

カテゴリ別特徴抽出による高速スパース 最小自乗サポートベクトルマシン

北村 拓也^{1,a)} 関根 崇正¹ 塚越 勇規¹

受付日 2014年5月28日, 再受付日 2014年7月18日,
採録日 2014年9月5日

概要: 本論文では, カテゴリ標本特徴空間を用いた高速スパース最小自乗サポートベクトルマシン (FSLs-SVM) を提案する. FSLs-SVM では, カテゴリごとの教師ベクトルから一次独立な教師ベクトルを選択し, カテゴリごとにそれらの教師ベクトルを基底ベクトルとした標本特徴空間を生成する. これらの標本特徴空間上に全教師ベクトルを写像し, それぞれ識別基準を決定する. このとき, これらの識別基準は, 各カテゴリの一次独立な教師ベクトルのみを用いて決定できるため, 通常の LS-SVM とは異なり, 解にスパース性を付与できる. また, カテゴリごとに独立して一次独立な教師ベクトルの選択を行えるため, 選択における計算コストは小さく, 全データから選択する従来手法に比べて高速な学習が可能となる. ベンチマークデータセットを用いた計算機実験により提案手法の有効性を示す.

キーワード: 機械学習, パターン認識, サポートベクトルマシン, スパース

Fast Sparse Least Squares Support Vector Training by Feature Extraction for Each Category

TAKUYA KITAMURA^{1,a)} TAKAMASA SEKINE¹ YUKI TSUKAGOSHI¹

Received: May 28, 2014, Revised: July 18, 2014,
Accepted: September 5, 2014

Abstract: In this paper, we propose fast sparse least squares support vector training in category empirical feature spaces. In the proposed method, the empirical feature space for each category is generated by the independent training data. The discriminant criteria are determined in these empirical feature spaces. Then, the solution is sparse because the discriminant criterion for each category is determined by the independent training data. And, because the independent training data are selected independently for each class, the computational cost of the selection is lower than that of the conventional method which selects from all the training data. Using benchmark data sets, we evaluate the effectiveness of the proposed method over the conventional methods.

Keywords: machine learning, pattern recognition, support vector machine, sparse

1. はじめに

最小自乗サポートベクトルマシン (LS-SVM: least squares support vector machine) [1], [2] とはサポートベクトルマシン (SVM: support vector machine) [2], [3], [4] のモデルの1つである. 通常の SVM における最適化問題が二次計画問題であるが, LS-SVM においては連立一次方程式とな

り, 計算コストの削減が期待できる. さらにパターン認識問題 [5] において, LS-SVM は通常の SVM と比べて汎化能力の観点から同等程度の精度を示すため, 広く用いられている. しかしながら LS-SVM では, 全教師ベクトルをサポートベクトル (SV: support vector) として取り扱うため, 通常の SVM の利点である解のスパース性が欠如している. そのため, LS-SVM に対してスパース性を付与させたスパース LS-SVM (SLS-SVM: sparse least squares support vector machine) [2], [6], [7], [8], [9], [10], [11], [12] の研究が広く行われている. Abe が開発した SLS-SVM [2], [11]

¹ 富山高等専門学校
National Institute of Technology, Toyama College, Toyama
939-8630, Japan

a) kitamura@nc-toyama.ac.jp

では、全教師ベクトルから一次独立な教師ベクトルを選択し、それらを基底ベクトルとして構成した標本特徴空間 [12] 上において LS-SVM を用いることにより、一次独立な教師ベクトルのみを SV と見なし、解にスパース性を付与している。このとき、一次独立な教師ベクトルの選択はコレスキー分解 [13] または線形判別分析 (LDA: linear discriminant analysis) [14] を用いている。以後、“従来の SLS-SVM” はこの SLS-SVM を表す。しかしながら、一次独立な教師ベクトルの選択における計算コストが大きくなり、計算機実験により通常の LS-SVM と比べて学習時間が大幅に増加する傾向にあることが確認されている [2], [11].

そこで本論文では、一次独立な教師ベクトルの選択に要する計算コストを削減した高速 SLS-SVM (FSLs-SVM: fast sparse least squares support vector machine) を提案する。FSLs-SVM では、従来の SLS-SVM とは異なり、カテゴリごとに一次独立な教師ベクトルを選択する。全教師ベクトル数を M 、カテゴリ数を n 、各カテゴリにおける教師ベクトル数を M_i ($i = 1, \dots, n$) としたとき、従来の SLS-SVM コレスキー分解を用いた選択に要する計算コストは $O(M^3)$ となることに対して、FSLs-SVM では $O(\sum_{i=1}^n M_i^3)$ となる。このため、各カテゴリの教師ベクトル数が等しい場合、計算コストが n^2 倍小さくなり、特にカテゴリ数が多いほど有効であるといえる。ここで、各カテゴリにおいて一次独立な教師ベクトルを基底ベクトルとした標本特徴空間を生成し、それらの空間をカテゴリ別標本特徴空間とよぶ。次に、各カテゴリ別標本特徴空間上でそれぞれの識別基準を LS-SVM により決定する。ここで、多クラス問題に拡張するため、広く用いられている 1 対多 (OAA: one-against-all) 方式 [15] とペアワイズ方式を用いる。OAA 方式では各カテゴリ別標本特徴空間上において他カテゴリの教師ベクトルと分類するための識別基準を 1 つずつ設定し、ペアワイズ方式では他のカテゴリのうち、1 つのカテゴリに属する教師ベクトルと分類する識別基準を 1 つずつ設定する。すなわち、ある 1 つのカテゴリ別標本特徴空間上において、OAA 方式では 1 個、ペアワイズ方式では $(n-1)$ 個の識別基準を決定する。また、標本特徴空間の次元数は選択された一次独立な教師ベクトル数と等しくなるため、FSLs-SVM における各カテゴリ別標本特徴空間の次元数は従来の SLS-SVM における標本特徴空間の次元数に比べて小さくなる。そのため、各識別基準を決定するために求めるべき変数の数も少なく、最適化問題における計算コストの削減も期待できる。次に、各カテゴリ別標本特徴空間を他カテゴリに属する教師ベクトルを考慮せず生成することによる情報の欠損を防ぐため、各空間上における識別基準を線形結合することにより新たに識別基準を決定する。すなわち、各空間上において決定された識別基準を 1 つの特徴量として取り扱い、それらの特徴量に重みづけを行い線形結合することにより他のカテゴリ別

標本特徴空間を考慮した識別基準を求められる。ここで識別基準の重みの最適化は OAA, ペアワイズ, 同時定式化 (AAO: all at once) [2], [16], [17] 方式を用いた LS-SVM により行う。

本論文の構成は、2 章に LS-SVM と従来の SLS-SVM について述べ、3 章に FSLs-SVM について説明し、4 章に計算機実験による結果と考察を述べる。最後に 5 章でまとめを述べる。

2. スパース最小自乗サポートベクトルマシン

2.1 LS-SVM

LS-SVM は通常の SVM と同様にマージン最大化の概念を基に識別基準を決定する。通常の SVM と比べて同等の汎化能力を持つとともに、連立一次方程式を解くことにより識別基準を決定するため、学習の高速化が期待できる。

教師ベクトル数 M の m 次元 n クラス問題において、教師ベクトルとクラスラベル $\{\mathbf{x}_j, y_j\}$ ($j = 1, \dots, M$) が与えられているとき、LS-SVM における最適化問題は以下のように定義できる。

$$\min \quad \frac{1}{2} \mathbf{w}^\top \mathbf{w} + \frac{C}{2} \sum_{j=1}^M \xi_j^2 \quad (1)$$

$$\text{s.t.} \quad z_j (\mathbf{w}^\top \mathbf{g}(\mathbf{x}_j) + b) = 1 - \xi_j \quad (2)$$

for $j = 1, \dots, M$

ここで、 \mathbf{w} , C , ξ_j , b , $\mathbf{g}(\mathbf{x})$ はそれぞれ重みベクトル、マージンパラメータ、スラック変数、バイアス項、高次元特徴空間への写像関数である。また、LS-SVM は通常の SVM と同様に 2 クラス識別器であり、 z_j は $y_j = 1$ のとき $z_j = 1$ となり、 $y_j = 2$ のとき $z_j = -1$ とする。式 (1), (2) にラグランジュ乗数 α_j ($j = 1, \dots, M$) を導入し、双対問題により最適化問題を解くと重みベクトル \mathbf{w} は以下のように求められる。

$$\mathbf{w} = \sum_{j=1}^M z_j \alpha_j \mathbf{g}(\mathbf{x}_j) \quad (3)$$

識別基準である決定関数 $D(\mathbf{x}) = \mathbf{w}^\top \mathbf{g}(\mathbf{x}) + b$ に式 (3) を代入すると、

$$D(\mathbf{x}) = \sum_{j=1}^M z_j \alpha_j K(\mathbf{x}_j, \mathbf{x}) + b \quad (4)$$

となる。ここで、 $K(\mathbf{x}, \mathbf{x}')$ は高次元特徴空間上における \mathbf{x} と \mathbf{x}' の内積値であり、カーネル関数とよぶ。

LS-SVM において、 $\alpha_j \neq 0$ ($j = 1, \dots, M$) となり、式 (4) において、全教師ベクトルを用いる必要がある。そのため、図 1 の 2 次元 2 クラス問題を例としてあげるように、SVM においては識別面付近または誤認識しうる教師ベクトルが SV となることに対して、LS-SVM ではすべての教師ベクトルを SV と見なす。すなわち、解のスパース性が欠損しているといえる。

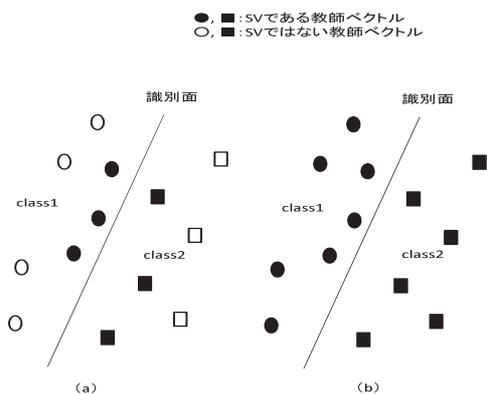


図 1 SVM と LS-SVM における SV. (a) SVM, (b) LS-SVM
 Fig. 1 Support vectors for SVM and LS-SVM. (a) SVM, (b) LS-SVM.

2.2 SLS-SVM

LS-SVM のスパース性の欠損を解消するため、広く SLS-SVM の研究がなされている。代表的な 1 つとして、Suykens らの提案した *pruning* 法を用いた SLS-SVM [6], [8] があげられる。本手法では、全教師ベクトルを用いて LS-SVM を学習し、ラグランジュ乗数 α_j ($j = 1, \dots, M$) の絶対値を基に一部の教師ベクトルを選択し、それらを用いて再び LS-SVM により識別基準を決定する。しかしながら、この手法では解にスパース性を与えられるが、全教師ベクトルを考慮した識別基準となっておらず、通常の LS-SVM の識別基準と大きく異なりうる。文献 [9], [10] で提案されている SLS-SVM においても、*pruning* 法を基にスパース性を付与しているため同様のことがいえる。

そこで Abe が提案した SLS-SVM [2], [11] では、標本特徴空間 [12] 上で LS-SVM により学習させることによりスパース性を付与している。全教師ベクトル \mathbf{x}_j ($j = 1, \dots, M$) から一次独立な教師ベクトル \mathbf{x}'_j ($j = 1, \dots, N$) をコレスキー分解 [13] または LDA [14] により選択し、それらを基底ベクトルとして標本特徴空間を生成する。ここで、 N は一次独立な教師ベクトル数である。このとき、標本特徴空間への写像関数 $\mathbf{h}(\mathbf{x})$ は以下のように求められる。

$$\mathbf{h}(\mathbf{x}) = (K(\mathbf{x}, \mathbf{x}'_1), \dots, K(\mathbf{x}, \mathbf{x}'_N))^T \quad (5)$$

標本特徴空間上と高次元特徴空間上の内積値は異なり、すなわちデータ分布は変化するが、適切なカーネル関数の種類とパラメータ値を決定することにより、汎化能力への影響は考慮しなくてもよいことが計算機実験により確認されている [2], [11]。

式 (5) で求められた標本特徴空間への写像関数 $\mathbf{h}(\mathbf{x})$ を用いて以下の最適化問題を解く。

$$\min \quad \frac{1}{2} \mathbf{v}^T \mathbf{v} + \frac{C}{2} \sum_{j=1}^M \xi_j^2 \quad (6)$$

$$\text{s.t.} \quad z_j (\mathbf{v}^T \mathbf{h}(\mathbf{x}_j) + b) = 1 - \xi_j \quad (7)$$

for $j = 1, \dots, M$

ここで、 \mathbf{v} は N 次元重みベクトルである。高次元特徴空間とは異なり、標本特徴空間の次元数は N は有限かつ全教師ベクトル数 M 以下となる。すなわち、主問題は双対問題に比べて計算コストが小さくなるため、主問題により最適解を求める。このとき、求められた重みベクトル \mathbf{v} 、バイアス項 b を用いて、決定関数は以下のように求められる。

$$D(\mathbf{x}) = \mathbf{v}^T \mathbf{h}(\mathbf{x}) + b \quad (8)$$

$$= \mathbf{v}^T (K(\mathbf{x}, \mathbf{x}'_1), \dots, K(\mathbf{x}, \mathbf{x}'_N))^T + b$$

式 (8) より、決定関数の導出に用いられる教師ベクトルは一次独立な教師ベクトルのみであり、それらを SV と見なせる。そのため、一次従属な教師ベクトルが存在する限り、SV 数は通常の LS-SVM と比べて少なく、スパース性を付与できるといえる。また、Suykens らの SLS-SVM [6], [7], [8] と異なり、全教師ベクトルを考慮したうえで識別基準を決定している。しかしながら、スパース性を付与できるが、一次独立な教師ベクトルの選択による計算コストの増加により、通常の LS-SVM に比べて全体の学習コストが大きくなる。本論文では、阿部氏の SLS-SVM [2], [11] を“従来の SLS-SVM” と定義する。

3. カテゴリ標本特徴空間による高速学習法

3.1 カテゴリ別選択による計算コストの削減

従来の SLS-SVM において、一次独立な教師ベクトルの選択に要する計算コストは $O(M^3)$ となる。すなわち、SV 数にかかわらず、与えられた教師ベクトル数が多ければ多いほど計算コストは大きくなる。ここで、選択には LDA と比べて計算コストの小さいコレスキー分解を用いた場合を示している。

そこで、カテゴリ別に一次独立な教師ベクトルを選択することにより学習コストの削減を図る。カテゴリごとの教師ベクトル数を M_i ($i = 1, \dots, n$) としたとき、クラス i における一次独立な教師ベクトルの選択に要する計算コストは $O(M_i^3)$ となる。すなわち、全カテゴリにおける選択に要する計算コストの総和は $O(\sum_{i=1}^n M_i^3)$ となる。 $M = \sum_{i=1}^n M_i$ となるため、 $M^3 \geq \sum_{i=1}^n M_i^3$ となり、計算コストが削減される。各カテゴリの教師ベクトル数が等しい場合、選択における計算コストは従来の SLS-SVM の $\frac{1}{n^2}$ となり、カテゴリ数が多ければ多いほど削減割合が大きくなる。また、1 つの最適化問題を解くために必要な計算量は、従来の SLS-SVM では N^3 、提案手法においては N_i^3 ($i = 1, \dots, n$) となる。ここで、 N_i はクラス i の一次独立な教師ベクトル数である。 $N \leq \sum_{i=1}^n N_i$ であるが、 $N \geq N_i$ であるため、1 つの最適化問題を解くために必要な計算量も提案手法は従来の SLS-SVM に比べて削減される。

3.2 FSLs-SVM

FSLs-SVMでは、3.1節におけるカテゴリ別に生成された標本特徴空間上に写像し、最適化問題を解くことにより識別基準を決定する。式(5)と同様にクラス*i*標本特徴空間への写像関数 $\mathbf{h}_i(\mathbf{x})$ を以下のように定義する。

$$\mathbf{h}_i(\mathbf{x}) = (K(\mathbf{x}, \mathbf{x}'_{i1}), \dots, K(\mathbf{x}, \mathbf{x}'_{iN_i}))^\top \quad (9)$$

for $i = 1, \dots, n$

ここで、 \mathbf{x}'_{ik} ($k = 1, \dots, N_i$) はクラス*i*標本特徴空間における k 番目の一次独立な教師ベクトルである。このとき、クラス*i*標本特徴空間上における最適化問題は以下のように定義できる。

$$\min \quad \frac{1}{2} \mathbf{t}_i^\top \mathbf{t}_i + \frac{C}{2} \sum_{j=1}^M \xi_{ij}^2 \quad (10)$$

$$\text{s.t.} \quad z_j(\mathbf{t}_i^\top \mathbf{h}_i(\mathbf{x}_j) + b_i) = 1 - \xi_{ij} \quad (11)$$

for $i = 1, \dots, n, j = 1, \dots, M$

ここで、 \mathbf{t}_i , b_i , ξ_{ij} はそれぞれクラス*i*標本特徴空間上で求める N_i 次元重みベクトル、バイアス項、スラック変数である。また、 $y_j = i$ のとき $z_j = 1$, $y_j \neq i$ のとき $z_j = -1$ である。クラス*i* ($i = 1, \dots, n$) 標本特徴空間上での決定関数 $D_i(\mathbf{x})$ は

$$D_i(\mathbf{x}) = \mathbf{t}_i^\top \mathbf{h}_i(\mathbf{x}) + b_i \quad \text{for } i = 1, \dots, n \quad (12)$$

である。

それぞれのカテゴリ別標本特徴空間へ対応するカテゴリに属する教師ベクトルを写像した場合、それらのデータ分布は元空間におけるデータ分布と比べて変化は小さい。しかしながら、これらはカテゴリごとに独立して生成しているため、他のカテゴリに属する教師ベクトルを写像した場合、それらのデータ分布は元空間におけるデータ分布と比べて変化が大きくなりうる。すなわち、他のカテゴリの教師ベクトルは標本特徴空間へ写像することにより情報の欠損が起こりうる。このとき、欠損された情報が識別の観

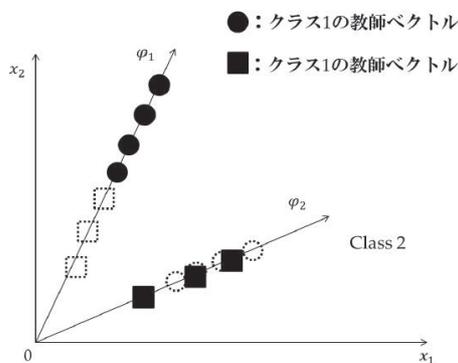


図2 カテゴリ別標本特徴空間生成による情報の欠損

Fig. 2 The loss of information by generating the empirical feature space for each class.

点から重要である場合、汎化能力の劣化につながりうる。図2に2次元2クラス問題における各カテゴリにおいて1次元標本特徴空間へ写像された場合の情報の欠損の例をあげる。それぞれ破線で示されている○と□はそれぞれ他のカテゴリ別標本特徴空間へ写像された場合のクラス1, 2に属する教師ベクトルを表す。ここで、 ϕ_1 , ϕ_2 はそれぞれクラス1またはクラス2標本特徴空間を構成する基底ベクトルであり、図2の例ではそれぞれ1次元である。図2より各カテゴリ別標本特徴空間上における各カテゴリ間のマージンは元空間におけるマージンと比べて明らかに小さくなっており、特にクラス2標本特徴空間上においては識別が困難なデータ分布となっている。

そこで、各カテゴリ別標本特徴空間上で求められた決定関数に対して重みづけをし、線形統合することにより新たに全カテゴリの情報を考慮した決定関数を求める。また2クラス識別器である各種SVMにおいて多クラス問題などの複数のカテゴリを含む場合、一般的にOAA方式[2], [3], [4], [15]やペアワイズ方式[2], [3], [4]などの拡張方式を用いられる。最適化問題(6), (7)はOAA方式における最適化問題である。ペアワイズ方式では、クラス*i* ($i = 1, \dots, n$) と他のある1つのクラス*l* ($l \neq i$) との識別基準を全組合せにおいて決定するため、最適化問題(6), (7)において \mathbf{t}_i , M , ξ_{ij} , b_i ($i, l = 1, \dots, n, i \neq l$) をそれぞれ \mathbf{t}_{il} , $M_i + M_l$, ξ_{ilj} , b_{il} と置き換えて計算すればよい。ここで z_j は $y_j = i$ のとき1, $y_j = l$ のとき-1とする。

3.2.1 OAA方式における決定関数の線形結合

OAA方式を用いたFSLs-SVM[15]において新たに求めるクラス*i*とその他カテゴリを分類する決定関数 $D'_i(\mathbf{x})$ を式(12)を用いて

$$D'_i(\mathbf{x}) = \mathbf{u}_i^\top \mathbf{f}(\mathbf{x}) + b_i^e \quad (13)$$

$$\mathbf{f}(\mathbf{x}) = (D_1(\mathbf{x}), \dots, D_n(\mathbf{x}))^\top \quad (14)$$

for $i = 1, \dots, n$

とする。ここで、 \mathbf{u}_i , b_i^e は n 次元重みベクトル、バイアス項である。 \mathbf{u}_i と b_i^e はそれぞれ通常のSLS-SVMと同様に最適化問題(6), (7)の主問題を解くことにより求める。このとき求めるべき変数の数は $(n+1)$ 個となり、計算量の増加は $O(n^3)$ となるため、 n が非常に大きくならない限り、決定関数の統合による計算量の増加を考慮しなくてもよい。識別部では、ある入力 \mathbf{x} において、

$$\arg \max_{i=1, \dots, n} D'_i(\mathbf{x}) \quad (15)$$

となるクラス*i*に分類する。

3.2.2 ペアワイズ方式における決定関数の線形結合

ペアワイズ方式では、クラス*i* ($i = 1, \dots, n$) における標本特徴空間上において、クラス*i*とクラス*j* ($j \neq i, j = 1, \dots, n$) を分類する識別基準を決定するため、各カテゴリ別標本特徴空間上において $(n-1)$ 個の識

別基準を有する．また，クラス i またはクラス l 標本特徴空間上でクラス $i, l (l \neq i)$ を分類する決定関数 $D_{il}(\mathbf{x})$, $D_{li}(\mathbf{x})$ の関係は $D_{il}(\mathbf{x}) \neq D_{li}(\mathbf{x})$ となる．すなわち，計 $n(n-1)$ 個の識別基準を決定する必要がある．決定関数 $D_{il}(\mathbf{x})$ は

$$D_{il}(\mathbf{x}) = \mathbf{t}_{il}^\top \mathbf{h}_i(\mathbf{x}) + b_{il} \quad \text{for } i, l = 1, \dots, n, i \neq l \quad (16)$$

と定義できる． $D_{il}(\mathbf{x})$ はクラス i 標本特徴空間上で決定されており，クラス l に属する教師ベクトルにおいて情報の欠損が発生しうするため，クラス l 標本特徴空間上におけるクラス i と分類する決定関数 $D_{li}(\mathbf{x})$ と線形結合する．新たにクラス i, l を分類する決定関数 $D'_{il}(\mathbf{x})$ を式 (16) を用いて

$$D'_{il}(\mathbf{x}) = \mathbf{u}_{il}^\top \mathbf{f}_{il}(\mathbf{x}) + b_{il}^e \quad (17)$$

$$\mathbf{f}_{il}(\mathbf{x}) = (D_{il}(\mathbf{x}), D_{li}(\mathbf{x}))^\top \quad (18)$$

for $i, l = 1, \dots, n, i \neq l$

とする．ここで， \mathbf{u}_{il} , b_{il}^e は 2 次元重みベクトル，バイアス項である． $D'_{il}(\mathbf{x})$ は $D_{il}(\mathbf{x})$ と異なり， $D'_{il}(\mathbf{x}) = D'_{li}(\mathbf{x})$ となる． \mathbf{u}_{il} と b_{il}^e はそれぞれ通常の SLS-SVM と同様に最適化問題 (6), (7) の主問題を解くことにより求める．このとき各カテゴリ対において求めるべき変数は 2 となり，決定関数の統合による計算量の増加を考慮しなくてもよい．識別部では，ある入力 \mathbf{x} において，

$$D'_i(\mathbf{x}) = \min_{l=1, \dots, n, l \neq i} D'_{il}(\mathbf{x}) \quad \text{for } i = 1, \dots, n \quad (19)$$

とおき，式 (15) を満たすクラス i に分類する．

3.2.3 AAO 方式による決定関数の線形結合

OAA 方式は一般的にペアワイズ方式に比べて汎化能力が劣る傾向がある．ペアワイズ方式では，ある 1 つのカテゴリとその他のカテゴリのうちの 1 つの識別基準を全組合せにおいて決定しているため，より細かな識別基準となっているためである．しかしながら，決定関数の線形結合を行う前の全識別基準の決定における計算コストは $O(n \sum_{i=1}^n M_i^3)$ となり，ペアワイズ方式の $O(n(n-1) \sum_{i=1}^n M_i^3)$ と比べて $(n-1)$ 倍小さいメリットがある．そこで，決定関数の線形結合において AAO 方式 [2], [16], [17] を用いることにより汎化能力を向上させる．AAO 方式とは，全識別基準を考慮して同時に決定する多クラス問題への拡張方式である．AAO 方式はペアワイズ方式と同程度以上の汎化能力を示すが，全識別基準を同時に決定することから，一度に求める変数の数が膨大となる傾向にあり計算コストの観点から一般的に実用的ではない．しかしながら，決定関数の線形結合において，1 つの識別基準を決定するために必要な求めるべき変数の数は n であり，すなわち同時に決定する変数の数は n^2 となり，カテゴリ数が膨大でない限り計算コストの問題は考慮しなくてよい．

はじめに OAA 方式を用いてクラス $i (i = 1, \dots, n)$ における標本特徴空間上でクラス i と他カテゴリを分類するための決定関数 $D_i(\mathbf{x}) (i = 1, \dots, n)$ を求める．新たに線形結合することにより求めるクラス i とその他カテゴリを分類する決定関数 $D'_i(\mathbf{x})$ は

$$D'_i(\mathbf{x}) = \mathbf{u}_i^\top \mathbf{f}(\mathbf{x}) \quad (20)$$

for $i = 1, \dots, n$

と定義する．ここで， $\mathbf{f}(\mathbf{x})$ は式 (14) で定義できる．式 (13) と異なる点はバイアス項の有無のみであり，AAO 方式においてバイアス項を除くことにより計算を簡略化させている．AAO 方式において最適化問題は以下のように定義できる．

$$\min \quad \frac{1}{2} \sum_{i=1}^n \mathbf{u}_i^\top \mathbf{u}_i + \frac{C}{2} \sum_{j=1}^M \sum_{i=1, i \neq y_j}^n \xi_{ij}^2 \quad (21)$$

$$\text{s.t.} \quad \mathbf{u}_{y_j}^\top \mathbf{f}(\mathbf{x}_j) - \mathbf{u}_i^\top \mathbf{f}(\mathbf{x}_j) = 1 - \xi_{ij} \quad (22)$$

for $i = 1, \dots, n, j = 1, \dots, M$

上式における制約条件は， \mathbf{x}_j が属するカテゴリにおける決定関数が全カテゴリの中で最大とすることを意味している．式 (21), (22) の解は以下のように表される．

$$\begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_n \end{pmatrix} = E^{-1} \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_n \end{pmatrix} \quad (23)$$

$$E = \begin{pmatrix} A_1 & B_{12} & \cdots & B_{1n} \\ B_{21} & A_2 & & B_{2n} \\ \vdots & & \ddots & \vdots \\ B_{(n-1)1} & & & B_{(n-1)n} \\ B_{n1} & B_{n2} & \cdots & A_n \end{pmatrix} \quad (24)$$

$$A_i = \frac{1}{C} + \sum_{j=1, y_j \neq i}^M \mathbf{f}(\mathbf{x}_j) \mathbf{f}^\top(\mathbf{x}_j) \quad (25)$$

$$+ (n-1) \sum_{j=1, y_j = i}^M \mathbf{f}(\mathbf{x}_j) \mathbf{f}^\top(\mathbf{x}_j)$$

$$B_{ik} = - \sum_{j=1, y_j = i \text{ or } k}^M \mathbf{f}(\mathbf{x}_j) \mathbf{f}^\top(\mathbf{x}_j) \quad (26)$$

$$\mathbf{a}_i = \sum_{j=1}^M \mathbf{f}(\mathbf{x}_j) - \sum_{j=1, y_j = i}^M \mathbf{f}(\mathbf{x}_j) \quad (27)$$

for $i, k = 1, \dots, n$

式 (20) に式 (23) を代入することにより，決定関数 $D'_i(\mathbf{x})$ を求める．識別部では，式 (15) を満たすクラス i に分類する．

4. 計算機実験

本章では、ベンチマークデータセット [2], [15], [18], [19], [20] を用いた計算機実験により従来の SLS-SVM と FLSLS-SVM の学習時間, SV 数, 汎化能力の比較・評価を行う。さらに, それぞれの計算機実験において参考のため通常の LS-SVM との比較・評価も行う。計算機の仕様は OS: Windows 7 (64bit), CPU: Intel (R) Core (TM) i7 3.40GHz, メモリ: 16.00GB であり, 識別器の実装は Microsoft 社の Visual C を用いた。

4.1 ベンチマークデータセット

表 1, 表 2 に計算機実験に用いた 2 クラスまたは多クラスベンチマークデータセット [2], [15], [18], [19], [20] を示す。

4.1.1 2 クラス問題

表 1 において, “Inputs” は次元数, “Training” は教師データ数, “Test” はテストデータ数, “Sets” は各データセットのファイル数を示す。また, すべてのデータセットにおいて 2 クラス問題である。

4.1.2 多クラス問題

表 2 に多クラスデータセットを示し, ここで “Classes” はクラス数を表す。また, Thyroid (3) データセットと 2 クラスベンチマークデータセットにおける Thyroid データセットとの違いは 3 クラス問題の甲状腺に関するデータ

表 1 2 クラスベンチマークデータセット
Table 1 Two-class benchmark data sets.

Data	Inputs	Training	Test	Sets
Banana	2	400	4900	100
B. cancer	9	200	77	100
Diabetes	8	468	300	100
German	20	700	300	100
Heart	13	170	100	100
Image	18	1300	1010	20
Ringnorm	20	400	7000	100
F. solar	9	666	400	100
Splice	60	1000	2175	20
Thyroid	5	140	75	100
Titanic	3	150	2051	100
Twonorm	20	400	7000	100
Waveform	21	400	4600	100

表 2 多クラスベンチマークデータセット
Table 2 Multi-class benchmark data sets.

Data	Classes	Inputs	Training	Test
Iris	3	4	75	75
Numeral	10	12	810	820
Blood-cell	12	13	3097	3100
Thyroid (3)	3	21	3772	3428
Hiragana-50	39	50	4610	4610

セットである点である。

4.2 パラメータ設定

LS-SVM, SLS-SVM, FLSLS-SVM はそれぞれ人為的に決定する必要のあるハイパーパラメータとしてカーネルパラメータとマージンパラメータ C がある。ここで, 用いるカーネル関数は線形カーネル: $K(\mathbf{x}, \mathbf{x}') = \mathbf{x}^\top \mathbf{x}'$, 多項式カーネル: $K(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^\top \mathbf{x}' + 1)^d$, RBF カーネル: $K(\mathbf{x}, \mathbf{x}') = \exp\left\{\frac{\gamma}{m} \|\mathbf{x} - \mathbf{x}'\|^2\right\}$ とする。ここで, それぞれ d, γ がカーネルパラメータとなる。

それぞれのハイパーパラメータの設定する値によって, それぞれの識別器の汎化能力が大きく左右される。そのため, 最適なハイパーパラメータを推測する必要がある。本実験ではそれぞれのハイパーパラメータを五分割交差検定法により決定する。カーネルパラメータは $d = \{2, 3, 4, 5\}$ または $\gamma = \{0.1, 0.5, 1, 1.5, 3, 5, 10, 15, 20, 30, 50, 100, 200\}$ から選択する。ただし, 線形カーネルを用いる際, カーネルパラメータは存在しない。マージンパラメータ C は $C = \{0.1, 1, 5, 10, 50, 100, 500, 10^3, 5 \times 10^3, 10^4\}$ から選択する。また SLS-SVM, FLSLS-SVM では一次独立な教師ベクトルの選択においてコレスキー分解を用いており, カーネル行列の分解途中に対角成分が閾値 μ 以下となれば正則性を乱すと判断し, 対応する教師ベクトルを一次従属として削除する。そのため, 閾値 μ を決定する必要があり, $\mu = \{10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}\}$ から選択する。

表 3, 表 4 に 2 クラス, 多クラス問題における五分割交差検定により決定されたハイパーパラメータ値を示す。ここで, “linear” は線形カーネルを用いていることを意味する。2 クラス問題における OAA, ペアワイズ方式は同様の結果が得られるため, 表 3 において 2 方式をまとめて “FLSLS-SVM” と記す。また, “FLSLS (AAO)” は FLSLS-SVM (AAO) を意味する。表 4 における “LS (OAA)”, “LS (pairwise)”, “SLS (OAA)”, “SLS (pairwise)”, “FLSLS (OAA)”, “FLSLS (pairwise)” は OAA, ペアワイズ方式を用いた LS-SVM, SLS-SVM, FLSLS-SVM をそれぞれ意味する。また, Hiragana-50 データセットにおいて, FLSLS-SVM (AAO) は学習時間が膨大となることから計測不能となったため, それぞれの欄には “-” を記す。

4.3 学習時間の比較

表 5, 表 6 に 2 クラス, 多クラス問題における LS-SVM, SLS-SVM, FLSLS-SVM の学習時間 (s) をそれぞれ示す。各方式において最短学習時間を太字でそれぞれ記す。

4.3.1 2 クラス問題

表 5 において, 従来の SLS-SVM と比べて短い FLSLS-SVM の学習時間の前に*を記す。

表 5 より, OAA 方式またはペアワイズ方式を用いた FLSLS-SVM は LS-SVM と比べて 4 個のデータセットにお

表 3 2 クラス問題におけるパラメータ値

Table 3 Parameter values for two-class problems.

Data	Methods	Kernel	C	μ
Banana	LS-SVM	$\gamma = 20$	10^3	-
	SLS-SVM	$\gamma = 20$	5×10^3	10^{-3}
	FSLs-SVM	$\gamma = 20$	5×10^3	10^{-3}
	FSLs (AAO)	$\gamma = 20$	5×10^3	10^{-5}
B. cancer	LS-SVM	$\gamma = 10$	1	-
	SLS-SVM	$d = 2$	100	10^{-3}
	FSLs-SVM	$\gamma = 0.1$	5×10^3	10^{-2}
	FSLs (AAO)	$\gamma = 10$	100	10^{-2}
Diabetes	LS-SVM	$\gamma = 10$	5	-
	SLS-SVM	$d = 3$	5	10^{-5}
	FSLs-SVM	$\gamma = 10$	50	10^{-4}
	FSLs (AAO)	$d = 3$	10	10^{-6}
German	LS-SVM	$\gamma = 3$	50	-
	SLS-SVM	$\gamma = 5$	50	10^{-3}
	FSLs-SVM	linear	100	10^{-3}
	FSLs (AAO)	$\gamma = 1.5$	100	10^{-4}
Heart	LS-SVM	$\gamma = 1.5$	10	-
	SLS-SVM	$\gamma = 15$	0.1	10^{-2}
	FSLs-SVM	$\gamma = 3$	100	10^{-2}
	FSLs (AAO)	$\gamma = 5$	50	10^{-2}
Image	LS-SVM	$d = 5$	10^4	-
	SLS-SVM	$d = 5$	10^4	10^{-5}
	FSLs-SVM	$d = 5$	10^4	10^{-6}
	FSLs (AAO)	$d = 5$	10^4	10^{-6}
Ringnorm	LS-SVM	$\gamma = 20$	5	-
	SLS-SVM	$\gamma = 0.1$	50	10^{-4}
	FSLs-SVM	$\gamma = 0.5$	0.1	10^{-5}
	FSLs (AAO)	$\gamma = 0.1$	50	10^{-4}
F. solar	LS-SVM	$d = 3$	1	-
	SLS-SVM	$d = 3$	10	10^{-2}
	FSLs-SVM	$d = 5$	1	10^{-6}
	FSLs (AAO)	$d = 5$	1	10^{-6}
Splice	LS-SVM	$\gamma = 10$	50	-
	SLS-SVM	$\gamma = 10$	100	10^{-2}
	FSLs-SVM	$\gamma = 1$	5×10^3	10^{-4}
	FSLs (AAO)	$\gamma = 1$	10^4	10^{-3}
Thyroid	LS-SVM	$\gamma = 20$	10^3	-
	SLS-SVM	$\gamma = 20$	5×10^3	10^{-2}
	FSLs-SVM	$\gamma = 20$	10^3	10^{-3}
	FSLs (AAO)	$\gamma = 20$	10^3	10^{-5}
Titanic	LS-SVM	linear	5	-
	SLS-SVM	$\gamma = 10$	10	10^{-2}
	FSLs-SVM	linear	500	10^{-2}
	FSLs (AAO)	$\gamma = 0.1$	10^3	10^{-5}
Twonorm	LS-SVM	$d = 2$	0.1	-
	SLS-SVM	$\gamma = 1.5$	5	10^{-4}
	FSLs-SVM	linear	5	10^{-3}
	FSLs (AAO)	5	10^{-3}	
Waveform	LS-SVM	$\gamma = 15$	1	-
	SLS-SVM	$\gamma = 3$	100	10^{-4}
	FSLs-SVM	$\gamma = 10$	50	10^{-3}
	FSLs (AAO)	$\gamma = 20$	5	10^{-2}

表 4 多クラス問題におけるパラメータ値

Table 4 Parameter values for multi-class problems.

Data	Methods	Kernel	C	μ
Iris	LS (OAA)	$d = 3$	500	-
	LS (pairwise)	linear	5	-
	SLS (OAA)	$d = 2$	500	10^{-2}
	SLS (pairwise)	$d = 4$	100	10^{-2}
	FSLs (OAA)	$\gamma = 3$	5×10^3	10^{-3}
	FSLs (pairwise)	$d = 5$	10	10^{-2}
	FSLs (AAO)	$d = 2$	5	10^{-3}
Numeral	LS (OAA)	$\gamma = 10$	10	-
	LS (pairwise)	$\gamma = 3$	500	-
	SLS (OAA)	$\gamma = 10$	5×10^3	10^{-2}
	SLS (pairwise)	$\gamma = 3$	10^3	10^{-2}
	FSLs (OAA)	$\gamma = 10$	500	10^{-2}
	FSLs (pairwise)	$\gamma = 10$	50	10^{-2}
	FSLs (AAO)	$\gamma = 10$	500	10^{-4}
Blood-cell	LS (OAA)	$\gamma = 5$	10^3	-
	LS (pairwise)	$\gamma = 5$	500	-
	SLS (OAA)	$\gamma = 50$	50	10^{-3}
	SLS (pairwise)	$\gamma = 3$	50	10^{-2}
	FSLs (OAA)	$\gamma = 50$	5×10^3	10^{-4}
	FSLs (pairwise)	$d = 4$	100	10^{-3}
	FSLs (AAO)	$\gamma = 50$	10^4	10^{-5}
Thyroid (3)	LS (OAA)	$\gamma = 20$	5×10^3	-
	LS (pairwise)	$\gamma = 30$	5×10^3	-
	SLS (OAA)	$d = 4$	10^4	10^{-5}
	SLS (pairwise)	$\gamma = 200$	10^4	10^{-6}
	FSLs (OAA)	$\gamma = 100$	10^4	10^{-4}
	FSLs (pairwise)	$d = 3$	10^4	10^{-6}
	FSLs (AAO)	$\gamma = 30$	500	10^{-3}
Hiragana-50	LS (OAA)	$\gamma = 15$	5×10^3	-
	LS (pairwise)	$\gamma = 5$	5×10^3	-
	SLS (OAA)	$d = 4$	0.1	10^{-2}
	SLS (pairwise)	$\gamma = 10$	10^4	10^{-6}
	FSLs (OAA)	$\gamma = 200$	1	10^{-2}
	FSLs (pairwise)	$d = 2$	10^3	10^{-3}
	FSLs (AAO)	-	-	-

いて計算時間が短いですが、9 個のデータセットにおいて計算時間が長くなった。しかしながら、それらの計算時間の増加は従来の SLS-SVM と比べて非常に小さく、Image データセットや Splice データセットのようなデータ数が 1000 を超えるデータセットにおいて特に増加率が小さい。また、従来の SLS-SVM と比べて 10 個のデータセットにおいて計算時間が短くなった。3 個のデータセットでは計算時間が長くなっているが非常に小さい増加幅であり、考慮する必要はないと考えられる。2 クラス問題において従来の SLS-SVM における一次独立な教師ベクトルの選択と標本特徴空間上において最適化問題を解くために必要な計算量は FSLs-SVM と比べて、それぞれ n^2 倍、 $\frac{N^3}{2 \sum_{i=1}^n N_i^3}$ 倍となるが、FSLs-SVM では学習部において式 (12) と式 (17) を解く必要があり、必ずしも計算時間が短くなるとは限ら

表 5 2 クラス問題における学習時間 (s) の比較

Table 5 Comparison of training time in seconds for two-class problems.

Data	LS-SVM	SLS-SVM	FSLs	FSLs (AAO)
Banana	0.219	0.211	*0.108	*0.052
B. cancer	0.008	0.023	0.025	*0.016
Diabetes	0.058	0.294	*0.111	*0.105
German	0.162	1.257	*0.439	*0.421
Heart	0.007	0.013	*0.005	*0.008
Image	0.803	7.597	*1.787	*4.121
Ringnorm	0.344	0.228	* 0.174	*0.175
F. solar	0.091	0.883	*0.248	*0.263
Splice	0.731	6.076	*2.142	*2.150
Thyroid	0.004	0.011	*0.007	*0.008
Titanic	0.004	0.011	0.028	0.033
Twonorm	0.139	0.316	0.317	*0.305
Waveform	0.245	0.277	*0.235	*0.234

表 6 多クラス問題における学習時間 (s) の比較

Table 6 Comparison of training time in seconds for multi-class problems.

Data	Form	LS-SVM	SLS-SVM	FSLs-SVM
Numeral	OAA	0.265	1.65	0.094
	pairwise	0.218	0.500	0.156
	AAO	-	-	0.128
Blood-cell	OAA	8.58	142	7.61
	pairwise	4.51	32.3	3.68
	AAO	-	-	9.59
Thyroid (3)	OAA	11.9	177	139
	pairwise	18.8	472	150
	AAO	-	-	158
Hiragana-50	OAA	38.5	708	8.36
	pairwise	16.0	35.3	10.5
	AAO	-	-	-

ないが、結果から確認できるように学習時間が短縮する傾向にある。これは次節で表す多クラス問題においても同様のことがいえる。FSLs-SVM (AAO) では LS-SVM と比べて 4 個のデータセットにおいて計算時間が短い、9 個のデータセットにおいて計算時間が長くなった。しかしながら、OAA 方式またはペアワイズ方式を用いた FSLs-SVM と同様に増加幅は小さい。また、SLS-SVM と比べて 12 個のデータセットにおいて計算時間が短くなった。

これらの結果より、2 クラス問題において各種 FSLs-SVM の学習コストは従来の SLS-SVM に比べて小さくなり、学習コストの観点から有効的な識別器であるといえる。また、通常の SVM において学習コストが膨大となる傾向にある AAO 方式を用いた FSLs-SVM はその他の方式における FSLs-SVM と比べて同程度の学習時間を示したため、FSLs-SVM の決定関数統合部における AAO 方式の導入は問題ないと考えられる。

4.3.2 多クラス問題

Iris データセットにおいて、いずれの識別器においても学習時間が非常に短く、計測不可能であったため省略している。また、Hiragana-50 データセットにおいて、FSLs-SVM (AAO) は学習時間が非常に長く、計測不可能であったため省略している。

表 6 より、OAA 方式とペアワイズ方式を用いた FSLs-SVM は 3 個のデータセットにおける各方式の最短学習時間を計測しており、Thyroid (3) データセットにおいても従来の SLS-SVM と比べて短い学習時間を計測した。Thyroid (3) データセットは 3 クラス問題であり、そのうち 1 個のカテゴリのデータ数が全データ数の 9 割程度を占めている不均衡データセットであるため、FSLs-SVM における一次独立な教師ベクトルの選択に要する学習時間の削減割合が小さくなり、LS-SVM と比べて学習時間が長くなったと考えられる。FSLs-SVM (AAO) は全データセットにおいて従来の各種 SLS-SVM と比べてほぼ同程度の高速な学習が行われていることが確認された。

これらの結果より、多クラス問題においても各種 FSLs-SVM は従来の SLS-SVM に比べて高速な学習を行え、さらに不均衡データセットを除く多クラス問題において LS-SVM と比べて OAA またはペアワイズ方式における学習時間は短く、学習時間の観点から有効的な識別器であるといえる。

4.4 汎化能力と SV 数の比較

表 7、表 8 に 2 クラス、多クラス問題における各手法の認識率と SV 数を示す。各データセットまたは各方式において最大平均認識率と最少 SV 数を太字でそれぞれ記す。

4.4.1 2 クラス問題

表 7 より、Welch の t 検定 (有意水準: 5%) により SLS-SVM と比べて OAA 方式またはペアワイズ方式を用いた FSLs-SVM は 11 個のデータセットにおいて従来の SLS-SVM と比べて同等以上の平均認識率であることが確認できた。また、LS-SVM と比べて 9 個のデータセットにおいて同等以上の平均認識率であることが確認できた。しかしながら、5 個のデータセットにおいて LS-SVM と比べて平均認識率が劣っていると判定されており、従来の SLS-SVM においてもほぼ同様の結果となるため、標本特徴空間が元空間と比べて分布が変化していることが原因として考えられる。SV 数に関して、OAA 方式またはペアワイズ方式を用いた FSLs-SVM は 5 個のデータセットにおいて最少であり、全データセットにおいて LS-SVM と比べて少ないため、スパース性が付与されていることが確認できた。これらより、OAA 方式またはペアワイズ方式を用いた FSLs-SVM は 2 クラス問題において、従来の SLS-SVM と比べて同等の汎化能力とスパース性を維持できていることを確認できた。

表 7 2 クラス問題における汎化能力と SV 数の比較

Table 7 Comparison of the average recognition rates in percent, standard deviations of the rates, and the average number of support vectors for two-classes problems.

Data	LS-SVM		SLS-SVM		FSLs-SVM		FSLs-SVM (AAO)	
	Rec.	SVs	Rec.	SVs	Rec.	SVs	Rec.	SVs
Banana	89.5 ± 0.5	400	89.2 ± 0.5	44	89.0 ± 0.5	69	89.5 ± 0.4	276
B. cancer	73.6 ± 4.5	200	74.1 ± 4.5	52	73.3 ± 4.5	5	74.1 ± 4.4	105
Diabetes	77.0 ± 1.6	468	77.0 ± 1.7	165	76.9 ± 1.1	284	77.0 ± 1.7	324
German	76.2 ± 2.1	700	75.9 ± 2.1	189	75.8 ± 2.0	40	75.9 ± 2.0	268
Heart	84.2 ± 3.1	170	84.2 ± 3.3	126	84.1 ± 3.2	47	83.7 ± 3.1	75
Image	95.5 ± 0.7	1300	91.7 ± 1.2	279	91.0 ± 1.3	359	95.9 ± 0.5	1210
Ringnorm	96.3 ± 0.5	400	94.2 ± 3.0	22	94.3 ± 2.1	317	98.5 ± 0.1	400
F. solar	66.6 ± 1.6	666	66.6 ± 1.6	32	66.7 ± 1.6	104	66.6 ± 1.6	109
Splice	89.4 ± 0.7	1000	89.3 ± 0.7	977	86.8 ± 0.6	977	86.9 ± 0.6	974
Thyroid	93.8 ± 2.8	140	92.7 ± 2.8	29	92.6 ± 2.8	53	95.7 ± 2.1	108
Titanic	77.3 ± 1.2	150	77.2 ± 0.8	10	77.3 ± 1.1	6	77.2 ± 1.1	13
Twonorm	97.4 ± 0.2	400	97.5 ± 0.2	306	97.6 ± 0.1	40	97.6 ± 0.1	40
Waveform	90.3 ± 0.4	400	89.6 ± 0.6	393	90.2 ± 0.4	399	90.4 ± 0.4	400

表 8 多クラス問題における汎化能力と SV 数の比較

Table 8 Comparison of the recognition rates in percent and the number of support vectors for multi-classes problems.

Data	Form	LS-SVM		SLS-SVM		FSLs-SVM	
		Rec.	SVs	Rec.	SVs	Rec.	SVs
Iris	OAA	93.3	75	92.0	12	94.7	37
	pairwise	97.3	50	97.3	23	97.3	31
	AAO	-	-	-	-	90.7	62
Numeral	OAA	99.3	810	99.5	222	98.9	249
	pairwise	99.8	162	99.8	67	99.2	50
	AAO	-	-	-	-	99.0	491
Blood-cell	OAA	93.3	3097	93.4	1684	93.2	2870
	pairwise	94.3	516	94.5	85	94.2	278
	AAO	-	-	-	-	93.4	3080
Thyroid (3)	OAA	94.4	3772	94.3	1043	93.6	1446
	pairwise	95.4	2515	95.0	1683	94.8	591
	AAO	-	-	-	-	93.7	843
Hiragana-50	OAA	99.2	4610	99.2	3797	97.4	3732
	pairwise	99.2	236	99.3	197	98.1	185
	AAO	-	-	-	-	-	-

FSLs-SVM (AAO) において、Welch の t 検定 (有意水準: 5%) により従来の SLS-SVM と比べて 6 個のデータセットにおいて優れた平均認識率であり、6 個のデータセットにおいて同等の平均認識率であることが確認できた。また、LS-SVM と比べて 12 個のデータセットにおいて同等以上の平均認識率であることが確認できた。特に Image, Ringnorm, Thyroid データセットのように SLS-SVM, FSLs-SVM の平均認識率が通常の LS-SVM に比べて大きな劣化がみられるデータセットにおいても劣化が解消された。また SV 数において、LS-SVM と比べて Ringnorm データセットと Waveform データセットを除いて少なくなっているため、スパース性が付与されていることが確認できる。Ringnorm データセットと Waveform

データセットにおいて、各カテゴリにおいて全教師ベクトルが一次独立と判定されたため、スパース性を付与できなかったと考えられる。カーネルを用いることにより各教師ベクトルが一次独立性が大きくなり、RBF カーネルにおいては特にその傾向があり、Ringnorm データセットと Waveform データセットにおいて一次独立性が非常に高まってしまったと考えられる。このような問題点は五分割交差検定において、汎化能力のみでなくスパース性の観点からハイパーパラメータの選択を行うことにより防げる。これらより、FSLs-SV (AAO) は 2 クラス問題において、LS-SVM に対してスパース性を付与しているのみでなく、従来の SLS-SVM とは異なり、LS-SVM と同等の汎化能力を維持できていることが確認された。

4.4.2 多クラス問題

表 8 より, 各種 FLS-SVM は LS-SVM と比べて SV 数が減少しており, スパース性を付与されていることが確認できた. OAA 方式において, FLS-SVM は 1 個のデータセットにおいて最大認識率を示しており, 4 個のデータセットにおいて LS-SVM, 従来の SLS-SVM と比べて劣化が確認された. 原因として, カテゴリ別に標本特徴空間を生成する際に生じる他カテゴリの情報欠損を決定関数の統合により完全に防ぐことはできなかったことが考えられる.

また, 決定関数結合の際, AAO 方式を用いることにより OAA 方式を用いた FLS-SVM と比べて 3 個のデータセットにおいて認識率の向上が確認できたが, その差は小さい. Iris データセットにおいて, AAO 方式を用いることにより OAA 方式と比べて 4% と大きな劣化が確認されたが, Iris データセットはテストデータ数が 75 個であり, 誤認識の個数が 3 個増加したのみである. これらより, 多クラス問題において AAO 方式と OAA 方式を用いた各 FLS-SVM の汎化能力はほぼ同等であるといえる.

ペアワイズ方式において, FLS-SVM は 1 個のデータセットにおいて最大認識率を示しており, 4 個のデータセットにおいて LS-SVM, 従来の SLS-SVM と比べて劣化が確認された. しかしながら, その劣化割合は OAA 方式における劣化割合と比べて小さく, 他カテゴリの情報欠損は OAA 方式に比べて小さい.

これらの結果より多クラス問題において, OAA 方式と AAO 方式を用いた FLS-SVM は従来の SLS-SVM と比べて汎化能力の観点から劣化しているが, ペアワイズ方式を用いることにより小さくできると考えられる.

5. おわりに

本論文では, カテゴリ別標本特徴空間を用いることによる SLS-SVM の高速学習法 (FLS-SVM) を提案した. 多クラス問題の拡張方式として OAA 方式, ペアワイズ方式, AAO 方式を用いた.

計算機実験により, 提案手法と従来手法の比較・評価を行った. 計算機実験結果により, FLS-SVM は汎化能力とスパース性を維持すると同時に学習時間の削減がなされていることが確認できた. さらに FLS-SVM (AAO) では, 2 クラス問題において従来の SLS-SVM と比べて優れた汎化能力を示した. これらのことから, FLS-SVM はスパース性を付与でき, 学習コストも小さい実用的な識別器であるといえる.

謝辞 本研究は, JSPS KAKENHI 25871033 の助成を受けたものである.

参考文献

[1] Suykens, J.A.K. and Vandewalle, J.: Least squares support vector machine classifiers, *Neural Processing Letters*, Vol.9, No.3, pp.293-300 (1999).

[2] Abe, S.: *Support vector machines for pattern classification (advances in pattern recognition)*, Springer-Verlag, London (2010).

[3] Vapnik, V.N.: *Statistical learning theory*, John Wiley & Sons, New York (1998).

[4] Vapnik, V.N.: *The nature of statistical learning Theory*, Springer, New York (1995).

[5] Bishop, C.M.: *Pattern recognition and machine learning*, Springer, New York (2006).

[6] Suykens, J.A.K., Lukas, L. and Vandewalle, J.: Sparse squares support vector machine classifiers, *European Symposium on Artificial Neural Networks (ESANN2000)*, pp.37-42 (2000).

[7] Suykens, J.A.K., Lukas, L. and Vandewalle, J.: Sparse approximation using least squares support vector machines, *IEEE Symposium on Circuits and Systems (ISCAS2000)*, pp.757-760 (2000).

[8] Hoegaerts, L., Suykens, J.A.K. and Vandewalle, J.: A comparison of pruning algorithms for sparse least squares support vector machines, *11th International Conference on Neural Information Processing (ICONIP2004)*, LNCS3316, pp.1247-1253 (2004).

[9] Cawley, G.C. and Talbot, N.L.C.: Improved sparse least-squares support vector machines, *Neurocomputing*, Vol.48, No.1-4, pp.1025-1031 (2002).

[10] Zeng, X. and Chen, X.: SMO-based pruning methods for sparse least squares support vector machines, *IEEE Trans. Neural Networks*, Vol.16, No.6, pp.1541-1546 (2005).

[11] Abe, S.: Sparse least squares support vector training in the reduced empirical feature space, *Pattern Analysis and Applications*, Vol.10, No.3, pp.203-214 (2007).

[12] Xiong, H., Swamy, M.N.S. and Ahmad, M.O.: Optimizing the kernel in the empirical feature space, *IEEE Trans. Neural Networks*, Vol.16, No.2, pp.460-474 (2005).

[13] Kershaw, D.S.: The incomplete Cholesky - conjugate gradient method for the iterative solution of systems of linear equations, *Journal of Computational Physics*, Vol.26, No.1, pp.43-65 (1978).

[14] Mika, S., Rätsch, G., Weston, J., Schölkopf, B. and Müller, K.R.: Fisher discriminant analysis with kernels, *IEEE Signal Processing Society Workshop*, pp.41-48 (1999).

[15] Kitamura, T. and Sekine, T.: A novel method of sparse least squares support vector machines in class empirical feature space, *19th International Conference on Neural Information Processing (ICONIP2012)*, LNCS7664, Part II, pp.475-482 (2012).

[16] Hsu, C.W. and Lin, C.J.: A comparison of methods for multiclass support vector machines, *IEEE Trans. Neural Networks*, Vol.13, No.2, pp.415-425 (2002).

[17] Tatsumi, K., Hayashida, K., Kawachi, R. and Tanino, T.: Multiobjective multiclass support vector machines maximizing geometric margins, *Pacific Journal of Optimization*, Vol.6, No.1, pp.115-140 (2010).

[18] Rätsch, G., Onoda, T., and Müller, K.R.: Soft Margins for AdaBoost, *Machine Learning*, Vol.42, No.3, pp.287-320 (2001).

[19] available from <http://ida.first.fraunhofer.de/projects/bench/benchmarks.htm>.

[20] available from <http://cs.nyu.edu/roweis/data.html>.

北村 拓也

1985年生。2008年神戸大学工学部電気電子工学科卒業。2009年同大学大学院修士課程修了。2011年同大学院博士課程修了。博士（工学）。同年富山高等専門学校助教。機械学習，パターン認識，関数近似，ファジィ，データマイニングの研究に従事。電子情報通信学会，IEEE 各会員。

関根 崇正

1992年生。2013年富山高等専門学校電気工学科卒業。パターン認識の研究に従事。

塚越 勇規

1993年生。2014年富山高等専門学校電気工学科卒業。パターン認識の研究に従事。