

# 抑制型カーネル標本空間射影法によるパターン認識

鷲沢 嘉一<sup>†</sup> 山下 幸彦<sup>†</sup>

<sup>†</sup> 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

E-mail: †{washizawa,yamasita}@ide.titech.ac.jp

あらまし パターン認識の一手法であるカーネル標本空間射影 (KSP) 法を拡張した抑制型カーネル標本空間射影 (SKSP) 法を提案する。カーネル法は、マースカーネルを用いて入力ベクトルを高次元の特徴空間へ写像し、認識を行う手法で SVM や PCA などに応用され高い性能を示している。KSP 法は、マースカーネルが作る特徴空間に写像されたサンプルが張るカーネル標本空間へ未知の入力を正射影し、そのノルムを比較することによりクラスを決定する 1 クラス識別器である。SKSP 法は、KSP 法を拡張し未知の入力を斜射影することにより他のクラスの影響を抑制し、KSP 法よりも高精度な認識を実現する。また、手書き数字認識実験と 2 クラス公開データベースの実験を行いその有効性を示す。

キーワード カーネル法, カーネル標本空間射影 (KSP) 法, サポートベクタマシン (SVM), カーネル主成分分析 (KPCA), Tikhonov 正則化

## Suppressed kernel sample space projection method for pattern recognition

Yoshikazu WASHIZAWA<sup>†</sup> and Yukihiro YAMASHITA<sup>†</sup>

<sup>†</sup> Tokyo Institute of Technology, 2-12-1, Ookayama, Meguro-ku, Tokyo 152-8552, Japan

E-mail: †{washizawa,yamasita}@ide.titech.ac.jp

**Abstract** We propose SKSP; suppressed kernel sample space projection method which is extended from KSP; kernel sample space projection method. By kernel based methods, after an input vector is mapped to a high dimensional feature space by a Mercer kernel function, it is classified. A kernel based method is applied to SVM and PCA, and achieves high performance. KSP is an one-class classifier. It classifies an unknown input vector by comparing the projection norms onto kernel sample spaces which are spanned by samples in the feature space. By SKSP, the effect of other classes is suppressed and useful features are extracted with an oblique projection. We show experimental results of hand written digits recognition problem and some two-class classification problems, and show its advantages.

**Key words** Kernel based method, Kernel sample projection method, Support vector machine (SVM), Kernel PCA (KPCA), Tikhonov regularization

### 1. はじめに

手書き文字認識などの認識問題において、部分空間法が広く用いられている [1]。その代表的な手法である CLAFIC (Class Feature Information Compression) 法は、未知の入力を各クラスのサンプルが作る Karhunen-Loève (KL) 部分空間へ正射影し、そのノルムを比較することにより、クラスを決定する [2]。前田ら、津田は Support Vector Machine (SVM) などに適用されているカーネル法を CLAFIC 法へ適用し、非線形な識別器

へ拡張した [3] [4]。本論文ではこれらを KPCA 法と表記する。

カーネル法は入力ベクトル  $x$  を非線形な写像  $\Phi$  によって高次元の特徴空間へ写像し、識別を行う手法である。カーネル法では写像  $\Phi$  は陽に定義されず、以下の式を満たすカーネル関数を用いて計算を行う。

$$k(x, y) = \langle \Phi(x), \Phi(y) \rangle \quad (1)$$

即ち、写像後の内積が入力空間の関数となるような関数  $k$  を用いて  $\Phi$  を用いることなく、特徴空間での内積を計算する。マ

サーカーネル [5] には、式 (1) を満たす写像  $\Phi$  が存在する．一般的にマースカーネルとして以下の関数がよく用いられる．

$$\text{Linear} : k(x, y) = \langle x, y \rangle \quad (2)$$

$$\text{Polynomial} : k(x, y) = (\langle x, y \rangle + 1)^d \quad (3)$$

$$\text{Gaussian} : k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right) \quad (4)$$

SVM では識別器は、サンプルと未知の入力との内積で表されるため、内積をカーネル関数に置き換えることによりカーネル法を実現できる [6]．しかし、Principal Component Analysis (PCA) は非常に次元の大きい特徴空間での固有値問題を解く必要がある．Schölkopf らは特異値分解を用いることにより特徴空間での固有値問題を解き、主成分を求める手法を提案した [7]．

CLAFIC 法ではサンプルが全空間に対して大きな次元の空間に分布しているため、クラス間での重なりを消し、クラスに固有の特徴を取り出すために PCA あるいは KL 変換を用いている．KL 変換のランクは、識別率に対し非常に敏感であり、ランクが低すぎると抽出されるクラス特徴が不十分であり、ランクが高すぎるとクラス間の重なりによって識別を行うことができなくなる．一方、カーネル法では写像  $\Phi$  によって特徴空間へ写像されたサンプルには重なりがないため、クラス間の重なりをなくすという意味では KL 部分空間を用いる必要はない．カーネル標本空間射影 (KSP; Kernel sample space projection) 法は特徴空間でのサンプルの KL 部分空間ではなく、サンプルが張る空間へ正射影し、そのノルムを比較することによりクラスを決定する手法である [8]．KPCA では KL 部分空間を求めるためにサンプル数の次元の固有値問題を解く必要があるが、KSP 法ではサンプル数の次元の行列の逆を求めるだけで設計ができる．このため、KSP 法は設計時において KPCA 法よりも計算コストが小さく、また追加学習が容易に行うことができるという特徴がある．

一方、PCA, KL 変換の拡張として相対 KL 変換が提案され、パターン認識にも応用されている [9] [10]．本論文ではこれを RPCA 法と表記する．CLAFIC 法は、自クラスの特徴を取り出すだけであるのに対し、RPCA 法は、他クラスの特徴を抑制しながら自クラスの特徴を取り出すことができる．これにより RPCA 法は CLAFIC 法よりも高い性能を示す．また、RPCA 法をカーネル法により拡張した Kernel RPCA (KRPCA) 法も提案されている [11]．しかし、前述のように特徴空間ではクラス間の重なりを消すためにランクを下げる必要はない．

本論文では、KSP 法に他クラスの抑制を導入し、拡張を行った抑制型カーネル標本空間射影 (SKSP) 法を提案する．このために、第 2 節で KSP 法について述べ、別の視点からの特徴付けを行い、正則化項を導入する．次に第 3 節において SKSP 法を定義し、解を求め、その性質について述べる．第 4 節では、手書き数字認識実験と公開データベースによる 2 クラス識別実験を行い、提案手法の有効性を示す．第 5 節で他の手法と比較し、提案手法の有効性を述べる．

本論文で用いる記号、表記を表 1 に示す．

表 1 表記、記号一覧

$\Phi$	特徴空間への写像
$k(\cdot, \cdot)$	カーネル関数
$A^*$	作用素 $A$ の共役作用素
$A^\dagger$	作用素 $A$ の MP 一般逆
$\langle \cdot, \cdot \rangle$	内積
$\Omega_i$	クラス $i$ のパターンの集合
$f_n^i$	$\Omega_i$ からの $n$ 番目のサンプル
$L_i$	$\Omega_i$ のサンプル数
$\Psi_i$	クラス $i$ 以外の抑制すべきパターンの集合
$g_n^i$	$\Psi_i$ からの $n$ 番目のサンプル
$M_i$	$\Psi_i$ のサンプル数
$S_i$	$S_i = [\Phi(f_1^i) \Phi(f_2^i) \dots \Phi(f_{L_i}^i)]$
$T_i$	$T_i = [\Phi(g_1^i) \Phi(g_2^i) \dots \Phi(g_{M_i}^i)]$
$U_i$	$U_i = [S_i T_i]$
$K_{S_i}$	自クラスのカーネルグラム行列 $K_{S_i} = S_i^* S_i$
$K_{U_i}$	全クラスのカーネルグラム行列 $K_{U_i} = U_i^* U_i$
$x$	未知の入力
$\mathcal{R}(A), \mathcal{N}(A)$	$A$ の値域、核
$P_{\mathcal{R}(A)}$	$\mathcal{R}(A)$ への正射影作用素
$\ f\ $	$l_2$ ノルム $\ f\  = \sqrt{\langle f, f \rangle}$
$\ A\ _{l_{ub}}$	作用素ノルム $\ A\ _{l_{ub}} = \sup_{f \neq 0} \frac{\ Af\ }{\ f\ }$
$\ A\ _2$	Frobenius ノルム $\ A\ _2 = \sqrt{\text{tr}(A^* A)}$
$I$	単位行列、恒等作用素

## 2. カーネル標本空間射影法

クラス  $\Omega_i$  からのサンプル  $f_1^i \dots f_{L_i}^i$  が得られたとき、これらのベクトルを特徴空間へ非線形写像したベクトル  $\Phi(f_1^i) \dots \Phi(f_{L_i}^i)$  が張る空間をクラス  $i$  のカーネル標本空間と呼ぶ．即ち  $S_i = [\Phi(f_1^i) \dots \Phi(f_{L_i}^i)]$  とおくと、 $\mathcal{R}(S_i)$  がクラス  $i$  のカーネル標本空間となる．

カーネルグラム行列  $K_{S_i}$  を以下のように定義する．

$$K_{S_i} = S_i^* S_i \quad (5)$$

$$= \begin{bmatrix} k(f_1^i, f_1^i) & \dots & k(f_1^i, f_{L_i}^i) \\ \vdots & \ddots & \vdots \\ k(f_{L_i}^i, f_1^i) & \dots & k(f_{L_i}^i, f_{L_i}^i) \end{bmatrix} \quad (6)$$

このときクラス  $i$  のカーネル標本空間への正射影作用素  $P_{\mathcal{R}(S_i)}$  は以下の式で表される．

$$P_{\mathcal{R}(S_i)} = S_i K_{S_i}^\dagger S_i^* \quad (7)$$

KSP 法では未知の入力  $x$  を各クラスのカーネル標本空間へ正射影し、そのノルムが最大となるクラスへ決定する． $x$  のクラス  $i$  のカーネル標本空間への正射影ノルムは

$$\|P_{\mathcal{R}(S_i)} x\|^2 = \langle h(x), K_{S_i}^\dagger h(x) \rangle \quad (8)$$

で与えられる．ここで

$$h(x) = S_i^* \Phi(x) = (k(x, f_1^i), k(x, f_2^i), \dots, k(x, f_{L_i}^i))^\top \quad (9)$$

である．作用素  $S_i$  は、列が非常に大きな行列あるいは無限の作用素であるが、 $K_{S_i}$ 、 $h(x)$  はそれぞれ  $(L_i \times L_i)$  の実対称行列、

$L_i$  次元実ベクトルとなるため計算機上での計算が可能である．

カーネル標本空間への正射影作用素は，サンプルに対してはそのサンプルを与える作用素であり，未知の入力に対してはサンプルの線形結合としての最良の近似を与える作用素である．即ち以下の評価基準を最小にする作用素として定義することができる．

$$\min_{X_i} : J[X_i] = \frac{1}{L_i} \sum_{s=1}^{L_i} \|\Phi(f_s^i) - X_i \Phi(f_s^i)\|^2 \quad (10)$$

$$\text{subject to} : \mathcal{N}(X_i) \supset \mathcal{R}(S_i)^\perp \quad (11)$$

KSP 法は，カーネルグラム行列が正則であれば常にサンプルをすべて正しく分離する．しかし，サンプルは一般にノイズや外れ値を含んでいるためサンプルをすべて正しく分離できるとオーバーフィッティングにより汎化能力が下がってしまうことがある．オーバーフィッティングは識別器の識別境界が複雑になりすぎる場合に生じる．オーバーフィッティングの問題を回避するため，SVM や Ada-Boost ではソフトマージンなどの正則化の手法が提案されている [12] ．

KSP 法では，識別境界の複雑さは正射影作用素の作用素ノルムで測ることができる．しかし，作用素ノルムを計算することは難しいため，Frobenius ノルムを用いる．任意の作用素に対して  $\|A\|_{lub} \leq \|A\|_2$  が成り立つため，Frobenius ノルムを小さくすることにより，作用素ノルムを小さくすることができる．この正則化を導入した以下の評価基準を最小にする作用素  $X_i = \tilde{P}_{\mathcal{R}(S_i)}$  を正則化 KSP 作用素と呼ぶ．

$$\min_{X_i} : J[X_i] = \frac{1}{L_i} \sum_{s=1}^{L_i} \|\Phi(f_s^i) - X_i \Phi(f_s^i)\|^2 + \epsilon \|X_i\|_2^2 \quad (12)$$

$$\text{subject to} : \mathcal{N}(X_i) \supset \mathcal{R}(S_i)^\perp \quad (13)$$

ここで  $\epsilon > 0$  は正則化パラメータを表す．正則化 KSP 作用素  $\tilde{P}_{\mathcal{R}(S_i)}$  は以下の式で与えられる．

$$\tilde{P}_{\mathcal{R}(S_i)} = S_i(K_{S_i} + \epsilon L_i I)^{-1} S_i^* \quad (14)$$

式 (12) の正則化は逆問題の分野では Tikhonov 正則化と呼ばれ，広く用いられている [13] ．

### 3. 抑制型カーネル標本空間射影法

KSP 法は 1 クラスの識別器であり，クラス  $i$  の識別関数を設計するにはクラス  $i$  のサンプルしか用いていない．しかし，クラス  $i$  以外のサンプルも設計に使用した方が高い識別率を示すことが期待できる．例えば 2 つ以上のクラスに同じような特徴がいくつかの次元に存在する場合や 1 つのクラスには存在するが他のクラスには全く存在しない特徴があるような場合を考える．このような場合は自クラスと他クラスの両方の特徴を使い，前者は抑制し，後者は増幅することにより，より有用な特徴を取り出すことができると考えられる．

逆問題における Wiener Filter [14] や RPCA, KRPCA 法などは雑音や他のクラスを抑制しながら推定や特徴抽出を行っている．抑制型カーネル標本空間射影 (SKSP; Suppressed KSP)

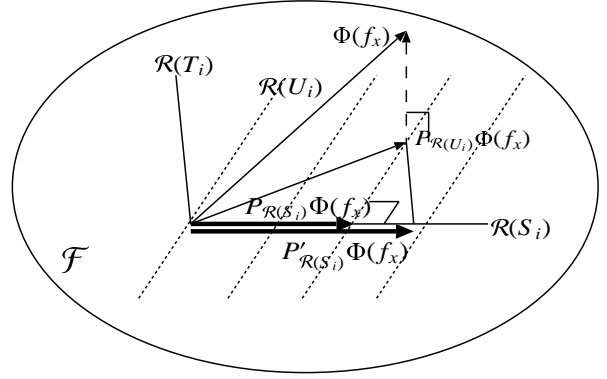


図 1 抑制型カーネル標本空間射影法

法は，KSP 法にこのような抑制項を導入し，拡張したものである．

$\Psi_i$  をクラス  $i$  に対して抑制すべき他クラスのパターンの集合とし， $\Psi_i$  からのサンプルを  $g_1^i \dots g_{M_i}^i$  とする． $T_i = [\Phi(g_1^i) \dots \Phi(g_{M_i}^i)]$ ， $U_i = [S_i \ T_i]$  とおく．このとき以下の評価基準を最小にする作用素  $X_i = P'_{\mathcal{R}(S_i)}$  を SKSP 作用素と呼ぶ．

$$\min : J[X_i] = \frac{1}{L_i} \sum_{s=1}^{L_i} \|\Phi(f_s^i) - X_i \Phi(f_s^i)\|^2 + \frac{1}{M_i} \sum_{t=1}^{M_i} \|X_i \Phi(g_t^i)\|^2 \quad (15)$$

$$\text{subject to} : \mathcal{N}(X_i) \supset \mathcal{R}(U_i)^\perp \quad (16)$$

全サンプルに対するカーネルグラム行列  $K_{U_i} = U_i^* U_i$  が正則のとき，SKSP 作用素  $P'_{\mathcal{R}(S_i)}$  は以下の式で与えられる．

$$P'_{\mathcal{R}(S_i)} = U_i D_i K_{U_i}^{-1} U_i^* \quad (17)$$

$$D_i = \begin{bmatrix} I_{L_i} & \mathbf{0}_{L_i M_i} \\ \mathbf{0}_{M_i L_i} & \mathbf{0}_{M_i M_i} \end{bmatrix} \quad (18)$$

また， $K_{U_i}$  が正則のとき， $P'_{\mathcal{R}(S_i)}$  は以下の式を満たす．

$$P'_{\mathcal{R}(S_i)} P'_{\mathcal{R}(S_i)} = P'_{\mathcal{R}(S_i)} \quad (19)$$

$$P'_{\mathcal{R}(S_i)} P_{\mathcal{R}(U_i)} = P'_{\mathcal{R}(S_i)} \quad (20)$$

$$P_{\mathcal{R}(U_i)} P'_{\mathcal{R}(S_i)} = P'_{\mathcal{R}(S_i)} \quad (21)$$

$$P'_{\mathcal{R}(S_i)} v = 0 \quad \forall v \in \mathcal{R}(T_i) \quad (22)$$

ここで  $P_{\mathcal{R}(U_i)}$  は， $\mathcal{R}(U_i)$  への正射影作用素を表す．式 (19),(22) より， $P'_{\mathcal{R}(S_i)}$  は  $\mathcal{R}(T_i)$  に沿った  $\mathcal{R}(S_i)$  への斜射影作用素である．また，式 (20),(21) より，未知のサンプルに対しては  $\mathcal{R}(U_i)$  へ正射影されたあと  $\mathcal{R}(T_i)$  に沿って  $\mathcal{R}(S_i)$  へ斜射影されることが分かる．これを図 1 に示す．

SKSP 法に対しても KSP 法と同様に正則化を導入することができる．以下の評価基準を最小にする作用素  $X_i = \tilde{P}'_{\mathcal{R}(S_i)}$  を正則化 SKSP 作用素と呼ぶ．

$$\min : J[X_i] = \frac{1}{L_i} \sum_{s=1}^{L_i} \|\Phi(f_s^i) - X_i \Phi(f_s^i)\|^2 + \frac{1}{M_i} \sum_{t=1}^{M_i} \|X_i \Phi(g_t^i)\|^2 + \epsilon \|X_i\|_2^2 \quad (23)$$

$$\text{subject to : } \mathcal{N}(X_i) \supset \mathcal{R}(U_i)^\perp \quad (24)$$

ここで  $\epsilon > 0$  は正則化パラメータを表す．正則化 SKSP 作用素  $\tilde{P}'_{\mathcal{R}(S_i)}$  は以下の式で与えられる．

$$\tilde{P}'_{\mathcal{R}(S_i)} = U_i D_i (K_{U_i} + \epsilon \tilde{D}_i)^{-1} U_i^* \quad (25)$$

$$\tilde{D}_i = \begin{bmatrix} L_i I_{L_i} & \mathbf{0}_{L_i, M_i} \\ \mathbf{0}_{M_i, L_i} & M_i I_{M_i} \end{bmatrix} \quad (26)$$

CLAFIC 法や KPCA 法, KSP 法においては正射影作用素を用いているので射影ノルム  $\|P_i x\|$  あるいは  $\|P'_i \Phi(x)\|$  が最大となるクラスと投影距離  $\|x - P_i x\|$  あるいは  $\|\Phi(x) - P'_i \Phi(x)\|$  が最小となるクラスは一致する．しかし, SKSP 法では斜射影作用素であるので射影ノルムが最大となるクラスと投影距離が最小となるクラスは必ずしも一致するとは限らない．そこで SKSP 法では次の 2 通りの識別規則が考えられる．

$$\|P'_{\mathcal{R}(S_i)} \Phi(x)\|^2 > \|P'_{\mathcal{R}(S_j)} \Phi(x)\|^2 \quad \forall i \neq j \Rightarrow x \in \Omega_i \quad (27)$$

$$\|\Phi(x) - P'_{\mathcal{R}(S_i)} \Phi(x)\|^2 < \|\Phi(x) - P'_{\mathcal{R}(S_j)} \Phi(x)\|^2 \quad \forall i \neq j \Rightarrow x \in \Omega_i \quad (28)$$

$\Psi_i = \phi$  とすれば SKSP 法は KSP 法に一致する．このため, SKSP 法は KSP 法の拡張であると考えられる．KSP 法は 1 クラス識別器でありクラスに含まれる特徴を十分に取出すことができる．このため, SKSP 法は SVM などの 2 クラス識別器のようにすべての他クラスのサンプルを  $\Psi_i$  に入れる必要はない．これは多クラスの識別の応用において非常に有利な特徴である．

## 4. 計算機実験

### 4.1 手書き数字認識実験

実験には MNIST データベースを用いた．MNIST は 28x28 ピクセルの '0'-'9' の画像が学習用に 60,000 文字, 試験用に 10,000 文字が用意されている．

MNIST の学習用サンプルは 60,000 文字あり, 現在のコンピュータではこの大きさの逆行列を求めることは難しい．このため MNIST で SKSP 法の実験を行うためには抑制集合  $\Psi_i$  に含めるパターンを選び出す必要がある．抑制集合  $\Psi_i$  には自クラスのサンプルに類似したサンプルを含めれば少ないサンプルでも高い性能が期待できる．そこで他クラスのサンプルについて以下の評価基準値が大きいサンプル  $n$  個を抑制集合とした．

$$t(g_i) = \|P_{\mathcal{R}(S_i)} \Phi(g_i)\|^2 \quad (29)$$

ここで  $P_{\mathcal{R}(S_i)}$  は KSP 作用素を表す．

KSP 法, SKSP 法, KPCA 法, SVM での実験結果を表 2 に示す．KPCA 法は, ランクが大きいときには KSP 法と等価になるため, ランクは 500 以下で実験を行った．また, SVM の

表 2 各手法の誤認識率

識別手法	パラメータ	誤認識率 [%]
KSP with Gaussian kernel	$\sigma = 7$	2.25
SKSP with Gaussian kernel	$\sigma = 7, n = 5000$	1.79
KPCA with Gaussian kernel	$\sigma = 10, \text{rank} = 500$	2.32
SVM with Polynomial kernel	$d = 4, C = 10$	1.49

実験には SVM light [15] を用いた．

$n$  の値を変化させたときの誤認識率の変化を図 2 に示す．図中の 1 は式 (27) による識別規則, 2 は式 (28) による識別規則を表す．

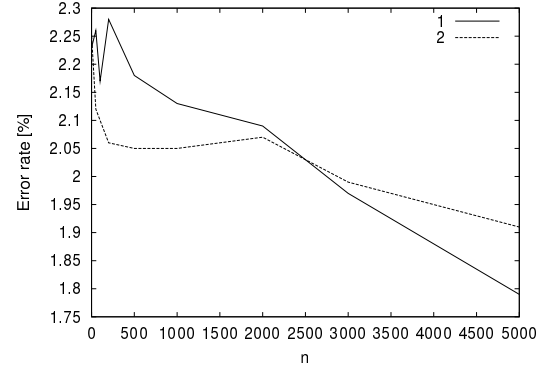


図 2 SKSP の実験結果

### 4.2 2 クラス識別問題

次に文献 [12] [16] [17] で使用されている 13 種類の 2 クラス問題に対して KSP 法, SKSP 法の実験を行った<sup>(注1)</sup>．これらのデータセットは学習用, 試験用の組がそれぞれ 100 あるいは 20 セット用意されている．

SKSP 法, KSP 法と既存の識別器の実験結果を表 3 に示す．表中の数値は 100 あるいは 20 セットの識別率の平均と標準偏差である．カーネル関数は Gaussian カーネルを用い, 抑制集合は他クラスのすべてのサンプルとした．表中の KSP 法と SKSP 法以外の実験結果は, 前述の論文からの引用である．最も識別率の平均の高い結果を Bold font で表し, 次に高いものを Emphasized font で表す．

## 5. 検討, 考察

手書き数字認識実験より, SKSP 法は KSP 法よりも安定して良い識別率を示すことが分かった．識別規則は  $n$  が小さいときには式 (27),  $n$  が大きいときには式 (28) が有効であることが分かった．これは  $n$  が小さいときには自クラスのサンプルの影響が大きいため射影ノルムが有効であり,  $n$  が大きくなると他クラスのサンプルの影響が大きくなり, 投影距離が有効になったためであると考えられる．

2 クラス識別問題の実験では, 最良および 2 番目の識別率を与えた問題の数は SKSP 法が最も多い結果であった．また, SKSP 法は多くの識別問題において KSP 法よりも高い識別率

(注1): データセットは以下の URL よりダウンロードした．

<http://ida.first.fraunhofer.de/projects/bench/benchmarks.htm>

表 3 2 クラス識別問題の実験結果

dataset	SKSP	KSP	RBF	AB Reg	SVM	KFD
Banana	<b>10.4 ± 0.5</b>	<b>10.4 ± 0.5</b>	10.8 ± 0.4	10.9 ± 0.4	11.5 ± 0.7	10.8 ± 0.5
Breast-cancer	26.0 ± 4.6	29.7 ± 4.5	27.6 ± 4.7	26.5 ± 4.5	26.0 ± 4.7	<b>24.5 ± 4.6</b>
Diabetis	<b>23.0 ± 1.6</b>	24.5 ± 1.9	24.3 ± 1.9	23.8 ± 1.8	23.5 ± 1.7	23.2 ± 1.6
Flare-solar	37.2 ± 4.5	39.1 ± 2.4	34.4 ± 2.0	34.2 ± 2.2	<b>32.4 ± 1.8</b>	33.2 ± 1.7
German	<b>23.4 ± 2.1</b>	31.3 ± 2.5	24.7 ± 2.4	24.7 ± 2.4	23.6 ± 2.1	23.7 ± 2.2
Heart	15.8 ± 3.1	<b>15.4 ± 3.3</b>	17.6 ± 3.3	16.5 ± 3.5	16.0 ± 3.3	16.1 ± 3.4
Image	2.8 ± 0.4	2.9 ± 0.5	3.3 ± 0.7	<b>2.7 ± 0.6</b>	3.0 ± 0.6	4.8 ± 0.6
Ringnorm	18.0 ± 2.3	19.9 ± 1.8	1.7 ± 0.2	1.6 ± 0.1	1.7 ± 0.1	<b>1.5 ± 0.1</b>
Splice	11.2 ± 0.7	12.6 ± 0.7	10.0 ± 0.8	<b>9.5 ± 0.7</b>	10.9 ± 0.7	10.5 ± 0.6
Thyroid	<b>4.0 ± 2.3</b>	4.2 ± 2.3	4.5 ± 2.1	4.6 ± 2.2	4.8 ± 2.2	4.2 ± 2.1
Titanic	29.4 ± 10.3	28.3 ± 9.4	23.3 ± 1.3	22.6 ± 1.2	<b>22.4 ± 1.0</b>	23.3 ± 2.1
Twonorm	2.4 ± 0.1	<b>2.3 ± 0.1</b>	2.9 ± 0.3	2.7 ± 0.2	3.0 ± 0.2	2.6 ± 0.2
Waveform	<b>9.6 ± 0.4</b>	11.2 ± 0.6	10.7 ± 1.1	9.8 ± 0.8	9.9 ± 0.4	9.9 ± 0.4
# of bold	5	3	0	2	2	2
# of emph.	4	1	1	3	2	3

を示した．これは他クラスの影響を抑制し，有効な特徴が抽出されているためであると考えられる．

同じカーネル法による識別器に SVM が提案されており，様々な問題において高い汎化能力を示している．SVM と SKSP 法の最も大きな違いは，識別関数である．SVM の識別関数は，特徴空間における重みベクトル  $w$  と入力との内積で表される．つまり，特徴空間での 1 次識別関数と考えることができる．一方，SKSP 法の特徴量は特徴空間での射影ノルムにより与えられるため識別関数は特徴空間での 2 次関数と考えられる．よって SKSP 法の識別関数は SVM よりも多くの自由度を持っていると考えられる．

SVM の識別超平面は Support Vector (SV) と呼ばれるごく少数のサンプルによって決められる．SV は通常 2 クラスの境界付近に存在するサンプルである．通常，サンプルにはノイズや外れ値が混入している．特に識別境界付近のサンプルはこれらの影響が大きいと考えられる．SVM はサンプルの分布に関わらず，境界付近のサンプルのみによって識別を行っているため，ノイズの影響を受けやすいと考えられる．Soft margin や  $\nu$ -SVM などノイズの影響を抑えるための正則化の手法も提案されているが，境界付近のサンプルのみを使っている以上，ノイズに対して頑健性は低い．文献 [8] には実験によって KSP 法の頑健性が SVM よりも優れていることが示されている．

しかし，一方で SVM のスパース性は認識時における計算コストを低くすることができる．乗算とカーネル関数の計算コストを  $s, k$  とし，サンプル数を  $L$ ，SV の数を  $L_{SV}$  とすると認識時に必要な主な計算のコストは以下のようになる．

$$\text{SVM} : (k + s)L_{SV} \quad (30)$$

$$\text{SKSP} : sL^2 + (s + k)L \quad (31)$$

一般に  $s < k$ ， $L_{SV} < L$  である．SKSP 法は前述の通り 2 次識別関数を持つため  $L$  に対し 2 次のオーダーとなる．しかし，SKSP 法は全クラスのサンプルを  $\Psi$  に含める必要はないため， $L$  の数を減らし，計算コストを抑えることもできる．

逆に SVM の設計時の計算コストは非常に高い．SVM は設

計時に 2 次計画問題を解く必要があるためである．この 2 次計画問題を高速に解くためにさまざまな手法が提案されているが，それでも計算コストは大きい．一方，KSP 法，SKSP 法は逆行列の計算と行列の乗算のみで計算が行うことができる．また，前述の通り SKSP 法は全クラスのサンプルを使う必要はないため設計時の計算コストも抑えることができる．

サンプル数が非常に大きくなる場合には，計算量削減のために KSP 法，SKSP 法はマルチテンプレートを導入することが考えられる．マルチテンプレートは各クラス内でいくつかのサブクラスを用意し，このサブクラスに対して識別を行う手法である．2 クラス識別器では自クラスと他クラスのサンプルをすべて使う必要があるため，この手法を導入しても計算量を削減することはできないが，SKSP 法はすべての他クラスサンプルを使う必要はないため簡単に実現可能であると考えられる．

## 6. 結 び

本論文では，KSP 法を拡張した SKSP 法とその正則化の手法を提案し，解を与えた．手書き数字認識実験と 2 クラス識別問題についての実験を行い，他の識別手法との比較しその有効性を示した．最後に SVM との比較を行った．

今後は，SKSP 法の性能を確認するために多くの識別問題に対して実験を行う必要がある．また，マルチテンプレートの具体的な適用法についても検討する必要がある．

## 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．

## 文 献

- [1] E. Oja: "Subspace methods of pattern recognition", Wiley, New-York (1983).
- [2] S. Watanabe and N. Pakvasa: "Subspace method in pattern recognition", Proc. 1st Int. J. Conf on Pattern Recognition, Washington DC, pp. 25-32 (1973).
- [3] 前田, 村瀬: "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), **J82-D-II**, 4, pp. 600-612 (1999).

- [4] 津田：“ヒルベルト空間における部分空間法”，信学論 (D-II), **J82-D-II**, 4, pp. 592–599 (1999).
- [5] 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, **10**, 5, pp. 1000–1017 (1999).
- [6] V. N. Vapnik: “Statistical Learning Theory”, Wiley, New-York (1998).
- [7] B. Schölkopf, A. Smola and K.-R. Müller: “Nonlinear component analysis as a kernel eigenvalue problem”, Neural Computation, **10**, 5, pp. 1299–1319 (1998).
- [8] 鷲沢, 山下：“カーネル標本空間射影法によるパターン認識”，第6回情報論的学習理論ワークショップ IBIS2003 予稿集, pp. 149–154 (2003).
- [9] Y. Yamashita and H. Ogawa: “Relative Karhunen-Loève transform”, IEEE Transactions on signal processing, **44**, 2 (1996).
- [10] 池野, 山下, 小川：“相対 KL 変換法によるパターン認識”，信学論 (D-II), **J80-D-II**, 2, pp. 541–547 (1997).
- [11] 鷲沢, 疋田, 田中, 山下：“カーネル相対主成分分析による多クラスパターン認識”，第二回 FIT (情報科学技術フォーラム) 情報レターズ, pp. 207–208 (2003).
- [12] G. Rätsch, T. Onoda and K.-R. Müller: “Soft margins for AdaBoost”, Machine Learning, **42**, 3, pp. 287–320 (2001). also NeuroCOLT Technical Report NC-TR-1998-021.
- [13] A. N. Tikhonov and V. Y. Arsenin: “Solution of Ill-posed problems”, V. H. Winston and Sons (1977).
- [14] H. Ogawa and E. Oja: “Projection filter, Wiener filter, and Karhunen-Loève subspaces in digital image restoration”, J. Math Anal. Appl., **114**, pp. 37–51 (1986).
- [15] T. Joachims: “Making large-scale svm learning practical”, Advances in Kernel Methods – Support Vector Learning (Eds. by B. Schölkopf, C. Burges and A. Smola), MIT Press, pp. 169–185 (1998).
- [16] S. Mika, G. Rätsch, J. Weston, B. Schölkopf and K.-R. Müller: “Fisher discriminant analysis with kernels”, Neural Networks for Signal Processing IX (Eds. by Y.-H. Hu, J. Larsen, E. Wilson and S. Douglas), IEEE, pp. 41–48 (1999).
- [17] G. Rätsch: “Robust Boosting via Convex Optimization”, PhD thesis, University of Potsdam, Neues Palais 10, 14469 Potsdam, Germany (2001).

## 付 録

クラスを表す添字  $i$  は省略する .

### 1. 式 (17),(25) の証明

$\mathcal{N}(X) \supset \mathcal{R}(U)^\perp$  より,  $X$  は

$$X = UBU^*$$

と表すことができる . これを式 (23) に代入して

$$\begin{aligned} J &= \frac{1}{L} \sum_{s=1}^L \|\Phi(f_s) - UBU^*\Phi(f_s)\|^2 \\ &\quad + \frac{1}{M} \sum_{t=1}^M \|UBU^*\Phi(g_t)\|^2 + \epsilon \|UBU^*\|_2^2 \\ &= \frac{1}{L} \sum_{s=1}^L \{k(f_s, f_s) - \langle U^*\Phi(f_s), BU^*\Phi(f_s) \rangle \\ &\quad - \langle BU^*\Phi(f_s), U^*\Phi(f_s) \rangle + \langle BU^*\Phi(f_s), K_U BU^*\Phi(f_s) \rangle\} \\ &\quad + \frac{1}{M} \sum_{t=1}^M \langle BU^*\Phi(g_t), K_U BU^*\Phi(g_t) \rangle + \epsilon \text{tr}[B^* K_U B K_U] \end{aligned}$$

$U^*\Phi(f_s)$  と  $U^*\Phi(g_t)$  はそれぞれ  $K_U$  の第  $s$  列, 第  $L+t$  列の列ベクトルであるので

$$\begin{aligned} J &= \frac{1}{L} \text{tr}[K_U D - (K_U D)^* B K_U D - (K_U D)^* B^* K_U D \\ &\quad + (K_U D)^* B^* K_U B K_U D] \\ &\quad + \frac{1}{M} \text{tr}[(K_U D)^* B^* K_U B K_U D] + \epsilon \text{tr}[B^* K_U B K_U] \\ &= \text{tr}\left[\frac{1}{L}(K_U D - K_U B K_U D - K_U B^* K_U D) \right. \\ &\quad \left. + K_U B^* K_U B K_U \tilde{D}^{-1} + \epsilon B^* K_U B K_U\right] \end{aligned}$$

となる .  $J$  の  $B$  による変分は

$$\begin{aligned} \delta J &= 2\text{tr}[(\delta B)^*(K_U B K_U \tilde{D}^{-1} K_U \\ &\quad - \frac{1}{L} K_U D K_U + \epsilon K_U B K_U)] \end{aligned}$$

a)  $\epsilon \neq 0$  のとき

$\delta J = 0$  となるのは

$$K_U B (K_U + \epsilon \tilde{D}) \tilde{D}^{-1} K_U = \frac{1}{L} K_U D K_U$$

であり, これを満たす  $B$  は

$$\begin{aligned} B &= D(K_U + \epsilon \tilde{D})^{-1} + Y \tilde{D} (K_U + \epsilon \tilde{D})^{-1} \\ &\quad - K_U^\dagger K_U Y K_U K_U^\dagger \tilde{D} (K_U + \epsilon \tilde{D})^{-1} \end{aligned}$$

で与えられる . ここで  $Y$  は任意の  $(L_i + M_i)$  実正方形行列を表す .

$\mathcal{R}(K_U) = \mathcal{R}(U^*)$  より  $K_U K_U^\dagger U^* = U^*$  が成り立つ . よって

$$X = UBU = UD(K_U + \epsilon \tilde{D})^{-1} U^*$$

となる .

b)  $\epsilon = 0$  のとき

$K_U$  が正則であるとき  $\delta J = 0$  となるのは

$$B = DK_U^{-1}$$

$$X = UDK_U^{-1} U^*$$

で,  $X$  が与えられるときである .

### 2. 式 (19)-(22) の証明

式 (19)-(21):

$$P'_{\mathcal{R}(S)} P'_{\mathcal{R}(S)} = UDK_U^{-1} U^* UDK_U^{-1} U^* = UDK_U^{-1} U^*$$

$$P'_{\mathcal{R}(S)} P_{\mathcal{R}(U)} = UDK_U^{-1} U^* U K_U^{-1} U^* = UDK_U^{-1} U^*$$

$$P_{\mathcal{R}(U)} P'_{\mathcal{R}(S)} = U K_U^{-1} U^* UDK_U^{-1} U^* = UDK_U^{-1} U^*$$

式 (22):

$$v = \sum_{j=1}^{M_i} v_j \Phi(g_j) = U_i v', \quad v' = \underbrace{(0 \dots 0)}_{L \text{ 個}} v_1 v_2 \dots v_{M_i}^\top \text{ と}$$

表すことができるので

$$P'_{\mathcal{R}(S)} v = UDK_U^{-1} U^* U v' = U D v' = 0$$

となる .