベルト空間 (RKHS: Reproducing kernel hilbert space) といひ、関数 $k(x, x')$ を再生核 (Reproducing kernel) という。□

命題 1. 任意の Mercer カーネル k に対して、 k が再生核となるような再生核ヒルベルト空間 \mathcal{H} が存在する [10]. □

したがって、入力空間 \mathbb{R}^N から特徴空間 \mathcal{F} への写像 Φ を

$$\Phi : \mathbb{R}^N \rightarrow \mathcal{F}, \quad x \mapsto k(\cdot, x) \quad (3)$$

によって定義すれば、

$$\begin{aligned} \langle \Phi(x), \Phi(y) \rangle &= \langle k(\cdot, x), k(\cdot, y) \rangle \\ &= k(x, y) \end{aligned} \quad (4)$$

となり、特徴空間 \mathcal{F} への写像後の内積が原空間での関数として計算が可能となる。式 (1) を満たす Mercer カーネルとして以下のものが一般的に用いられている。

$$k(x, y) = \langle x, y \rangle \quad : \quad \text{Identical kernel} \quad (5)$$

$$k(x, y) = (\langle x, y \rangle + 1)^d \quad : \quad \text{Polynomial kernel} \quad (6)$$

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right) \quad : \quad \text{Gaussian kernel} \quad (7)$$

カーネル関数 k が式 (5) の場合、写像 Φ は恒等写像であり、式 (6) の多項式カーネルでは、入力ベクトル $f \in \mathbb{R}^N$ は、 N 次元空間から $N+dC_d - 1$ 次元空間へ写像される。式 (7) の場合は、入力ベクトルは無限次元の空間へ写像される。

3 カーネル標本空間射影法

カテゴリ i のサンプル数を L_i とし、カテゴリ i に属する入力パターンを $f_i \in \Omega_i$ とする。写像 Φ によって写像される特徴空間 \mathcal{F} の次元が L_i に比べて十分に大きいとき、特徴空間 \mathcal{F} に写像されたカテゴリ i に属する入力ベクトル $\Phi(f_i)$ は、特徴空間 \mathcal{F} に比べ疎な部分空間を張る。そして、その部分空間は、カテゴリごとに大きく異なり、偏った部分空間になると考えられる。

カーネル標本空間射影 (KSP: Kernel Sample Space Projection) 法は、未知の入力 x をこの偏った部分空間へ射影し、そのノルムをカテゴリとの類似度とし、ノルムが最大となるクラスへ決定する。

カテゴリ i に属するパターン $f_i^k (k = 1, 2, \dots, L_i)$ が特徴空間で張る空間をカーネル標本空間 (kernel sample space) と呼ぶ。作用素 S_i の以下のように定義する。

$$S_i = [\Phi(f_i^1) \quad \Phi(f_i^2) \quad \dots \quad \Phi(f_i^{L_i})] \quad (8)$$

S_i の値域 $R(S_i)$ が、カーネル標本空間となる。このときカーネルグラム行列 K_i を以下のように定義する。

$$K_i = S_i^* S_i = \begin{bmatrix} k(f_i^1, f_i^1) & \dots & k(f_i^1, f_i^{L_i}) \\ \vdots & \ddots & \vdots \\ k(f_i^{L_i}, f_i^1) & \dots & k(f_i^{L_i}, f_i^{L_i}) \end{bmatrix} \quad (9)$$

ここで $A^* \in \mathcal{B}(\mathcal{H}_2, \mathcal{H}_1)$ は以下で定義される $A \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ の共役作用素を表す .

$$\langle Af, g \rangle = \langle f, A^*g \rangle \quad \forall f \in \mathcal{H}_1, \forall g \in \mathcal{H}_2 \quad (10)$$

A に対し ,

$$ABA = A \quad (11)$$

$$BAB = B \quad (12)$$

$$(AB)^* = AB \quad (13)$$

$$(BA)^* = BA \quad (14)$$

を満たす作用素 B を A に対する Moore-Penrose 一般逆と呼び, A^\dagger で表す . A が正則であるとき $A^\dagger = A^{-1}$ となる [11] .

定理 2. $h(x) = S_i^* \Phi(x) = (k(f_i^1, x) \ k(f_i^2, x) \ \dots \ k(f_i^{L_i}, x))^T$ とおくと, 未知の入力 x の $\mathcal{R}(S_i)$ への射影ノルムは

$$\|P_{\mathcal{R}(S_i)} \Phi(x)\|^2 = \langle h(x), K_i^\dagger h(x) \rangle \quad (15)$$

で与えられる . ここで $P_{\mathcal{R}(S_i)}$ は, $\mathcal{R}(S_i)$ への正射影作用素を表す . □

射影, Moore-Penrose 一般逆作用素の性質により次の補題が与えられる .

補題 1. $\mathcal{R}(S_i)$ への正射影作用素 $P_{\mathcal{R}(S_i)}$ は以下のように与えられる .

$$P_{\mathcal{R}(S_i)} = S_i S_i^\dagger \quad (16)$$

$$= S_i K_i^\dagger S_i^* \quad (17)$$

□

定理 2. の証明 補題 1 より, 未知の入力 x の射影ノルム $\|P_{\mathcal{R}(S_i)} \Phi(x)\|^2$ は,

$$\begin{aligned} \|P_{\mathcal{R}(S_i)} \Phi(x)\|^2 &= \|S_i S_i^\dagger \Phi(x)\|^2 \\ &= \|S_i K_i^\dagger S_i^* \Phi(x)\|^2 \\ &= \langle S_i K_i^\dagger h(x), S_i K_i^\dagger h(x) \rangle \quad (18) \end{aligned}$$

$$= \langle (K_i^\dagger)^* S_i^* S_i K_i^\dagger h(x), h(x) \rangle \quad (19)$$

となる . K_i は対称なので $(K_i^\dagger)^* = K_i^\dagger$, また $S_i^* S_i = K_i$ より

$$\|P_{\mathcal{R}(S_i)} \Phi(x)\|^2 = \langle h(x), K_i^\dagger h(x) \rangle \quad (20)$$

となる . □

未知の入力 f_x は射影ノルムが最大となるクラスへ決定する .

$$\|P_{\mathcal{R}(S_j)} \Phi(f_x)\|^2 > \|P_{\mathcal{R}(S_i)} \Phi(f_x)\|^2 \quad \forall j \neq i \Rightarrow f_x \in \Omega_i \quad (21)$$

4 カーネル標本空間射影法の性質

4.1 カーネルグラム行列 K の正則性

カーネルグラム行列 K が正則である場合, K の Moore-Penrose 一般逆 K^\dagger は逆行列 K^{-1} となり計算を簡単に行うことができる .

カーネル関数 k が式 (7) の Gaussian カーネルで与えられる場合には, 次の定理が成り立つ [12] .

定理 3. f_1, f_2, \dots, f_L を \mathbb{R}^N 上の互いに異なる点とすると, 第 ij 要素が

$$\exp\left(-\frac{\|f_i - f_j\|^2}{2\sigma^2}\right) \quad (\sigma > 0) \quad (22)$$

である $n \times n$ 行列は正定値である . □

したがって, カーネル関数 k が式 (7) の Gaussian カーネルの場合には, 同じ学習サンプルがない限り, K は正則となる .

カーネル関数 k が式 (6) の多項式カーネルで与えられる場合についても, 特徴空間の次元 $N+dC_d-1$ がサンプル数に比べ十分に大きいときは, 同じ学習サンプルがない限り, K は正則になると考えられる .

また, K が正則にならない場合でも以下の式により K を正則化して近似的に計算することができる .

$$K' = K + \epsilon I \quad (23)$$

ここで ϵ は正則化パラメータを表す .

4.2 追加学習

KSP 法では標本ベクトルが与えられたとき, カーネルグラム行列 K の逆行列 K^{-1} または Moore-Penrose 一般逆行列 K^\dagger をあらかじめ計算しておき, 未知の入力に対して式 (15) の射影ノルムを計算すればよい .

学習サンプルが追加された場合でも K^\dagger を計算し直す必要はなく, 以下の方法により計算することができる .

$$S_i^L = [\Phi(f^1) \ \Phi(f^2) \ \dots \ \Phi(f^L)] \quad (24)$$

$$S_i^{L+1} = [\Phi(f^1) \ \Phi(f^2) \ \dots \ \Phi(f^L) \ \Phi(f^{L+1})] \quad (25)$$

$$K_L = (S_i^L)^* S_i^L, \quad (26)$$

$$K_{L+1} = (S_i^{L+1})^* S_i^{L+1} \quad (27)$$

とおく . 標本点が追加された後の K_{L+1}^\dagger は $\tau_{L+1} = k(f_{L+1}, f_{L+1}) - \langle h(f_{L+1}), K_L^\dagger h(f_{L+1}) \rangle$ とおくと

$$K_{L+1}^\dagger = \begin{bmatrix} K_L^\dagger & \mathbf{0} \\ \mathbf{0}^\top & 0 \end{bmatrix} + \frac{1}{\tau_{L+1}} \begin{bmatrix} K_L^\dagger h(f_{L+1}) \\ -1 \end{bmatrix} \begin{bmatrix} K_L^\dagger h(f_{L+1}) \\ -1 \end{bmatrix}^\top \quad (28)$$

で与えられる . K が正則な場合も同様に K^{-1} が計算できる .

5 計算機実験

KSP法を用いて手書き数字認識実験を行い、他の手法との比較を行った。実験にはMNISTデータベースを用いた。MNISTは28x28ピクセルの画像で学習用に60,000文字、認識用に10,000文字があるが、KSP法は各クラスで部分空間の次元数であるサンプル数をそろえる必要があるため各クラス5,421文字で学習を行った。カーネル関数は式(6)の多項式カーネルと式(7)のGaussianカーネルを用い、パラメータを変化させて実験を行った。

最良のパラメータでの結果と他手法での結果を表1に示す。また、カーネル関数のパラメータ(d, σ)と誤認識率の関係を図1に示す。

また、KPCA法は提案手法では、忠実度が一定となるように各カテゴリの作用素のランクを決めているが、今回の実験では各カテゴリで作用素のランクを一定とした。

表1: 各手法の誤認識率

識別手法	パラメータ	誤認識率
KSP with Polynomial kernel	$d = 8$	2.29
KSP with Gaussian kernel	$\sigma = 7$	2.25
KPCA with Polynomial kernel	$d = 7, \text{rank}=460$	2.36
KPCA with Gaussian kernel	$\sigma = 10, \text{rank}=500$	2.32
3-NN [13]	-	2.4
Poly SVM [14]	$d = 4$	1.1

(SVM lightでの実験では、Poly SVM ($d = 4$)で1.49%の誤認識率であった。)

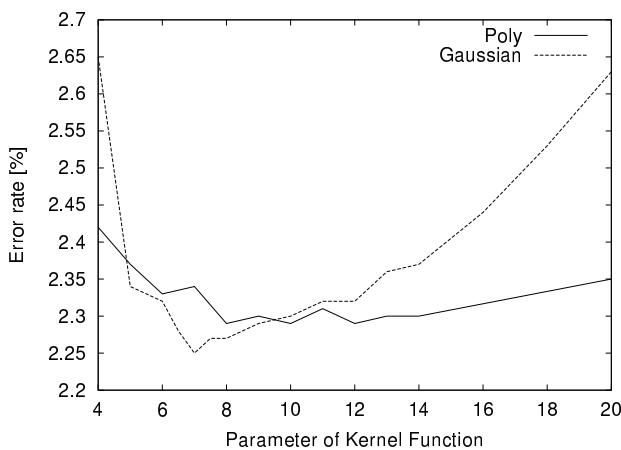


図1: カーネル関数のパラメータと誤認識率

多項式カーネルでは、 $d = 3$ (784次元から80,931,144次元への写像)から実験を行ったが、 K はすべて正則となった。

設計時1クラスあたりと認識時1サンプルあたりの平均の計算時間を表2に示す。実験にはKSP法とKPCA

法の設計にはMatlabを用い、認識にはCのプログラムを用いた。SVMは設計、認識ともThorsten JoachimsによるSVM light [5]を用いた。計算機は、CPU:Pentium4 3GHz、メモリ:2Gbyte、OS:Redhat linux9を用いた。

表2: 計算時間の比較

識別手法	設計時間 [s]	認識時間 [s]
KSP	3.2×10^2	2.9
KPCA	1.5×10^3	1.6×10^{-1}
SVM	2.8×10^3	5.8×10^{-2}

表1の実験結果より、KSP法は、SVMには及ばないが従来手法であるKPCA法よりも高い識別率を示すことが分かった。また、計算時間は表2より、認識の時間は最も遅いが設計時間では、他の手法よりも1桁近く小さいことが分かる。

6 他の手法との比較

6.1 KPCA法

KPCA法 [3],[4]は、クラス i について以下の評価量 J_{KPCA} を最小にする作用素 X_i を求める。

$$J_{KPCA} = \frac{1}{L_i} \sum_{j=1}^{L_i} \|\Phi(f_i^j) - X_i \Phi(f_i^j)\|^2 \quad (29)$$

$$\text{subject to rank}(X_i) = d \quad (30)$$

ここで f_i^j ($j = 1, 2, \dots, L_i$)はクラス i のサンプルである。 J_{KPCA} を最小にする作用素 X_i は、カーネル標本空間 S_i のKarhunen-Loève部分空間への正射影作用素となり、カーネルグラム行列 K_i の固有値を降順に λ_j ($j = 1, 2, \dots$)とし、対応する正規化固有ベクトルを v_j とすると

$$X_i = \sum_{j=1}^d \frac{1}{\lambda_j} S_i v_j v_j^T S_i^* \quad (31)$$

で表される。

KSP法は、KPCA法の式(30)の制約が無い場合と等価である。また、 K が正則な場合には、 X_i がフルランクの場合と等価である。KPCA法は、 X_i のランクや忠実度により部分空間の次元を調整できるという利点があるが、 $L \times L$ 行列を固有値分解する必要があり、計算コストは大きい。一方、KSP法では、 K が正則であれば K の逆行列を求めるだけで計算が可能となる。

詳しく計算時間を比較するため、計算機実験と同じ環境で、適当な実対称で正則な行列の固有値分解と逆行列にかかる時間を調べた。結果を図2に示す。図2より、逆行列の方が固有値分解よりも計算時間が格段に短いことが分かる。また、ガウス消去法などのアルゴリズムを

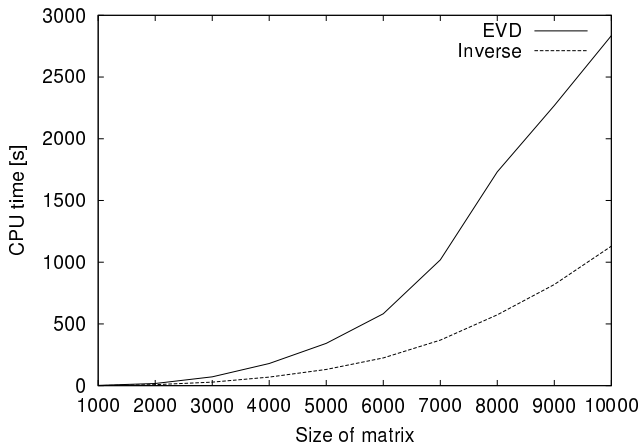


図 2: 固有値分解 (EVD) と逆行列の計算時間

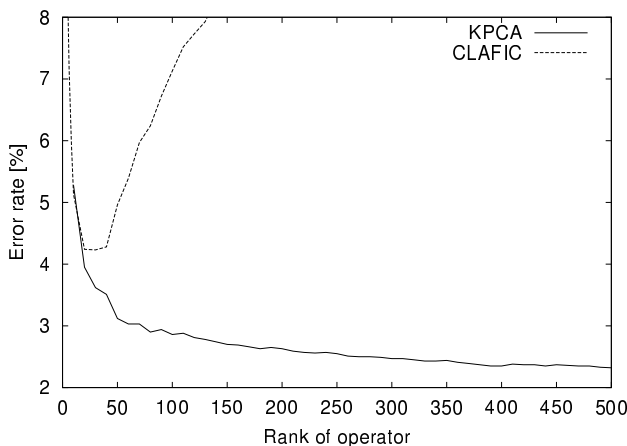


図 3: KPCA の作用素とランクの関係 (Gaussian kernel, $\sigma = 10$)

使えば、式 (15) の $K^{-1}h(x)$ を直接、高速に求めることもできる。

代表的な部分空間法に CLAFIC (CLAss-Featuring Information Compression) 法がある。KPCA 法はこの手法にカーネル法を適用したものである。KPCA 法と CLAFIC 法の作用素のランクと誤認識率の関係を図 3 に示す。CLAFIC 法は、ランクが上がり過ぎると各クラスの部分空間の重なりが大きくなり、識別を行うことができなくなるため急激に認識率が下がっている。しかし、KPCA 法の場合は、特徴空間の次元が十分に大きいため、部分空間の重なりはランクの増加とともに大きくなる。ランクを下げることによって部分空間が取り出せる特徴を消してしまっているとも考えられる。したがって、カーネル法により高次元へ写像し、部分空間法を行う場合はフルランクで部分空間へ射影する KSP 法の方が、良い認識率が期待できる。

6.2 Support Vector Machine

SVM は設計時に 2 次計画問題を解く必要があり、これには多くの計算コストが必要とする。また、基本的に SVM は 2 クラス識別器であるため、多クラス識別ではある 1 つのクラスとそれ以外のクラスでの 2 クラス識別器をクラス数分だけ作るという手法が一般的である。そのため 1 つの識別器を設計するために全クラスすべてのサンプルを用いなければならない。しかし、KSP 法は 1 クラス識別器であるため対象クラスのサンプルのみで設計ができる。したがって、実験結果でも KSP 法の方が SVM よりもかなり少ない時間で設計することができた。また、KPCA 法は 1 クラス識別器であるため、マルチテンプレートなどの手法も簡単に導入できる。

SVM は識別超平面を Support Vector (SV) と呼ばれる少数のサンプルで決める。このため認識時の計算コストは小さくなる。しかし、サンプルに少数の外れ値やエラーが混入している場合、これらは他のサンプルに比べ特異であるので SV となる可能性が高く、識別超平面を大きく劣化させてしまう可能性がある。

そこで MNIST の学習サンプルに少数のエラーを混入させ、識別実験を行った。エラー数 n だけ自クラスと他クラスのサンプルを入れ替え、実験を行った。KSP 法は Gaussian カーネル $\sigma = 7$ 、SVM は多項式カーネル $d = 4$ で行った。誤認識率と 1 クラス当りの平均の設計時間を表 3 に示す。1 段目括弧内の数字は全サンプル数に占める誤サンプルの割合である。

実験結果より、学習用サンプルにわずか 1% 以下のエラーが混入するだけでも SVM は認識率が大きく低下し、設計に要する時間も長くなるが、KSP 法ではほとんど変わらないことが確認できる。

7 まとめ

本論文ではカーネル法による新しい識別手法であるカーネル標本空間射影 (KSP) 法を提案し、その特徴、有効性を示した。本手法は、MNIST の実験においては KPCA 法以上の精度で認識ができ、設計時の計算コストも非常に小さく抑えることができる。さらに、解が単純な形であり、追加学習も簡単に計算することができる。したがって、本手法は従来提案されているカーネル法による識別手法よりも様々な面において優位な識別手法であると考えられる。また、SVM との比較により、学習サンプルのエラーに対しても頑健な識別手法であることが分かった。

今後、KSP 法の優位性を示すにはより多くの問題に対する性能を評価する必要がある。また、追加学習のアル

表 3: エラーを混入させたときの誤認識率と1クラスあたりの設計時間

n	90 (0.3%)			180 (0.6%)			270 (0.9%)		
識別手法	誤認識率 [%]	設計時間	SV の数	誤認識率 [%]	設計時間	SV の数	誤認識率 [%]	設計時間	SV の数
KSP	2.22	6 分	–	2.35	6 分	–	2.25	6 分	–
SVM	2.21	2 時間	3739	2.85	4 時間	4965	3.91	7 時間	6067

ゴリズムによる能動学習についても今後の検討課題の1つである。

謝辞

本研究は、文部科学省科学研究費補助金 (No.15500101) の補助を受けた。

参考文献

- [1] V.N. Vapnik, *Statistical Learning Theory*, Wiley, New-York, 1998.
- [2] B. Schölkopf, A. Smola, and K.R. Müller, “Nonlinear component analysis as a kernel eigenvalue problem,” *Neural Computation*, vol.10, no.5, pp.1299-1319, 1998.
- [3] 前田英作, 村瀬洋, “カーネル非線形部分空間法によるパターン認識,” *信学論 (D-II)*, vol.J82-D-II, no.4, pp.600-612, Apr. 1999.
- [4] 津田宏治, “ヒルベルト空間における部分空間法,” *信学論 (D-II)*, vol.J82-D-II, no.4, pp.592-599, Apr. 1999.
- [5] T. Joachims, “Making large-scale svm learning practical,” in *Advances in Kernel Methods – Support Vector Learning*, eds. B. Schölkopf, C. Burges, and A. Smola, pp.169–185, MIT Press, 1998.
- [6] J. Platt, “Fast training of support vector machines using sequential minimal optimization,” in *Advances in Kernel Methods — Support vector learning*, eds. B. Schölkopf, C. Burges, and A. Smola, pp.185-208, MIT Press, 1999.
- [7] J. Mercer, “Functions of positive and negative type, and their connection with the theory of integral equations.,” *Trans. Lond. Phil. Soc. (A)*, vol.209, pp.415-446, 1909.
- [8] S. Saitoh, *Integral transforms, reproducing kernels and their applications*, Pitman research notes in mathematics series BA00276264 369, Harlow : Addison Wesley Longman, 1997.
- [9] 斎藤三郎, *再生核の理論入門*, 牧野書店, 2002.
- [10] B. Schölkopf, S. Mika, C. Burges, P. Knirsch, K.R. Müller, G. Rätsch, and A. Smola, “Input space vs. feature space in kernel-based methods,” *IEEE Transactions on Neural Networks*, vol.10, no.5, pp.1000–1017, Sept. 1999.
- [11] 竹内 啓柳井 晴夫, *射影行列・一般逆行列・特異値分解*, UP 応用数学選書, no.10, 東京大学出版会, 1983.
- [12] S. Haykin, ed., *Neural Networks: A Comprehensive foundation*, IEEE Computer Society Press, 1995.
- [13] Y. LeCun, L.D. Jackel, L. Bottou, A. Brunot, C. Cortes, J.S. Denker, H. Drucker, I. Guyon, U.A. Muller, E. Sackinger, P. Simard, and V. Vapnik, “Comparison of learning algorithms for handwritten digit recognition,” *International Conference on Artificial Neural Networks*, eds. F. Fogelman, and P. Gallinari, pp.53-60, Paris, 1995, EC2 & Cie.
- [14] C. Cortes, and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol.20, no.3, pp.273-297, 1995.