

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

北村 拓也^{†1,a)} 関根 崇正¹ 塚越 勇規¹

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 (FSL-SVM) を提案する。FSL-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。

1. はじめに

最小自乗サポートベクトルマシン (LS-SVM) [1], [2] とサポートベクトルマシン (SVM) [2]-[4] のモデルのひとつである。通常の SVM において解くべき最適化問題が二次計画問題であることと異なり、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において同等程度の精度を示すことから、広く用いられている。しかしながら LS-SVM では、すべての教師ベクトルをサポートベクトル (SV) 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) [2], [6]-[9] の研究が広く行われている。阿部氏が開発した SLS-SVM では、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 [2], [8], [9] 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 とみなせ、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または線形判別分析 (LDA) [2], [10] を用いている。以後、“従来の SLS-SVM” は阿部氏の開発した SLS-SVM を表す。しかしながら、これらの選択における学習コストが大きくな

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。

そこで本論文では、一次独立な教師ベクトルの選択における学習コストを削減した高速 SLS-SVM (FSL-SVM) を提案する。FSL-SVM では、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 としたとき、従来の SLS-SVM コレスキー分解を用いた選択における計算コストは $O(M^3)$ となることに対して、FSL-SVM では $O(n \sum_{i=1}^n M_i^3)$ 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 倍小さくなり、特にカテゴリ数が多いほど有効であることが確認できる。次に、各カテゴリ上の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一対他 (OAA) 方式 [11] とペアワイズ方式を用いる。すなわち、OAA 方式ではあるひとつのカテゴリ標本特徴空間上に他のすべてのカテゴリの教師ベクトルと分類するための識別基準をひとつ設定し、ペアワイズ方式では他のカテゴリのうち、ひと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ひとつずつ設定する。そのため、あるひとつのカテゴリ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なるため、FSL-SVM における各カテゴリ標本特徴空間の次元数は従来の SLS-SVM もおける標本特徴空間の次元数に比べて小さくなるため、各識別基準を決定するために求めるべき変数の数も小さくなり、最適化問題

¹ 富山高等専門学校

13, Hongo-cho, Toyama 939-8630, Japan

^{†1} 現在、富山高等専門学校

Presently with Toyama National College of Technology

^{a)} kitamura@nc-toyama.ac.jp

においても学習コストの削減が期待できる。次に、各カテゴリ標本特徴空間を他の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ひと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にはLS-SVMを用いる。

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

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、教師データ 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。

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

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

$$\text{for } j = 1, \dots, M$$

ここで、 \mathbf{w} , C , ξ_j , b 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、パイアス項を表し、 $\mathbf{g}(x)$ は高次元特徴空間への写像関数を表す。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} は以下のように表せる。

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

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

$$\mathbf{w} = \sum_{j=1}^M z_j \alpha_j K(x_j, x) \quad (4)$$

となる。ここで、 $K(x, x')$ は高次元特徴空間上における x と x' の内積値であり、カーネル関数とよぶ。またLS-SVMにおいて、 $\alpha_j \neq 0$ ($j = 1, \dots, M$) となり、式(4)より決定関数の導出において、すべての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

挙げるように、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とみなす。すなわち、解のスパース性が欠損しているといえる。

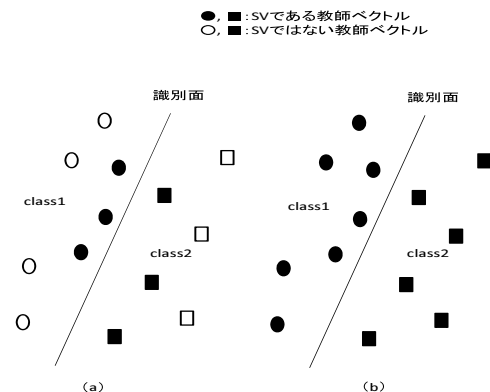


図1 SVMとLS-SVMにおけるSV (a) SVM (b) LS-SVM
Fig. 1 Support vectors for SVM and LS-SVM (a) SVM (b) LS-SVM

2.2 SLS-SVM

LS-SVMのスパース性の欠損における問題点を解決するため、広くSLS-SVMの研究がなされている。阿部氏の開発したSLS-SVMでは、標本特徴空間上で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 x_j ($j = 1, \dots, M$) から一次独立な教師ベクトル x'_j ($j = 1, \dots, N$) をコレスキー分解またはLDAにより選択し、それらの教師ベクトル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を表す。このとき、標本特徴空間への写像関数 $\mathbf{h}(x)$ は以下のように表せる。

$$\mathbf{h}(x) = (K(x, x'_1), \dots, K(x, x'_N))^\top \quad (5)$$

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。標本特徴空間上において以下の最適化問題を解く。

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

$$\text{s.t.} \quad y_j (\mathbf{v}^\top \mathbf{h}(x_j) + b) = 1 - \xi_j \quad (7)$$

$$\text{for } j = 1, \dots, M$$

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 以下となることから、上式の主問題を解くことにより最適解を求める。このとき最適解は以下のように表せる。

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

$$= \mathbf{v}^\top (K(\mathbf{x}, \mathbf{x}'_1), \dots, K(\mathbf{x}, \mathbf{x}'_N))^\top + b$$

式 (8) より, 決定関数の決定において一次独立な教師ベクトルのみが用いられており, それらの教師ベクトルのみを SV とみなせる. 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.

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

2 節における 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 $O(\frac{1}{6}M^3)$ となる. ここで, 選択にはコレスキー分解を用いた場合を示している. すなわち, SV 数の数に関わらず, 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.

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 $O(\frac{1}{6}M_i^3)$ 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 $O(\frac{1}{6}\sum_{i=1}^n M_i^3)$ となる. $M = \sum_{i=1}^n M_i$ となるため, $\frac{1}{6}M^3 \geq \frac{1}{6}\sum_{i=1}^n M_i^3$ 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 $\frac{1}{n}$ 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.

3.2 FLSLS-SVM

FLSLS 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, N_i はクラス i 標本特徴空間の一次独立な教師ベクトル数を表す. 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表せる.

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

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

$$\text{for } i = 1, \dots, n, \quad j = 1, \dots, M$$

ここで, $\mathbf{t}_i, b_i, \xi_{ij}$ はそれぞれクラス i 標本特徴空間上で求める重みベクトル, バイアス項, スラック変数である. 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標本特徴空間はカテゴリごとに独立して生成しているため, 他のカテゴリの情報が欠損しており, 汎化能力の劣化につながりうる. そこで, それぞれのカテゴリ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 方式やペアワイズ方式などの拡張方式を用いられる. 最適化問題 (6), (7) において OAA 方式では, $y_j = i$ のとき $z_j = 1$, $y_j \neq i$ のとき $z_j = -1$ とすればよい. ペアワイズ方式では, クラス i ($i = 1, \dots, n$) と他のあるひとつのクラス l ($l \neq i$) との識別基準を全組み合わせにおいて決定するため, 最適化問題 (6), (7) において $\mathbf{t}_i, M, \xi_{ij}, b_i$ をそれぞれ $\mathbf{t}_{il}, M_i + M_l, \xi_{ilj}, b_{il}$ 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

新たに求めるクラス i とその他のクラスを分類する決定関数 $D'_i(\mathbf{x})$ を式 (12) を用いて

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

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

$$\text{for } i = 1, \dots, n$$

とする. ここで, \mathbf{u}_i は n 次元重みベクトルである. b_i^e はバイアス項である. \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クラス数 n と等しいため, 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

ペアワイズ方式において各カテゴリ標本特徴空間上にて, 他カテゴリのうちひとつと分類する識別基準を決定するため, $n-1$ の識別基準を有する. すなわち, クラス i またはクラス l 標本特徴空間上でのクラス 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. ここで, 決定関数 $D_{il}(\mathbf{x})$ は

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

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, 情報の欠損が発生しうるため, 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

結合することにより防ぐ。新たにクラス i, l を分類する決定関数 $D'_{il}(x)$ を式 (16) を用いて

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

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

とする。 $D'_{il}(x)$ は $D_{li}(x)$ と異なり、 $D_{il}(x) = D_{li}(x)$ となる。 \mathbf{u} と b^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 x において、

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

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、あるひとつのカテゴリとその他のカテゴリのうちのひとつの識別基準を全組み合わせにおいて決定しているため、より細かい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、ひと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 方式を用いて各カテゴリ標本特徴空間上で多カテゴリと分類するための決定関数 $D_i(x)$ ($i = 1, \dots, n$) 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 とその他のクラスを分類する決定関数 $D'_i(x)$ は

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

$$\text{for } i = 1, \dots, n$$

と定義する。ここで、 $\mathbf{f}(x)$ は式 (14) である。 $D'_i(x)$ における 3.2.1 節の OAA 方式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。

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

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

$$\text{for } i = 1, \dots, n, \quad j = 1, \dots, M$$

上式における制約条件は、 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。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

本節では、ベンチマークデータセット [2], [12], [13] を用いた計算機実験により従来の SLS-SVM と FLS-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。計算機実験において用いる計算機の仕様は OS : Windows 7 (64bit), CPU : Intel (R) Core (TM) i7 3.40GHz, メモリ : 16.00GB である。

4.1 パラメータ設定

LS-SVM, SLS-SVM, FLS-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がある。ここで、用いるカーネル関数は線形カーネル: $K(x, x') = x^\top x'$, 多項式カーネル: $K(x, x') = (x^\top x' + 1)^d$, RBF カーネル: $K(x, x') = \exp\left(\frac{\gamma}{m} \|x - x'\|^2\right)$ 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、汎化能力が最大となる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を存在しない。マージンパラメータは $C = \{0.1, 1, 5, 10, 50, 100, 500, 10^3, 5 \times 10^3, 10^4\}$ から選択する。また SLS-SVM, FLS-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、本実験では $\mu = \{10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}\}$ から五分割交差検定法により決定する。

4.2 学習時間の比較

表 1, 2 に 2 クラス、多クラス問題における LS-SVM, SLS-SVM, FLS-SVM の学習時間 (s) を示す。各方式において最短学習時間を太字で記す。

4.2.1 2 クラス問題

2 クラス問題であるため、OAA 方式とペアワイズ方式に

において同様の結果が得られる。また、FSLs-SVM (AAO) は“FSLs (AAO)”と記載し、一対他方式とペアワイズ方式を用いた FSLs-SVM は“FSLs (AAO)”と記載する。ここで、FSLs-SVM において SLS-SVM と比べて短い学習時間の前に*を記す。

表 1 より、OAA 方式またはペアワイズ方式を用いた FSLs-SVM 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ている。しかしながら、それらの計算時間の増加は 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、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ている。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。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ているが、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ている。

これらの結果より、各種 FSLs-SVM の学習コストは従来の SLS-SVM に比べて非常に小さくなっていることが確認された。また、AAO 方式とその他の方式における FSLs-SVM における学習コストは 2 クラス問題においてほぼ同程度であることが確認された。

4.2.2 多クラス問題

LS-SVM, SLS-SVM では多クラス問題への拡張方式として OAA 方式とペアワイズ方式を用いており、FSLs-SVM では 2 方式に加えて 3.2.3 節で述べた AAO 方式を用いている。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 2 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いて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 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 割以上を占めている不均衡データセットであったため、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 SLS-SVM の各種方式と比べて高速な学習が行われていることが確認された。また、OAA 方式を用いた FSLs-SVM と比べて学習時間の増加は確認されたが、その増加幅は小さいといえる。

これらのことから、FSLs-SVM の各種方式は SLS-SVM

に比べて高速な学習を行い、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、より有効であるといえる。

表 1 2 クラス問題における学習時間 (s) の比較

Table 1 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 2 多クラス問題における学習時間 (s) の比較

Table 2 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	-	-	-

4.3 汎化能力と SV 数の比較

表 3, 4 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 数を太字で記す。

4.3.1 2 クラス問題

表 3 より、OAA 方式または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いて最大認識率を示し、SLS-SVM と比べて全データセットにおいて同等の認識率を示している。また SV 数において、5 個のデータセットにおいて最小であり、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る。FSLs-SVM (AAO) は 8 個のデータセットにおいて最大認識率を示しており、さらに 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、最大認識率を示している。また SV 数において、スパース性が付与されていることが確認できる。これらより、2 クラス問題において FSLs-SVM はスパース性と汎化能力を維持できており、FSLs-SVM (AAO) においてはスパース性付与による汎化能力の劣化を解消できたといえる。

4.3.2 多クラス問題

表 3 より、FSLs-SVM の全方式は 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る。OAA, ペアワイズ方式を用いた FSLs-SVM は LS-SVM と SLS-SVM とほぼ同程度の汎化能力を示しているが、FSLs-SVM (AAO) は汎化能力の劣化がみられる。この原因として、バイアス項の省略が考えられる。バイアス項を省略することにより最適化問題の簡易化が可能だが、多クラス問題において汎化能力が劣化すると考えられる。

5. おわりに

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

- least squares support vector machine classifiers. In: *European Symposium on Artificial Neural Networks (ESANN2000)*, pp. 37–42, (2000).
- [7] Suykens, J.A.K., ESAT, Lukas, L., Vandewalle, J.: Sparse approximation using least squares support vector machines. In: *IEEE Symposium on Circuits and Systems (ISCAS2000)*, pp. 757–760, (2000).
- [8] 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).
- [9] 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).
- [10] Mika, S., Ratsch, G., Weston, J., Scholkopf, B., and Muller, K.R.: Fisher discriminant analysis with kernels. In: *IEEE Signal Processing Society Workshop*, pp. 41–48, (1999).
- [11] Kitamura, T. and Sekine, T.: A Novel Method of Sparse Least Squares Support Vector Machines in Class Empirical Feature Space. In: *19th International Conference on Neural Information Processing (ICONIP2012), Part II, LNCS7664*, pp. 475–482, (2012).
- [12] Rätsch, G., Onda, T., and Müller, K.R.: Soft Margins for AdaBoost. *Machine Learning*, vol. 42, no. 3, pp. 287–320 (2001)
- [13] <http://archive.ics.uci.edu/ml>.

表 3 2 クラス問題における汎化能力と SV 数の比較

Table 3 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		FSL-SVM		FSL-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 4 多クラス問題における汎化能力と SV 数の比較

Table 4 Comparison of the average recognition rates in percent and the average number of support vectors for multi-classes problems

Data	Form	LS-SVM		SLS-SVM		FSL-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	-	99.3	197	98.1	185
	AAO	-	-	-	-	-	-