Vol.2010-MPS-77 No.28
2010/3/5

錘制約部分空間法に基づくサポートベクトルマシンの開発

北村 拓也[†1] 竹内 翔吾[†1] 阿部 重夫[†1]

本研究では, 1) 部分空間法に基づくサポートベクトルマシン (SS-SVM) による特徴選択と 2) 錘制約部分空間法に基づくサポートベクトルマシン (CRS-SVM) を提案する. 1) の手法では, SS-SVM を用いて識別の観点から最適な重みを決定し, 識別における重要度を意味する重みを基準として固有ベクトルの選択を繰り返し, 識別においてよりよい部分空間の軸の選択を行う. 2) の手法では, SS-SVM に各クラスの教師ベクトルの張る凸錘をそれぞれの部分空間とする錘制約部分空間法 (CRS) を適用することで固有値問題を避け, 学習コストを削減する. ベンチマークデータを用いて, 従来手法と比較することで提案手法の有効性を示す.

Cone Restricted Subspace Based Support Vector Machines

Takuya Kitamura,[†1] Shogo Takeuchi [†1] and Shigeo Abe[†1]

In this paper, we proposed two methods: 1) the feature selection using subspace based support vector machines (SS-SVM), and 2) cone restricted subspace based vector machines (CRS-SVM). In 1), we iteratively select the eigenvector according to the associated weights, which were obtained by using SS-SVM and mean importance for classification, as the dictionaries of the class subspace for classification. In 2), we apply CRS, in which the convex cone spanned by training data belonging to a class is a class subspace, to SS-SVM. Thus, we do not need to solve the eigenvalue problems, and the training cost is reduced. We demonstrate the effectiveness of the proposed methods with the conventional methods for benchmark datasets.

1. はじめに

パターン認識において, 代表的な識別手法の一つとして部分空間法[1),2)] が広く適用されてきた. 部分空間法では各クラス部分空間への射影長をクラス類似度とし, 重み付けされたクラス類似度の中で最大となるクラスに分類する. また, サポートベクトルマシン (SVM)[3)] の概念を取り入れ, 他のクラス部分空間とのオーバーラップを考慮することにより識別の観点から最適な重み付けを行う部分空間法に基づくサポートベクトルマシン (SS-SVM)[4)] が提案されている. SS-SVM では非線形問題に対応するためにカーネル法を用いたカーネル主成分分析 (KPCA)[5),6)] を用いて各クラス部分空間を生成し, ある入力ベクトルにおいて重み付けされた全クラス類似度の中で対応するクラス類似度が最大となるように制約条件をおき, マージン最大化と誤認識最小化とを考慮した最適化を行う. 現在まで適用する SVM のモデルとして最小自乗 SVM (LS-SVM)[3)] と線形計画 SVM (LP-SVM)[3)] が用いられており, それぞれの手法を部分空間法に基づく LS-SVM (SSLS-SVM), 部分空間法に基づく LP-SVM (SSLP-SVM) とよぶ. SSLP-SVM では学習時に特徴選択を同時に行うが, SSLS-SVM では軸における属するクラスのデータの分散を基に特徴選択を行うため, 他のクラスの部分空間の重なりなどを考慮しておらず, 識別において必ずしもよい軸の選択とはいえない. また SSLP-SVM では重みに非負制約を設けているが, 重みが負となる軸が識別において必要な可能性がある. さらに, 各クラスごとの共分散行列の固有値問題を解く必要があるため, 大規模データにおいて部分空間の生成における学習コストが膨大となる. 一方, 固有値分解を解く必要のない部分空間法として, 錘制約部分空間法 (CRS)[7)] が提案されている. CRS では, 教師ベクトルそのものを凸錘のクラス部分空間とおき, 入力ベクトルと各クラス凸錘部分空間との角度を類似度として識別を行う. しかしながら, 入力ごとに凸錘との角度を最小自乗法を用いて計算するため識別コストが膨大となりうる. 属するクラスから逸脱した教師ベクトルが存在したとき, この部分空間の張る凸錘は広くなり, クラス部分空間同士で重なりができる. その重複した領域は未分類領域となる.

そこで, 従来の SS-SVM と CRS のこれらの問題を解消するため, 1) SSLS-SVM における最適な特徴選択, 2) CSR に基づく SVM (CRS-SVM) を提案する. 1) の手法では, 全ての固有ベクトルを軸として選択し SSLS-SVM により重みを決定し, それらの重みを基に軸

†1 神戸大学大学院工学研究科, 神戸市
Graduate School of Engineering, Kobe University, Kobe

の選択を行う．この時, ある条件を満たすまで選択と学習を繰り返すが, 条件として二つの基準を設ける．一つ目は SSLP-SVM と同様に非負制約を設ける．負の重みとなる軸において属するクラスのパターンより他のクラスパターンにおける類似度が大きくなりやすいことを意味しており, これらの軸を削減することで属するクラスのパターンのみの類似度が大きくなる軸のみを選択できる．二つ目の基準として各軸の重みの絶対値の大小を用いる．重みの絶対値が小さい軸は識別に影響を与えにくく, 識別において冗長な軸として削除する．実験によりこれら二つの特徴選択の比較を行い, さらに SSLP-SVM においても非負制約を設けずに重みを最適化する手法を用いて従来の SSLP-SVM と比較し, SS-SVM において負の重みを持つ軸が識別において重要となりうるかを確認する．2) の手法では, CRS の識別関数を入力ベクトルの重み付けをした各軸への射影長の総和とすることにより SS-SVM の決定関数と同じ形にし, SS-SVM と同様に識別の観点から重みを最適化する．このとき他のクラス部分空間への射影長と等しい領域以外では未分類領域を存在しない．さらに識別時にはベクトルの内積のみを行うため, 識別コストが大幅に削減できる．この手法により, SS-SVM と比べ部分空間の生成における学習コストが削減でき, また CRS と比べ識別におけるコストを削減し, 汎化能力を向上させせることができる．

以下, 2 節では従来の SS-SVM を説明し, 3 節では提案する特徴選択について述べ, 4 節では CRS-SVM について述べる．5 節においては, ベンチマークデータセットを用いた計算機実験のによる従来の SS-SVM と提案手法を比較することにより提案手法の有効性を検証し, 6 節で結論を述べる．

2. 部分空間法に基づくサポートベクトルマシン

2.1 クラス部分空間の生成

SS-SVM ではクラス部分空間の生成に標本特徴空間[8] 上で定義した KPCA[5),6)] を用いている．ここで, n クラス m 次元の入力 $\mathbf{x}$ における識別問題について考える．入力ベクトル $\mathbf{x}$ を N 個の一次独立な教師ベクトル $\mathbf{x}_{k_i}$ $(i=1,...,N)$ の張る N 次元標本特徴空間に写像関数 $\mathbf{h}(\mathbf{x}) = (H(\mathbf{x}_{k_1},\mathbf{x}),...,H(\mathbf{x}_{k_N},\mathbf{x}))^T$ により写像する．ここで, 一次独立なデータはコレスキー分解を用いて選択する．このとき, 高次元特徴空間への写像関数を $\mathbf{g}(\mathbf{x})$ としたとき, $H(\mathbf{x},\mathbf{x}') = \mathbf{g}^T(\mathbf{x})\mathbf{g}(\mathbf{x}')$ である．このとき, 以下の共分散行列の固有値問題を解くことにより, クラス i 部分空間の r_i 個の軸 $\boldsymbol{\varphi}_{ik}$ を決定する．

$$\frac{1}{|X_i|}\sum_{i\in X_i}\mathbf{h}(\mathbf{x}_j)\,\mathbf{h}^T(\mathbf{x}_j)\boldsymbol{\varphi}_{ik} = \lambda_{ik}\ \boldsymbol{\varphi}_{ik} \tag{1}$$

ここで, X_i はクラス i の教師ベクトル集合, $|X_i|$ は X_i の教師ベクトル数, λ_{ik} は $\boldsymbol{\varphi}_{ik}$ に対応する固有値を示す．また, 固有値を大きい順に並び替え, 累積寄与率 a を用いて各クラス部分空間の軸の選択を行う．このとき $a(r_i) = \sum_{j=1}^{r_i}\lambda_{ij}/\sum_{j=1}^{N}\lambda_{ij}\times 100$ と定義し, a が閾値 κ 以上になるまで対応する固有ベクトルを軸として選択し, 部分空間の次元数 r_i を決定する．

2.2 重みの最適化

SS-SVM において, 各クラス部分空間の各軸における類似度を以下のような列ベクトルで定義する．

$$\mathbf{f}_i(\mathbf{x}) = \left(\frac{(\boldsymbol{\varphi}_{i1}^T\mathbf{h}(\mathbf{x}))^2}{\|\boldsymbol{\varphi}_{i1}\|^2\|\mathbf{h}(\mathbf{x})\|^2},...,\frac{(\boldsymbol{\varphi}_{ir_i}^T\mathbf{h}(\mathbf{x}))^2}{\|\boldsymbol{\varphi}_{ir_i}\|^2\|\mathbf{h}(\mathbf{x})\|^2}\right)^T \tag{2}$$

このとき, 決定関数を類似度の列ベクトルを用いて

$$D_i(\mathbf{x}) = \mathbf{w}_i^T\mathbf{f}_i(\mathbf{x}) + b_i \tag{3}$$

と表す．ここで, $\mathbf{w}_i$ をクラス i 部分空間の k 番目の軸の重み w_{ik} の列ベクトルとする．このとき, 入力ベクトル $\mathbf{x}$ を $D_i(\mathbf{x})$ が最大となるクラス i に分類する．また, SSLS-SVM においてバイアス項 b_i を省く．

通常の SS-SVM では, 各クラス間の部分空間を分離するためにマージンを最大し, 誤認識を最小化するように重みを最適化するため, 最適化問題は同時定式化方式となる．そのとき, SSLS-SVM の最適化問題は以下のように表せる．

$$\text{minimize}\ Q(\mathbf{w},\boldsymbol{\xi}) = \frac{1}{2}\sum_{i=1}^{n}\|\mathbf{w}_i\|^2 + \sum_{i=1}^{n}\sum_{\substack{y_j\neq i,\\ j=1}}^{M}\frac{CM}{2n\,|X_{y_j}|}\xi_{ji}^2 \tag{4}$$

$$\text{subject to}\ \mathbf{w}_{y_j}^T\mathbf{f}_{y_j}(\mathbf{x}_j) - \mathbf{w}_i^T\mathbf{f}_i(\mathbf{x}_j) = 1-\xi_{ji}$$
$$\text{for}\quad i\neq y_j, i=1,...,n, j=1,...,M \tag{5}$$

ここで M は全教師ベクトル数, C はマージン最大化と誤認識最小化のトレードオフを決定するパラメータであり, ξ_{ji} はスラック変数である．

SSLP-SVM では SSLS-SVM と異なり, 最適化問題に重みの非負制約を設けている．またバイアス項 b_i は正または負の値をとるため異なる正の変数によって正負の値をとるようにする．すなわち, $b_i = b_i^+ - b_i^-$ $(b_i^+\geq 0,\ b_i^-\geq 0)$ とする．さらに, 不等号制約を等号制約に変換するため, 非負のスラック変数 u_{ji} $(i=1,...,n,\ i\neq y_j,\ j=1,...,M)$ を用いる．このときの SSLP-SVM の最適化問題は以下のように表せる．

$$\text{minimize} \quad Q(\mathbf{w},\mathbf{b},\boldsymbol{\xi},\mathbf{u}) = \sum_{i=1}^{n}\sum_{k=1}^{r_k} w_{ik} + \sum_{j=1}^{M}\sum_{i\neq y_j, i=1}^{n} \frac{C\,M}{n\,|X_{y_j}|}\xi_{ji} \tag{6}$$

$$\text{subject to } \mathbf{w}_{y_j}^T \mathbf{f}_{y_j}(\mathbf{x}_j) + (b_{y_j}^+ - b_{y_j}^-) - \mathbf{w}_i^T \mathbf{f}_i(\mathbf{x}_j) - (b_i^+ - b_i^-) = 1 - \xi_{ji} + u_{ji}$$

$$\text{for } i \neq y_j,\ i = 1, ..., n,\ j = 1, ..., M \tag{7}$$

$$\xi_{ji},\ u_{ji} \geq 0 \ \text{ for } i \neq y_j,\ i = 1, ..., n,\ j = 1, ..., M \tag{8}$$

$$w_{ik} \geq 0 \text{ for } i = 1, ..., n,\ k = 1, ..., r_k \tag{9}$$

ここで, $\mathbf{b} = (b_1, \ldots, b_n)^T$, $\mathbf{u} = (\ldots, u_{ji}, \ldots)^T$ である．また SSLS-SVM とは異なり, 重み w_{ik} が 0 となりうる．$\mathbf{f}_i(\mathbf{x})$ の k 番目の要素を $f_{ik}(\mathbf{x})$ とすると, $w_{ik} = 0$ となる $f_{ik}(\mathbf{x})$ は識別に影響を与えないため, 対応する軸を削除することができる．すなわち SSLP-SVM では累積寄与率を用いることなく識別と同時に特徴選択を行うことができる．

3. 最適な重みに基づく特徴選択

SSLS-SVM では, 式 (1) により求めた固有値の大きい固有ベクトルを条件を満たすまでクラス部分空間の軸として選択する．しかしながら, これらの固有値は識別においての重要性を意味してはいないため, 固有値が大きい軸が識別の観点から冗長な軸であることがあり, 逆に小さい軸が識別において重要である場合がある．この点, 重みの最適化と同時に特徴選択を行える SSLP-SVM ではこのような問題は生じない．しかしながら, 重みに非負制約を設けているため, 識別において重要な軸であるが重みが負になりうるため削除される軸が存在しうる．

本節では, この問題点を解決するために SSLS-SVM における最適な重みに基づいた特徴選択と負の重みを許す SSLP-SVM の学習について述べる．

3.1 非負制約による SSLS-SVM の特徴選択

通常の部分空間法では各軸に対する重みは非負である．これに対して SSLS-SVM により最適化された重みの中でクラス i の負の重みをもつ軸の類似度 $f_{ik}(\mathbf{x})$ はある入力がクラス i のデータで場合より他のクラスのデータである場合の方が大きな値を示しやすいことを意味している．そこで, 重みが負である軸が有用であるか調べるために重みが負となる軸を削除する特徴選択を行う．まず, $\kappa = 99.9(\%)$ とおき, 式 (1) により求めた固有ベクトルの中から軸を選択し, これらを用いて式 (4), (5) の最適化問題を解く．これにより各固有ベクトルにおける重みが決定でき, これらの重みの中で正の重みをとる固有ベクトルのみを選択し, それらを用いて再び式 (4), (5) の最適化問題を解く．この処理を全重みが正となるまで繰り返す．

3.2 重みの絶対値に基づく SSLS-SVM の特徴選択

3.1 節では, 重みが負となる軸を削除することにより特徴選択を行った．しかしながら, 重みが負の軸においてもその絶対値が大きければ大きいほど, 識別により大きな影響を及ぼすが悪影響を及ぼすとは限らない．識別の観点から最適化した重みであるため識別によい影響を及ぼす可能性が高い．また, 重みの絶対値が小さければ小さいほど識別に影響をあまり与えない軸であり, 識別の観点から冗長な軸であるといえる．そこで, 重みの絶対値が小さい軸を削除することで特徴選択を行う．このとき, 選択となる絶対値の閾値 θ を以下に定義する．

$$\theta_i = \sqrt{T}\sum_{j=1}^{r_i} |w_{ij}| \tag{10}$$

ここで T は閾値の値を調整するパラメータを表す．また, $|w_{ij}| > \theta_i$ に対応する軸のみを選択し, 全ての重みが閾値以上になるまで学習を繰り返す．

3.3 負の重みを許す SSLP-SVM

前述に示したように, SSLP-SVM では学習過程で同時に特徴選択を行うため, 前節までに SSLS-SVM に適用した特徴選択を行う必要はない．しかしながら, 従来の SSLP-SVM は最適化問題に非負制約が設けてあり, SSLS-SVM と同様に識別において重要な軸を削除している恐れがある．そこで, SSLP-SVM の非負制約を省くことにより負の重みを許す手法を提案する．負の重みを許すために最適化問題 (6)-(9) から制約条件 (9) を省く．

5 節の計算機実験により SSLP-SVM において従来の非負制約を設けた手法と本手法を比較することにより, 非負制約を設けるべきかどうかを確認する．

4. 錘制約部分空間法に基づくサポートベクトルマシン

4.1 CRS における決定関数

CRS では各クラスのデータクラスごとに一部の凸錘の領域にのみ密集して存在していると考え, その領域をクラス部分空間とおき, 入力 $\mathbf{x}$ とその凸錘クラス部分空間との角度を類似度として識別を行う手法である．通常の CRS では以下のように入力 $\mathbf{x}$ とクラス i 部分空間への正射影ベクトルとの角度 $\beta_i(\mathbf{x})$ を決定関数 $D_i(\mathbf{x})$ として定義する．

$$D_i(\mathbf{x}) = \beta_i(\mathbf{x}) = \arcsin\left(\sqrt{\min_{\alpha_{ij}\geq 0} \|\mathbf{x} - \sum_{j\in X_i} \alpha_{ij}\mathbf{x}_j\|^2 / \|\mathbf{x}\|}\right) \tag{11}$$

ここで, α_{ij} は教師ベクトル $\mathbf{x}_j$ における係数を表す．この決定関数 $D_i(\mathbf{x})$ は未知の入力がある度に非負最小自乗法を用いて係数ベクトルを計算し決定する．このため, 大規模データにおいて識別コストが膨大となる．また, 通常のパターン認識の問題では線形分離可能な問題はほとんど現実にはなく, 非線形問題であり, 各クラス凸錘部分空間同士で重なりがあるとき, その領域内では対応するクラスの決定関数 $D_i(\mathbf{x})$ が 0 となるため未分類領域となる．

4.2 SVM による重みの最適化

CRS の問題点を解消するために重複する領域を小さくするためカーネル法を用い, さらに SS-SVM と同様にマージン最大化と誤認識最小化の SVM の概念を CRS に導入する CRS-SVM を提案する．

はじめに錘部分空間の重複を小さくするため高次元特徴空間への写像関数 $\mathbf{g}(\mathbf{x})$ を用いて入力空間から高次元特徴空間に写像する．各クラスの部分空間を張る N_i $(i=1,...,n)$ 個の一次独立なデータ $\mathbf{x}_{ik_j}$ $(j=1,...,N_i)$ のみを用いて次式のように変更できる．各クラスごとの教師ベクトルそのものをクラス部分空間の軸する．またこの空間は一次独立なデータのみを用いた場合と等価になるため, 次元数を削減することができる．次に SS-SVM の決定関数を以下の各クラスの N_i $(i=1,...,n)$ 個の一次独立なデータ $\mathbf{x}_{ik_j}$ $(j=1,...,N_i)$ である軸への重み付けされた射影長の総和の形に変更する．

$$D_i(\mathbf{x}) = \sum_{j=1}^{N_i} \frac{w_{ij} H(\mathbf{x}, \mathbf{x}_{ik_j})}{\|\mathbf{g}(\mathbf{x})\| \|\mathbf{g}(\mathbf{x}_{ik_j})\|} \tag{12}$$

ただし, w_{ij} はクラス i 部分空間の j 番目の軸における重みである．ここで, 高次元特徴空間上のベクトル $\mathbf{x}$ のノルムは $\|\mathbf{g}(\mathbf{x})\| = \sqrt{H(\mathbf{x},\mathbf{x})}$ と容易に求めることができる．また類似度 $\mathbf{f}_i(\mathbf{x})$ は

$$\mathbf{f}_i(\mathbf{x}) = \left(\frac{H(\mathbf{x}, \mathbf{x}_{ik_1})}{\|\mathbf{g}(\mathbf{x})\| \|\mathbf{g}(\mathbf{x}_{ik_1})\|}, ..., \frac{H(\mathbf{x}, \mathbf{x}_{ik_{N_i}})}{\|\mathbf{g}(\mathbf{x})\| \|\mathbf{g}(\mathbf{x}_{ik_{N_i}})\|} \right)^T \tag{13}$$

と定義でき, 決定関数を式 (3) と同様に表せる．ここで重み w_{ij} にある入力 $\mathbf{x}_j$ の決定関数 $D_{y_j}(\mathbf{x}_j)$ が他のクラスの決定関数 $D_i(\mathbf{x}_j)(i \neq y_j)$ と比べて大きくなるような制約条件を設け, マージン最大化と誤認識最小化の概念の基に SVM を用いて SS-SVM と同様に最適化を行うことで識別の観点から最適な値を設定でき, 入力ごとに非負最小自乗法を解く必要がなくなるため, CRS に比べ識別コストを大幅に削減することができる．また他のクラスの決定関数の値と等しくなる領域以外で未分類領域は存在しない．ここで, SS-SVM と同様に用いる SVM のモデルとして LS-SVM と LP-SVM を用い, それぞれの手法を CRSLS-SVM, CRSLP-SVM とよぶ．

5 節では前節に提案手法した特徴選択を行った CRSLS-SVM と非負制約を設ける CRSLP-SVM と非負制約を省いた CRSLP-SVM において実験を行い比較・評価を行う．

5. 実　　験

本節では, 2 クラス識別問題においてベンチマークデータ citecit:12,fraunhofer を用いた計算機実験を行い, 従来の特徴選択と提案した特徴選択を用いた SS-SVM と CRS-SVM と SVM の比較・評価を行った．それぞれの手法において, 必要となるパラメータ κ, γ, C, T は五分割交差検定を用いて決定する．

以下, 従来の寄与率により軸を決定する SSLS-SVM を SSLS (A), 非負の重みを持つ軸のみの選択を繰り返す手法を用いた SSLS-SVM, CRSLS-SVM を SSLS (P), CRSLS (P), 重みの絶対値を基にした特徴選択を用いた SSLS-SVM, CRSLS-SVM を SSLS (Ab), CRSLS (Ab), 非負制約を設けた SSLP-SVM, CRSLP-SVM を SSLS (P), CRSLP (P), 非負制約を省いた SSLP-SVM, CRSLP-SVM を SSLP (N), CRSLP (N) と呼ぶ．

5.1 重みに基づく特徴選択と負の重みの重要性の評価

従来の SSLS (A), SSLP (P) と SSLS (P), SSLS (Ab), SSLP (N) を比較・評価する．表 1 にテストデータセットにおける SSLS (A), SSLS (P), SSLS (Ab), SSLP (P), SSLP (N) の平均認識率と標準偏差を示す．ここで, 各データセットにおいて最大となる平均認識率を太字で示し, 従来手法である SSLS (A) または SSLP (P) と比べ, Welch の t 検定 (有意水準：5%) により有意性を示した各データセットにおける提案手法の結果の前にアスタリスク (*), 同等であると示した結果の前に三角印 (△) を記す．

表 1 より, SSLS (P) と SSLS (Ab) は大半のデータセットにおいて SSLS (A) と比べ有意性を示しており, そうでないデータセットにおいてもほぼ同等の結果を示している．これにより, 重みを基にした特徴選択手法が識別において優れているといえる．特に SSLS (Ab) は SSLS (P) と比べても大半のデータセットにおいて高い認識率を示しており, SSLS-SVM において重みが負となる軸は識別の観点から必要であるといえる．また, SSLP (N) は SSLP (P) と比べて大半のデータセットにおいて高い認識率を示しており, S SSLP-SVM においても重みが負となる軸は識別の観点から識別の観点から必要な可能性があるといえる．

表 1 SS-SVM のテストデータにおける認識率 (%) と標準偏差の比較
Table 1 Comparison of average recognition (%) and their standard deviations of test data set for SS-SVMs

Data	SSLS (A)	SSLS (P)	SSLS (Ab)	SSLP (P)	SSLP (N)
Banana	88.9±0.6	*89.3±0.4	***89.5**±0.4	89.0±0.6	*89.3±0.5
B.cancer	**75.1**±4.4	73.6±1.6	△74.6±4.4	73.3±4.6	68.6±4.7
Diabetis	72.0±2.3	***75.2**±2.0	*75.1±2.2	73.5±2.0	*74.7±2.0
German	73.8±2.2	***75.2**±2.3	△74.4±2.1	70.9±6.6	△71.1±2.3
Heart	82.6±3.7	*83.9±3.3	*83.9±3.2	83.1±3.8	△**84.0**±3.4
Image	95.4±0.6	*96.3±0.7	***97.0**±0.6	95.1±1.0	△95.7±0.9
Ringnorm	97.6±0.3	96.6±0.4	*98.2±0.2	98.2±0.2	***98.5**±0.1
F.solar	**66.9**±1.6	65.8±1.8	△66.5±1.6	64.7±1.9	*66.7±1.8
Splice	86.2±1.0	*88.9±0.6	***89.0**±0.5	88.2±0.6	*88.9±0.5
Thyroid	95.4±2.1	△95.3±2.3	*96.2±1.9	**96.3**±2.2	95.4±2.7
Titanic	76.9±1.0	△77.0±1.5	***77.3**±0.3	76.8±1.1	△76.9±1.0
Twonorm	**97.7**±0.1	97.6±0.1	△**97.7**±0.1	97.6±0.2	△97.6±0.1
Waveform	89.1±0.8	*89.9±0.5	***90.0**±0.5	89.3±0.7	△89.5±0.6

5.2 錘制約部分空間法に基づくサポートベクトルマシンの評価

5.2.1 テストデータにおける認識率の比較・評価

表 2 に CRSLS (P), CRSLS (Ab), CRSLP(P), CRSLP (N) のテストデータセットにおける平均認識率と標準偏差を示す．ここで, 各データセットにおいて最大となる平均認識率を太字で示し, 表 1 の SS-SVM のそれぞれの結果に比べ, Welch の t 検定 (有意水準：5%) により有意性を示した各データセットにおけるそれぞれの用いている特徴選択手法に対応する CRS-SVM の結果の前にアスタリスク (*), 同等であると示した結果の前に三角印 (△) を記す．

表 2 より CRSLS (P) は SSLS (P) と比べ, 11 個のデータセットにおいて同等以上の結果を示しており, CRSLS (P) が SSLS (P) よりも高い汎化能力を示すことが得られた．また CRSLS (Ab) は SSLS (Ab) と比べ, 大半のデータセットにおいて同等の結果を示しており, 劣る結果を示したデータセットにおいても平均認識率, 標準偏差ともにそれほど差がなくほぼ同等であるといえる．また, CRSLS (Ab) は 10 個のデータセットにおいて平均認識率が最大となり, その他の CRS-SVM と比べて最も汎化能力がよいといえる．また, CRSLP (P) は SSLP (P) と比べてほぼ同等の性能を示したが, CRSLP (N) は SSLP (N) と比べて若干劣る結果となった．SS-SVM と異なり CRSLS（P) と CRSLS (Ab), または CRSLP (P) と CRSLP (N) は大半のデータセットにおいてほぼ同等の識別能力を示した．この原因として, カーネル法を用いることによって錘部分空間同士の重なりが小さくなり, 重みが負となる軸は他のクラスに属する入力における類似度もそれほど大きくならなくなり, 識別にあまり影響を与えていないことが考えられる．

表 2 CRS-SVM のテストデータにおける平均認識率 (%) と標準偏差
Table 2 Average recognition (%) and their standard deviations of test data set for CRS-SVMs

Data	CRSLS (P)	CRSLS (Ab)	CRSLP (P)	CRSLP (N)
Banana	***89.5**±0.4	△**89.5**±0.4	△89.1±0.6	89.1±0.5
B.cancer	△73.9±4.7	△**74.1**±4.5	69.7±5.1	△69.0±4.5
Diabetis	△**75.0**±2.0	△**75.0**±2.0	*74.3±2.8	△75.0±2.1
German	74.3±2.1	△**74.7**±2.5	△71.6±2.6	△71.6±2.5
Heart	△84.0±3.3	△84.0.±3.2	***84.5**±3.1	△**84.5**±3.1
Image	*96.8±0.5	△**97.0**±0.6	*96.2±0.5	△95.8±0.6
Ringnorm	*96.8±0.4	96.7±0.4	***98.3**±0.2	98.0±0.3
F.solar	*66.5±1.5	△66.5±1.6	***67.3**±1.8	*67.2±1.7
Splice	△**89.1**±0.7	△**89.1**±0.6	87.7±0.9	87.7±0.9
Thyroid	*95.9±1.9	△**96.0**±2.0	94.5±2.8	△94.9±2.3
Titanic	***77.3**±0.4	△**77.3**±0.3	76.3±2.1	76.3±2.1
Twonorm	△**97.6**±0.1	**97.6**±0.1	97.3±0.2	97.3±0.2
Waveform	89.7±0.5	**89.8**±0.5	88.6±0.5	88.8±0.7

5.2.2 学習時間の比較・評価

表 3 に SSLS (P), SSLS (Ab), CRSLS (P), CRSLS (Ab), SSLP (N), CRSLP (N) の各データセットにおける 1 ファイルあたりの平均学習時間をそれぞれ示す．

表 3 より, CRSLS (P), CRSLS (Ab) において SSLS (P), SSLS (Ab) と比べて全データセットにおいて学習時間が短縮されている．また本実験で用いるデータの中でデータ数の多い German データセット, Image データセット, Splice データセットにおいても学習時間が約 $\frac{2}{3}$ に削減できており, 大規模データにおいても有効であることがいえる．また CRSLP (N) においても SSLP (N) と比べて大半のデータセットにおいて学習時間を大幅に削減できている．しかしながら, Image データセットや Splice データセットの大規模データセットにおいては SSLP と比べて学習時間が増加している．この原因として重みの最適化の際に用いている単体法による計算コストが膨大になるためであると考えられる．

5.3 SVM との比較

本小節では SS-SVM, CRS-SVM において高い汎化能力を示した SSLS (Ab), CRSLS (Ab) を SVM と比較する．表 4 にそれぞれの手法におけるテストデータセットにおける平均認識率と標準偏差を示す．ここで, 各データセットにおいて最大となる平均認識率を太字で示し, SVM と比べ, Welch の t 検定 (有意水準：5%) により有意性を示した各データセットにおける SSLS (Ab) と CRSLS (Ab) の結果の前にアスタリスク (*), 同等であると示した結果の前に三角印 (△) を記した．

表 3 学習時間 (s) の比較
Table 3 Comparison of training time (s)

Data	SSLS (P)	SSLS (Ab)	SSLP (N)	CRSLS (P)	CRSLS (Ab)	CRSLP (N)
Banana	0.39	0.39	4.59	0.29	0.29	4.21
B.cancer	0.23	0.23	0.87	0.13	0.12	0.61
Diabetis	0.64	1.33	8.59	0.39	0.57	4.49
German	9.38	9.67	65.4	6.45	6.51	32.2
Heart	0.13	0.13	0.55	0.09	0.09	0.46
Image	53.1	46.6	406	36.1	39.8	624
Ringnorm	1.76	1.78	11.9	1.25	1.26	4.5
F.solar	1.24	1.25	19.9	0.87	0.98	14.5
Splice	26.7	25.9	208	18.7	19.2	538
Thyroid	0.08	0.06	0.19	0.05	0.05	0.0.9
Titanic	0.02	0.02	0.18	0.02	0.02	0.09
Twonorm	1.73	1.73	9.34	1.24	1.25	2.79
Waveform	1.92	1.86	11.2	1.31	1.21	5.91

表 4 より, SVM と比べ SSLS (Ab) において 5 個のデータセットにおいて有意性を示し, 5 個のデータセットにおいて同等の結果を示している．これより SSLS (Ab) は SVM と比べ, 同等以上の高い汎化能力を持つ識別器であるといえる．また, SVM と比べ CRSLS (Ab) において 4 個のデータセットにおいて有意性を示し, 5 個のデータセットにおいて同等の結果を示している．これより CRSLS (Ab) は SVM と比べ, ほぼ同等の汎化能力を持つ識別器であるといえ, 前節における学習時間の比較を考慮すると CRSLS-SVM が高速かつ高い汎化能力を持つ識別器であることがいえる．

6. ま と め

本論文では, 1) SSLS-SVM における最適な重みに基づく特徴選択, 2) 部分空間生成における学習コストの削減のための CRS-SVM を提案した．1) の手法では, SSLS-SVM により最適化された軸の重みに基づき条件を満たすまで選択処理を繰り返す．その条件として二つ設定した．一つ目は非負制約であり, 最適化の度に重みが正となる軸のみを選択し続ける．二つ目は, 全ての重みの絶対値が閾値より大きくなることである．ここでは, 最適化と同時に特徴選択を行う SSLP-SVM から非負制約を省いた手法を考慮した手法も提案した．2) の手法では, 各クラスの教師ベクトルをそれぞれ各クラス部分空間を構成する軸とすることで CRS に SS-SVM と同様に各軸における類似度に対し識別において最適な重み付けを行った．

ベンチマークデータを用いた計算機実験により 1), 2) の提案手法の有効性を確認した．

謝辞 計算機実験において協力していただいた田尻康之氏 (神戸大学) に感謝します．

表 4 テストデータセットにおける SVM と SSLS (Ab) と CRSLS (Ab) の平均認識率 (%) と標準偏差
Table 4 average recognition (%) and their standard deviations of test data set for SVM, SSLS (Ab), and CRSLS (Ab)

Data	SSLS (Ab)	CRSLS (Ab)	SVM
Banana	$^{*}\mathbf{89.5}\pm0.4$	$^{*}\mathbf{89.5}\pm0.4$	89.3 ± 0.5
B.cancer	$^{*}\mathbf{74.6}\pm4.4$	$^{*}74.1\pm4.5$	72.4 ± 4.6
Diabetis	75.1 ± 2.2	75.0 ± 2.0	$\mathbf{76.3}\pm1.8$
German	74.1 ± 2.1	74.7 ± 2.5	$\mathbf{76.2}\pm2.2$
Heart	$^{\triangle}83.9\pm3.2$	$^{\triangle}\mathbf{84.0}\pm3.3$	83.7 ± 3.4
Image	$^{\triangle}97.0\pm0.6$	$^{\triangle}97.0\pm0.6$	$\mathbf{97.3}\pm0.4$
Ringnorm	$^{*}\mathbf{98.2}\pm0.2$	96.7 ± 0.4	97.8 ± 0.3
F.solar	66.5 ± 1.8	66.5 ± 1.6	$\mathbf{67.6}\pm1.7$
Splice	$^{\triangle}89.0\pm0.5$	$^{\triangle}89.1\pm0.6$	$\mathbf{89.2}\pm0.7$
Thyroid	$^{\triangle}\mathbf{96.2}\pm1.9$	$^{\triangle}96.0\pm2.0$	96.1 ± 2.0
Titanic	$^{*}\mathbf{77.3}\pm0.3$	$^{*}\mathbf{77.3}\pm0.3$	77.2 ± 1.1
Twonorm	$^{*}\mathbf{97.7}\pm0.1$	$^{\triangle}97.6\pm0.1$	97.6 ± 0.1
Waveform	$^{\triangle}\mathbf{90.0}\pm0.5$	89.8 ± 0.5	$\mathbf{90.0}\pm0.4$

参 考 文 献

1) S. Watanabe and N. Pakvasa, “Subspace methods of pattern recognition,” *Proc. 1st IJCPR*, pp. 283–328, 1973.
2) E. Oja, *Subspace Methods of Pattern Recognition*, Research Studies Press, 1983.
3) S. Abe, *Support Vector Machines for Pattern Classification*, Springer, 2005.
4) T. Kitamura, S. Takeuchi, S. Abe, and K. Fukui, “Subspace-based support vector machines for pattern classification,” *Neural Networks,* Vol. 22, pp. 558-567, 2009.
5) B. Shölkopf, S. Mika, C.J.C. Burges, P. Knirsch, K.R. Müller, G. Rätsch, and A.J. Smola, “Input space vs. feature space in kernel-based methods,” *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 1000–1017, 1999.
6) B. Shölkopf, A.J. Smola, and K.R. Müller, “Nonlinear component analysis as a kernel eigenvalue problem,” *Neural Computation*, vol. 10, pp. 1299–1319, 1998.
7) 小林 匠, 大津 展之, “パターン認識のための錘制約部分空間法,” 電子情報通信学会論文誌 (D), Vol. J92-D, No. 1, pp. 104-111, 2009.
8) S. Abe, “Sparse least squares support vector training in the reduced empirical feature space,” *Pattern Analysis and Applications*, vol. 10, no. 3, pp. 203–214, 2007.
9) H. Xiong, M.N.S. Swamy and M.O. Ahmad, “Optimizing the kernel in the empirical feature space,” *IEEE Transactions on Neural Networks*, vol. 16, no. 2, pp. 460–474, 2005.
10) http://ida.first.fraunhofer.de/projects/bench/benchmarks.htm.