社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2007－MPS－63 (3)
2007／3／3

# カーネル主成分分析を用いた学習機械のパラメタ自動決定法

関 口 涼 平† 高 橋 治 久†† 堀 田 一 弘††

本論文では，カーネル主成分分析 (KPCA) に基づいた新しい学習機械を提案する．KPCA は，パターン識別の前処理として用いられ，主成分分析を使う場合より良い認識性能が出せる場合も報告されている．KPCA と線形サポートベクトルマシンを合わせたカーネルプロジェクションマシン (KPM) は，モデル選択との併用により，少ない学習時間でサポートベクトルマシン (SVM) と同等の汎化性能が得られる利点があるが，その性能は SVM と同様カーネルパラメタに大きく依存する．本論文では，KPM に対し，KPCA の理論に基づいて最適なカーネルパラメタを求めるアルゴリズムを提案し，計算機実験によりその性能を評価する．SVM との計算機実験による比較により，提案手法が少ない計算時間でよりよい性能を達成していることを示す．

# The Automatic Parameter Tuning in Learning with Kernel PCA

RYOHEI SEKIGUCHI,† HARUHISA TAKAHASHI† and KAZUHIRO HOTTA†

This paper proposes a new learning machine based on Kernel Principal Component Analysis (KPCA). KPCA is usually used as a pre-processing process preceding application of learning machines, thereby better performance is achieved than linear Principal Component Analysis in some cases. Kernel Projection Machine (KPM), which is proposed by Blanchard etc., applies linear SVM after KPCA with a model selection process. Although KPM can perform equally to Support vector Machine (SVM) with smaller execution time, its performance heavily depends on the kernel parameter. We propose a novel algorithm to determine the optimal kernel parameter in the learning process. The algorithm is obtained based on the theory of KPCA, and we show that the proposed learning method show a better performance than SVM in both generalization and computation time through computer experiments.

## 1. はじめに

パターン識別の分野では，主成分分析 (PCA) やカーネル主成分分析 (KPCA) はノイズなどが入った高次元データにたいして前処理として適用することで，識別性能を上げるために用いられることが多い．一般に，ノイズがパターン情報を上回るところで，次元圧縮をすることにより，識別器において分類精度を改善する目的で用いられる．そこで Blanchard 等[1] は，KPCA を次元圧縮に，線形 SVM を識別器としたカーネル射影学習機械 (KPM) を提案した．KPM の識別性能は SVM と同等であり，次元圧縮により更にサイズの小さい，実働化に適したネットワークが得られる．

KPM については，KPCA 処理に対して SVM と同様の問題がある．すなわちカーネルとして，ガウシアンカーネルが主として用いられるが，そのカーネルパラメタに対しては，自動決定する方法が無く，膨大な予備実験なしに KPM の能力を最大限引き出すことは出来ない．このような学習機械をパラメタを持つ学習機械と言い，実用上経験的にパラメタを設定するのが一般的であり，パラメタを含めて自動的に学習により決定する方法が望まれる．

本論文では，KPM に対し，最適なカーネルパラメタを自動決定する方法を提案し，計算機実験により，その効果を検証する．KPCA を適用する段階で，どこの固有値で次元圧縮するかは，モデル選択にゆだねられる．基本的アイデアは，選択された固有値 (特徴) に対し，最適なパラメタを，適切な評価関数を定めて求めることである．評価関数は，基本的に選択された特徴に入る情報と選択されなかった特徴に入る情報の比が最大になるよう選ばれる．本論文では，特徴のそれぞれの領域での平均値の差を評価関数として選び，最適なパラメタを自動決定するアルゴリズムを提案する．

提案法の効果を検証するため，いくつかの問題につ

† 電気通信大学大学院 情報通信工学専攻
Department of Information and Communication Engineering, The University of Electro-Communications
†† 電気通信大学 情報通信工学科
The University of Electro-Communications

いて計算機実験を行い，その汎化性能と出力されたネットワークのサイズについて SVM と比較して検討する．

## 2. カーネル主成分分析

パターン識別問題において，パターン特徴空間の次元数，すなわち特徴の数を適切に設定することが重要である．一般に特徴の数を増やすと計算量が増大し，実用的でなくなる．

主成分分析（PCA）は，特徴量の相関関係を利用することによって次元圧縮された部分空間への射影を求める方法であり，パターン識別信号処理の基礎的手法である．PCA は線形変換であるため，本来，分布が線形性をもつデータに対して次元圧縮の効果を得ることができる．しかし，分布の非線形性が強いデータに対しては，非線形の特徴量を抽出することができず大きく情報を損失してしまうと考えられる．

Schölkopf[2) ]によって提案されたカーネル主成分分析（KPCA）は，カーネル写像によって高次元空間に非線形写像を行い写像先で PCA を行うことにより，分布の非線形性の強いデータについて PCA を行うことを可能にした．すなわち，非線形主成分分析を行うことになり，線形主成分分析では抽出できないデータの特徴量を見つけることができる．

### 2.1 カーネル法

#### 2.1.1 カーネル法による非線形写像

カーネル法はカーネル関数によって高次元空間に写像を行うことで，非線形成分に対しての処理を可能にする方法である．

入力された $n$ 次元特徴ベクトルを $\boldsymbol{x}=(x_1,\ldots,x_n)^t$ とし，サンプルを $S=\{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_l\}$ で表す．カーネル写像によりベクトル $\boldsymbol{x}$ は $\boldsymbol{\phi}(\boldsymbol{x})\in F$ によって $p$ 次元線形空間 $F$ に非線形写像される

$$\begin{aligned}\boldsymbol{x} &= (x_1,\ldots,x_n)^t\\ &\mapsto \boldsymbol{\phi}(\boldsymbol{x})=(\phi_1(\boldsymbol{x}),\ldots,\phi_p(\boldsymbol{x}))^t\end{aligned}$$

カーネル写像により，写像の成分 $\phi_i(i=1,\ldots,p)$ を直接扱うことなく，その内積であるカーネル関数（スカラー関数）

$$k(\boldsymbol{x},\boldsymbol{z})=\langle\boldsymbol{\phi}(\boldsymbol{x}),\boldsymbol{\phi}(\boldsymbol{z})\rangle=\boldsymbol{\phi}(\boldsymbol{x})^t\boldsymbol{\phi}(\boldsymbol{z})$$

により，任意次元の像空間を扱うことができる．カーネル行列 $K$ を

$$K_{ij}=\langle\boldsymbol{\phi}(\boldsymbol{x}_i),\boldsymbol{\phi}(\boldsymbol{x}_j)\rangle=k(\boldsymbol{x}_i,\boldsymbol{x}_j)$$

とする．写像先の特徴行列を $\Phi=(\phi(\boldsymbol{x}_1),\ldots,\phi(\boldsymbol{x}_l))^t$ とすれば，$K=\Phi^t\Phi$ であり，この行列は半正定値行列となる．

### 2.2 カーネル主成分分析の定式化

通常の線形 PCA は基準点をデータの平均ベクトルにおくことにより次元削減によって発生する原空間と部分空間の誤差を最小にすることがわかっている．しかし KPCA では、この性質は望めないため，重心を基準点として一般化はされていない．しかし，経験的には重心を基準とするほうが特徴をよく捉えられていると考えられる．

写像先での主成分分析を行うには，$p$ 次元線形空間から $d$ 次元線形部分空間への正規直交基底による線形変換を考えればよい．中心を平均ベクトルに置いて主成分軸を求めるには行列

$$G=\left(I_l-\frac{1}{l}\mathbf{1}\mathbf{1}^t\right)K\left(I_l-\frac{1}{l}\mathbf{1}\mathbf{1}^t\right)$$

の固有値固有ベクトルを用いればよい．ここでサイズ $l$ の単位行列を $I_l$，要素が全て 1 の $l$ 行 1 列ベクトルを$\mathbf{1}$とする．$K=\Phi^t\Phi$ が正定値行列であれば $G$ も正定値行列である．$G$ は写像先の重心を取ることから中心化グラム行列と呼ぶ．

### 2.3 カーネル主成分分析の問題点

カーネル主成分分析を使用する場合，カーネルの種類によってカーネルパラメタの値を事前に設定する必要がある．本論文では最も頻繁に応用されている Gaussian カーネル

$$K(\boldsymbol{x},\boldsymbol{z})=\exp\left(-\frac{||\boldsymbol{x}-\boldsymbol{z}||}{2\sigma^2}\right)$$

を使用している．この場合のカーネルパラメタは $\sigma$ である．

カーネル主成分分析の能力は，カーネルパラメタに非常に強く依存している．最適なパラメタを設定したとき，カーネル法の能力を最大限に引き出すことができるが，事前にこれを行う有効な手法は提案されていない．一般には事前実験を行うことで最適パラメタを調べるが，学習そのものよりも計算量が膨大になり，学習機械へ応用するための障害になる．

カーネルパラメタによって大きく性能が左右される例を示す．2 次元平面上に非線形成分を持つデータ（図 1 左上）を用意し，Gaussian カーネルによる KPCA を行った．第 1 主成分の結果を図 1 右上・図 1 左下・図 1 右下に示す．図 1 右上のようにパラメタ $\sigma$ が最適値よりも小さすぎる場合，データの近傍にしか特徴が現れず，未知データに対する情報を抽出できない．逆に図 1 左下のように $\sigma$ が最適値より大きすぎる場合は，出力が大まかになりすぎて細かい分布情報を抽出できなくなっている．後述する手法で求めた最適パラメタで同様に計算した結果を図 1 右下に示す．この

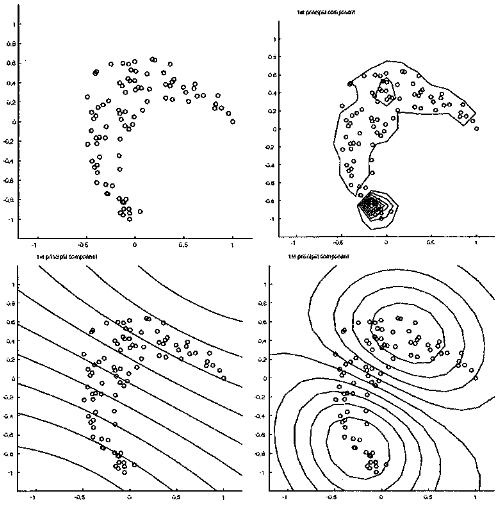

**図 1**　非線形データの例と KernelPCA の出力
Fig. 1　Example for non-linear dataset and Outputs of KernelPCA

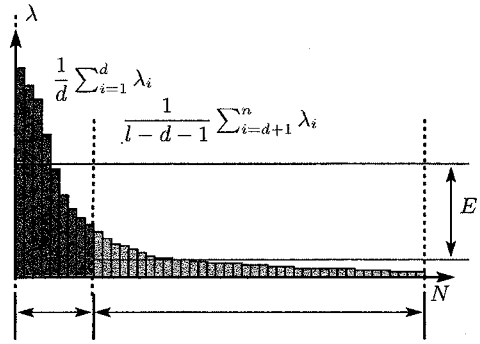


**図 2**　各主成分軸の分散と評価関数
Fig. 2　Eingenvalue and Evaluate function

図ではバランスよく特徴を捉えており，訓練サンプル以外の未知サンプルに対しても非線形な特徴を一般化することができる．

## 3. 最適パラメタの自動決定法

### 3.1 評価関数

パターン識別のための次元削除の観点から考えられる最適な写像空間は，残すべき主成分軸に分散が多く集まり，切り捨てるべき軸にはほぼ 0 に近い小さな分散が集中すればよい．なぜならば，このような写像空間であれば残すべき主成分が多くの圧縮された情報を持ち，切り捨てられる主成分によって損失する情報を最小限に抑えることができるからである．最適パラメタは，この条件を満たす評価関数を最適化することにより得られる．本論文では，最適パラメタの評価関数として図 2 のように，残す主成分軸の分散の平均と切り捨てる主成分軸の分散の平均の差をとり，これを最大化させる．この条件から得られる評価関数は，カーネル関数のパラメタを $\theta$ とし，第 $i$ 主成分の分散を $\lambda_i$ とすると

$$E(\theta)=\frac{1}{d}\sum_{i=1}^{d}\lambda_i-\frac{1}{l-d-1}\sum_{i=d+1}^{l}\lambda_i \tag{1}$$

となる．

主成分分析における第一主成分は，その分散の値を取り除いてから次元圧縮をすることで，分析能力が逆に上がるという現象が起こる場合がある．この場合の評価関数 $E(\theta)$ は，式 (1) と同じように

$$E(\theta)=\frac{1}{d}\sum_{i=2}^{d}\lambda_i-\frac{1}{l-d-2}\sum_{i=d+2}^{l}\lambda_i$$

となる．

### 3.2 評価関数 $E(\theta)$ の微分

評価関数 $E(\theta)$ の $\theta$ に関する微分は

$$\frac{\partial E}{\partial\theta}=\frac{1}{dl}\sum_{i=1}^{d}\frac{\partial\mu_i}{\partial\theta}-\frac{1}{l(l-d-1)}\sum_{i=d+1}^{l}\frac{\partial\mu_i}{\partial\theta}$$

となる．

ここで $\mu_i$ の $\theta$ による偏微分を求める．$\mu_i$ は直接微分することはできないので陰関数定理を用いる．$\mu_k$ は行列 $G(\theta)$ の固有値であり，固有方程式

$$\det(\mu_k I_l-G(\theta))=0 \tag{2}$$

を満たす．ここで $f_{\mu_k}(\cdot)$ を式 (2) の解，すなわち $f_{\mu_k}(\theta)=\mu_k$ とする．したがって $\det(\mu_k I_l-G)$ の転置余因子行列を $C$ とすると，$\Sigma_{i=1}^{l}C_{ii}\neq 0$ ならば陰関数が微分可能で，

$$\frac{\partial f_{\mu_k}}{\partial\theta}=-\frac{\frac{\partial\det(\mu_k I_l-G)}{\partial\theta}}{\frac{\partial\det(\mu_k I_l-G)}{\partial\mu_k}}=\frac{\Sigma_{i=1}^{l}\Sigma_{j=1}^{l}C_{ij}\frac{\partial G_{ij}}{\partial\theta}}{\Sigma_{i=1}^{l}C_{ii}}$$

と表すことができる．ここで転置余因子行列 $C$ の計算は多くの計算時間を必要とするため，計算を簡略化する必要がある．

転置余因子の定理を用いることで，

$$\begin{aligned}\frac{\partial f_{\mu_k}}{\partial\theta}&=\frac{\left(\prod_{\substack{i=1\\i\neq k}}^{l}(\mu_k-\mu_i)\right)\boldsymbol{v}_k^t\frac{\partial G_{ij}}{\partial\theta}\boldsymbol{v}_k}{\prod_{\substack{i=1\\i\neq k}}^{l}(\mu_k-\mu_i)}\\&=\boldsymbol{v}_k^t\frac{\partial G}{\partial\theta}\boldsymbol{v}_k\end{aligned}$$

となり，これを $\mu_k$ の微分として用いることで，パラメタによる微分値を求めることができる．

本論文では，評価関数 $E(\theta)$ を最大にする最適パラメタ $\theta=\theta^*$ を求めるために，極値を探索する二分法 (Bisection method) を適用した．

### 3.3 学習機械

KPCA の前処理の後，ハードマージンを持つ線形サポートベクトルマシン（LSVM）を識別器として用いる．最適パラメタになり次元削除で発生するデータの誤差を最小に抑えつつ，少ない次元数で高い汎化性能を実現することができる．

## 4. 計算機実験

本研究では提案手法を用いて，パターン識別の性能比較を行う．識別用サンプルデータである "heart","diabetes" について，提案手法と非線形サポートベクトルマシン (NLSVM) との識別性能の比較を行った．

また，サイズ 46×44 の 300 枚の顔画像☆を用意し，男性と女性の 2 クラスジェンダー判別も加えて行った．

### 4.1 カーネルパラメタ学習

提案手法については，式 (1) において取り出すべき主成分数 $d$ を $d=10,20,\ldots$ と 10 ずつに設定し，それぞれの $d$ における最適なカーネルパラメタ $\sigma_d$ を学習した．

NLSVM については，最適化アルゴリズムに SMO を使用した．また，カーネルのパラメタである $\sigma$ とソフトマージンのパラメタである $C$ を探索する必要があるため，$\sigma=[2,4,\ldots,10], C=[2,4,\ldots,10]$ の 25 の組み合わせについてそれぞれ 10 回のパラメタ探索を行い，その中での最も性能の良い値を選択した．

どちらの学習機械に対しても，汎化誤差の計算には 10-fold cross-validation を 5 回行った．

### 4.2 結果

データセット "heart","diabetes"，そして顔画像に対する識別器の性能比較を表 1 から表 3 に示す．

提案手法は毎回ほぼ同じカーネルパラメタに収束した結果，少ない試行回数で安定して高い汎化性能を出すことができた．一方通常の NLSVM は広い範囲でパラメタ探索をする必要があるため，最良の結果を出力するまでに非常に大きな時間を必要とする．したがって提案手法は非常に小さな計算規模で，SVM よりも高い汎化性能を実現することが可能になった．

☆ 本論文に使用した顔画像データは財団法人ソフトピアジャパン研究開発部地域結集型共同研究推進室から使用許諾を受けたものであり，権利者に無断で複写，利用，配布等を行う事は禁じられています．

**表 1** データセット "diabetes" の性能比較
Table 1 Result of the Dataset "heart"

| | NLSVM (SMO) | KPCA +LSVM |
|---|---|---|
| 最小汎化誤差 [%] | 26.37 | 22.91 |
| 平均パラメタ探索時間 [s] | 47.57 | 174.6 |
| 平均実動計算時間 [s] | 1190 | 217.3 |

**表 2** データセット "heart" の性能比較
Table 2 Result of the Dataset "diabetes"

| | NLSVM (SMO) | KPCA +LSVM |
|---|---|---|
| 最小汎化誤差 [%] | 24.32 | 16.98 |
| 平均パラメタ探索時間 [s] | 21.93 | 72.11 |
| 平均実動計算時間 [s] | 548.5 | 109.3 |

**表 3** 男女顔画像識別の性能比較
Table 3 Result of the Dataset "Gender"

| | NLSVM (SMO) | KPCA+LSVM (第 1 主成分削除) |
|---|---|---|
| 最小汎化誤差 [%] | 10.33 | 6.340 |
| 平均パラメタ探索時間 [s] | 26.57 | 121.7 |
| 平均実動計算時間 [s] | 664.3 | 168.2 |

### 4.3 考察

実験の結果，提案手法によって得られた最適カーネルパラメタは効率的に入力データの特徴抽出を行い，高い識別性能へ繋がるパラメタであることが分かった．

また KPCA に関しては，データの分布によって第 1 主成分を削除した方が高い識別を得る場合があるということが分かった．これは非線形性の特徴を持っているなど対象データの性質に依存するため，今後はデータの性質も含めて検討していく必要がある．

## 5. おわりに

本研究ではカーネルのパラメタを自動的に決定することで，カーネル主成分分析に基づいた識別器を提案した．LSVM と組み合わせることで，非線形の SVM よりも高速でよりよい性能を持つ識別器を実現することができた．

提案手法のパラメタ自動決定法は他のカーネルマシンについても応用できる可能性があり，マルチクラス SVM や KernelICA への応用も考えられる．

## 参考文献


1) Blanchard, G., Massart, P., Vert, R. and Zwald, L.: Kernel Projection machine: a new tool for pattern recognition, *Proc. 18th, Neural Information Processing System*, pp.1649–1656 (2004).
2) Scholkopf, B., Smola, A. and Muller, K.: Kernel Principal Component Analysis, *Advances in Kernel Methods*, pp.327–353 (1998).