

訓練セットのランダムサンプリングによるマルチカーネル学習法

西田 健次[†] 藤木 淳[†] 栗田多喜夫^{††}

[†] (独) 産業技術総合研究所 〒 305-8568 茨城県つくば市梅園 1-1-1

^{††} 広島大学 〒 739-8521 東広島市鏡山 1-7-1

E-mail: [†]{kenji.nishida,jun-fujiki}@aist.go.jp, ^{††}tkurita@hiroshima-u.ac.jp

あらまし 本稿では、訓練セットのランダムサンプリングによるマルチカーネル学習法 (Multiple Random Subset Kernel Learning: MRSKL) を提案する。MRSKL では、カーネルごとに訓練セットからランダムサンプリングされたサブセットを抽出し、ハイパーパラメータもランダムに設定される。そして、カーネルごとの重みを最適化することで、高い識別率を実現する。最適なカーネル重みを得るために線形 SVM を用いることにより、全体としては階層型 SVM となっている。MRSKL は、非常に少数の訓練サンプルを用いるだけでもかかわらず、最適なハイパーパラメータのもとに全サンプルを訓練に用いた SVM よりも高い識別性能を示した。具体的には、20,000 サンプルの訓練セットに対して、200 から 400 個の訓練サンプルを学習に用いるのみで、全サンプルを用いた単一の SVM よりも高い識別率を達成することができた。また、この例では、単一の SVM でのサポートベクター数は約 4,000 個であったが、MRSKL では 200 から 400 個の訓練サンプルを用いたのみであり、全体としての計算量も小さくなっている。

キーワード マルチ・カーネル学習法、汎化性、サポート・ベクター・マシン (SVM)

1. はじめに

近年、単一カーネルの識別器の性能を向上するために、複数種類のカーネルを用いた学習法、即ち、マルチカーネル学習法が提案されてきている [1], [2]。当初は、カーネルごとに重みを変えることなく統合する単純な手法がとられていたが、Lanckreit ら [3] によって最適なカーネル重みを求める手法が提案され、また Bach [1] によって逐次最適化法 (sequential minimal optimization: SMO) を利用した効果的なアルゴリズムが提案された。

マルチカーネル学習法 (以後、MKL) による識別関数は、カーネルの重み付き総和で以下のように表現される、

$$f(x) = \sum_{m=1}^p \beta_m \langle \mathbf{w}_m, \Phi_m(\mathbf{x}) \rangle + b \quad (1)$$

m はカーネルの番号を示し、 β_m はカーネルの重み係数、 \mathbf{w}_m はサンプルの重み係数である。 $\Phi_m(\mathbf{x})$ は特徴空間の写像を示し、 p はカーネルの総数である。

式 (1) を双対条件により変形することで、次式が得られる。

$$f(x) = \sum_{m=1}^p \beta_m \sum_{i=1}^n \alpha_{m_i} y_i \underbrace{\langle \Phi_m(\mathbf{x}), \Phi_m(\mathbf{x}_i) \rangle}_{\mathbf{K}_m(\mathbf{x}, \mathbf{x}_i)} + b \quad (2)$$

n は訓練サンプルの総数、 α_{m_i} はサンプルへの重み付け係数、 y_i は教師ラベル、そして、カーネルの重み係数は $\beta_m \geq 0$ かつ $\sum_{m=1}^p \beta_m = 1$ の条件を満たす。本稿の実験では、RBF カーネルを用いているが、異なる種類のカーネル (線形カーネル、多項式カーネル、ガウスカーネル

など) を統合することも可能である。

式 (1) で示される MKL アルゴリズムでは、個々のカーネルは全サンプルを用いたカーネル特徴を用いている。しかし、訓練サンプルの異なるサブセットに対してのカーネルを統合することも可能である。

$$f(x) = \sum_{m=1}^p \beta_m \sum_{i \in \dot{\mathbf{X}}} \alpha_{m_i} y_i \langle \mathbf{K}_m(\mathbf{x}, \mathbf{x}_i) \rangle + b \quad (3)$$

ここで、 \mathbf{X} を全訓練サンプルとした時に、 $\dot{\mathbf{X}}$ を m 番目のカーネルに対する訓練サンプルのサブセットとする。なお、サブセットのサンプリング方法は特定のものに限られないが、確率分布 $\eta_m(\mathbf{x})$ に従ってサンプリングを行う場合のカーネルマトリクスは次のように表される。

$$K_\eta(\dot{\mathbf{x}}_i, \dot{\mathbf{x}}_j) = \sum_{m=1}^p \langle \Phi_m(\dot{\mathbf{x}}_i), \Phi_m(\dot{\mathbf{x}}_j) \rangle. \quad (4)$$

ここで、 $\dot{\mathbf{X}} = \eta \mathbf{X}$ であり、 $K_\eta(\dot{\mathbf{x}}_i, \dot{\mathbf{x}}_j)$ となる確率は、 \mathbf{x}_i の確率と \mathbf{x}_j の確率の積となる。従って、サブセット・カーネルは、全サンプルを用いたカーネル・マトリクスとサンプリングに用いられる確率分布 η_m によって、次のように定義される。

$$\begin{aligned} K_\eta(\dot{\mathbf{x}}_i, \dot{\mathbf{x}}_j) &= \sum_{m=1}^p \langle \Phi_m(\dot{\mathbf{x}}_i), \Phi_m(\dot{\mathbf{x}}_j) \rangle \\ &= \sum_{m=1}^p \eta_m(\mathbf{x}_i) \underbrace{\langle \Phi_m(\mathbf{x}_i), \Phi_m(\mathbf{x}_j) \rangle}_{\mathbf{K}_m(\mathbf{x}_i, \mathbf{x}_j)} \eta_m(\mathbf{x}_j), \end{aligned} \quad (5)$$

これは、*Localized Multiple Kernel Learning (LKML)* [5] と

同等のものとなり、複数カーネルの統合時に特徴空間でのゲーティングを行うことなく、同様の効果が得られることを示している。

さて、マルチカーネル学習においては、最終的な識別性能が高くなるように、それぞれのカーネルに対して最適なサブセットとハイパーパラメータを設定するためには、膨大な探索が必要となる欠点がある。そこで、計算時間短縮のため、ランダムサンプリングによってサブセットを設定し、更に、ハイパーパラメータもランダムに設定することとした。そのため、個々のカーネルの識別性能は期待できないが、複数の識別器を統合することにより、高い識別性能を得ようとするものである。最終的な識別器は、それぞれのカーネルを線形結合で統合するものであり、カーネル重み β_m が与えられたサブセット・カーネルに対して最適なものであれば、識別性能に貢献するカーネルに対する重みが大きく、識別への貢献が小さいカーネルに対する重みが小さくなり、結果的に有効なカーネルが選択されることとなる。

そこで、本稿では、訓練セットのランダムサンプリングによるマルチカーネル学習法（Multiple Random Subset Kernel Learning : MRSKL）を提案する。本手法は、訓練サンプルから少数のサンプルをサブセットとしてランダムに抽出してカーネルを定義し、複数のカーネルを線形 SVM によって統合して識別器とする手法である。線形 SVM を用いることにより、与えられたサブセット・カーネルの組み合わせに対して、最適なカーネル重みが得られると考えられる。一般的に、マルチカーネル学習法においては、式 (3) 中のサンプル重み α_i とカーネル重み β_m の同時最適化が研究対象とされることが多いが ([1], [4] など)、本稿ではカーネル重みの最適化に主眼を置き、サンプル重みとは独立に最適化を行うこととした。

本稿の構成は、第二節において訓練サンプルのランダムサンプリングによるマルチカーネル学習法 (MRSKL) のアルゴリズムを紹介し、第三節において実験結果を示す。MRSKL は、人工データによる実験で、MRSKL は、その理論的上限であるベイズ推定に迫る性能を示すことができた。

2. ランダムサンプリングによるマルチカーネル学習アルゴリズム

2.1 訓練サンプルのサブセットを用いた学習法

訓練サンプルのサブセットを用いる学習法は、既に幾つか提案されている。それらの一部は汎化性の向上を目指しており、また一部は訓練時の計算量削減を目指している。*Feature Vector Selection* (FVS) [6] は、訓練サンプル F によって張られる特徴空間を、選択された *Feature Vectors* (FVs) によって近似しようとするものである。また、*Import Vector Machine* (IVM) は、カーネル・ロジスティック回帰 (Kernel Logistic Regression: KLR) において、より少ない数の *Import Vectors* (IVs) によってカーネ

ル特徴空間を近似しようとするものである。FVS, IVM ともに、選択されたサンプルによって特徴空間を近似しようとするものであったが、RANSAC-SVM [9] は選択されたサンプルを用いて、ハイパーパラメータを調整することで、SVM の識別面を近似しようとするものであった。FVS, IVM では、サンプルは逐次的に選択されていくが、RANSAC-SVM ではランダムサンプリングを行う。いずれにしても、これらのアルゴリズムでは全サンプルに対して単一のカーネル関数を用いている。

SABI [8] では、1 ペアのサンプルを逐次的に選択し、ペアのサンプル間を線形に補間することにより、識別面を決定しようとした。SABI はカーネル法を採用したものではないが、複数のペアの識別面を統合する手法は、複数種類のカーネルの統合手法と近いものである。

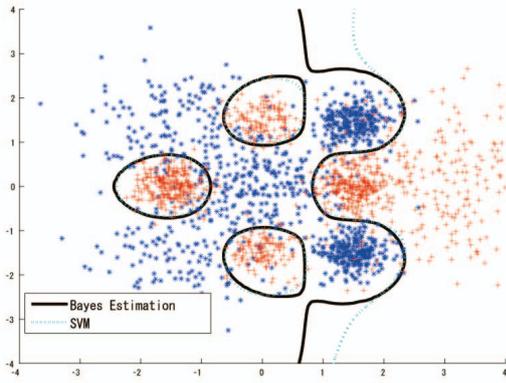
最適なサブセットと最適なハイパーパラメータを求めるためには、膨大な探索が必要となってしまう。そこで、本稿で提案する MRSKL では、ランダムサンプリングによりサブセットを抽出し、ハイパーパラメータもランダムに設定する。これは、個々のカーネルの識別性能を高くすることよりも、カーネルのバリエーションを大きくすることによりアンサンブル学習の効果で識別性能が向上することを期待するものである。さらに、線形 SVM によって複数のカーネルを統合する際に、識別に有効でないカーネルの重みが小さくなり、有効なカーネルのみが残ることも期待している。

2.2 MRSKL でのサブセットのサンプリング手法と学習手順

サブセット・カーネル (\mathbf{K}_m) は、訓練サンプルのサブセット (S_m) によって定義されるため、サンプリング手法によってそれぞれのカーネルの識別性能に影響が生じる。即ち、サブセットを用いるマルチカーネル学習においては、次の三つのパラメータ、サンプル重み α_{m_i} 、カーネル重み β_m 、サンプルのサブセット S_m が識別率に影響することとなる。

我々は、まず、サブセット S_m をランダムサンプリングによって抽出し、ランダムに設定したハイパーパラメータを用いて、サンプル重み α_{m_i} を求める。次に、カーネル重み β_m を求めることとした。サブセット、および、ハイパーパラメータがランダムに決められてしまうため、個々のカーネルには識別性能の低いものも含まれることがある。しかし、カーネル重み β_m が最適であれば、識別能力の低いカーネルに対する重みは小さくなるため、最終的な識別性能には影響しないと考えられる。

サンプル重みとカーネル重みの最適化を独立に行うこととし、 $\alpha_{m_i} y_i \langle \mathbf{K}_m(\mathbf{x}, \mathbf{x}_i) \rangle$ に $f_m(x)$ を代入することで、式 (2) は以下のように書き直すことができる。



黒線はベイズ推定による識別面、青線は単一 SVM での識別面
図 1 実験データとその識別面

$$\begin{aligned}
 f(x) &= \sum_{m=1}^p \beta_m \sum_{i \in S_m} \alpha_{m_i} y_i \langle \mathbf{K}_m(\mathbf{x}, \mathbf{x}_i) \rangle + b \\
 &= \sum_{m=1}^p \beta_m f_m(x) + b
 \end{aligned} \quad (6)$$

MRSKL のアルゴリズムを以下に示す。

- (1) N for the number of training samples T , P for the number of kernels, L for the number of samples in selected subsets S_m ,
- (2) Repeat following steps P times
 - (a) Determine Q training subsets S_m by randomly select samples from T
 - (b) Randomly set hyper-parameters (such as Gaussian width and regularization term for RBF kernel)
 - (c) Train m th classifier f_m over the subset S_m
 - (d) Predict whole training samples T by f_m determining probability output
- (3) Train linear SVM over $f_m: \{m = 1 \dots P\}$ to determine optimal β_m for the final classifier.

パラメータの選択を行う際には、ステップ 2b からステップ 2d までを繰り返し、最も識別性能の高かったものを選択する。 $f_m(x)$ に RBF-SVM を用いることで、MRSKL は階層型 SVM となる。

3. 実験結果

本節では、MRSKL による実験結果を示す。MRSKL において、複数種類のカーネルを統合することも可能だが、本稿の実験ではサブセットカーネルとして RBF-SVM のみを用い、ランダムサンプリングの効果を見ることとした。また、ハイパーパラメータ (LIBSVM における g パラメータ、 c パラメータ) は、適当な範囲からランダムに設定される。

3.1 実験データ

実験に用いた人工データは、10 個の混合ガウス分布

から成り、5 個が +1 クラスのサンプルを生成し、5 個が -1 クラスのサンプルを生成する。20,000 サンプル (+1 : 10,000 サンプル、-1 : 10,000 サンプル) を訓練サンプルとし、それとは別に 20,000 サンプルをテストサンプルとした。図 1 に、本実験で用いたデータセットの分布を示す。図中の識別面はベイズ推定によるものと、単一の SVM によるものである。ベイズ推定による識別率は、訓練サンプルに対して 92.25%、テストサンプルに対しては 92.15% である。単一 SVM による識別率は、訓練サンプルに対して 92.22%、テストサンプルに対して 92.15% である。単一 SVM のハイパーパラメータは、5-fold cross validation によって決定されたもの ($c = 3$, $g = 0.5$) を用いた。また、単一 SVM でのサポートベクター数は 4,257 個であった。

MRSKL の識別性能はサブセットのサイズに影響されると考えられる。そこで、まず、最小のサブセットとして +1 クラスと -1 クラスから一つずつ 1 ペアのサンプルを選択したものをサブセットとして評価を行った。実験は、3 回繰り返し、その平均を識別率として採用した。

3.2 単一ペア・サブセット・カーネル

図 2 は訓練サンプルに対する識別率、図 3 はテストサンプルに対する識別率を示す。正則化項 C は 2^{10} から 2^{-1} の範囲から、ガウス関数の幅 G は 2^2 から 2^{-5} の範囲内で 2 のべき乗部分が一樣となるようにランダムに設定された。3 試行での訓練サンプルに対する平均識別率は約 100 カーネルで単一 SVM の識別率と同等となり、150 カーネルで単一 SVM を越えるようになり、200 カーネルで 92.20% に達した。テストサンプルに対する 3 試行平均識別率は、150 カーネルで単一 SVM を越え、最終的に 200 カーネルで 91.97% に達した。一つのサブセットは 1 ペア (2 サンプル) しか含まれていないことを考えると、200 サンプル (100 ペア) のサンプルによって 4,257 個のサポートベクターによる SVM と同等の性能が得られたことになる。図 4 は MRSKL の識別面を示す。50 カーネルで単一 SVM と比べて、よりベイズ推定に近い識別面となっている。

一般に、RBF カーネルにおいて、ガウス関数の幅を小さくすることで、より高いフィッティング性能を得られることが多い。そこで、 G パラメータを 2^2 から $2^{1.5}$ の範囲に限定することで、その効果を検証した。図 5 は訓練サンプルに対する識別率、図 6 はテストサンプルに対する識別率を示す。訓練サンプルに対する識別率は、ガウス幅の広いものを含む場合よりも収束が遅くなり、150 カーネルでベイズ推定よりも高い値を示し、200 カーネルで 92.26% に達した。これは、訓練サンプルに対して過学習を起こしてしまったと考えられる。そのため、テストサンプルに対する識別率は 91.92% と単一 SVM よりも、やや低い値となってしまった。図 7 は、識別面を示す。この結果は、良い識別面を得るためにはより多くの

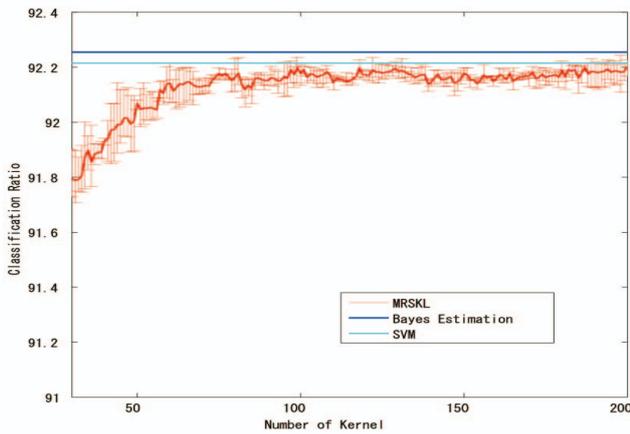


図2 単一ペアカーネルでの訓練サンプルに対する識別率 ($C = 2^{10}$ to 2^{-1} , $G = 2^2$ to 2^{-5})

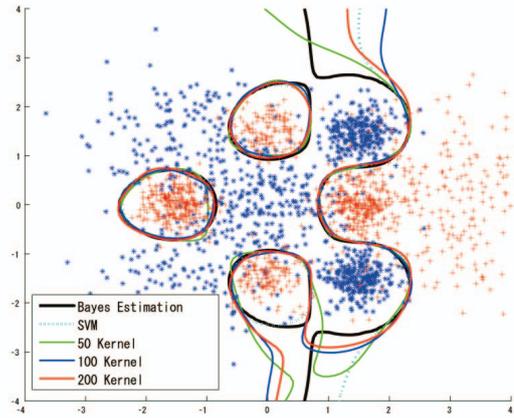


図4 単一ペアカーネルでの識別面 ($C = 2^{10}$ to 2^{-1} , $G = 2^2$ to 2^{-5})

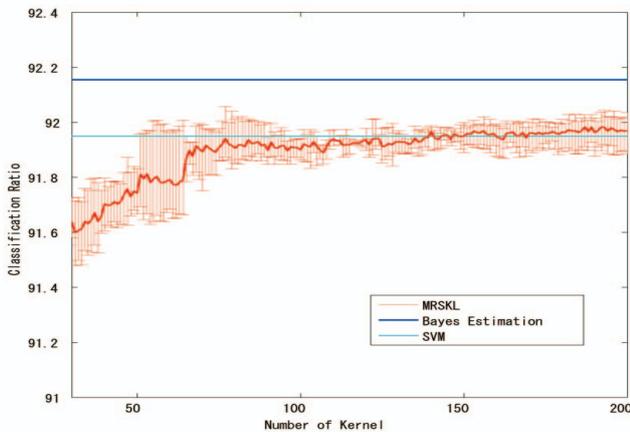


図3 単一ペアカーネルでのテストサンプルに対する識別率 ($C = 2^{10}$ to 2^{-1} , $G = 2^2$ to 2^{-5})

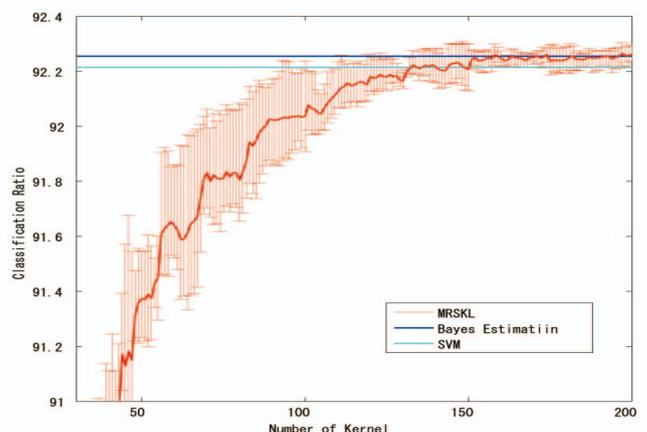


図5 単一ペアカーネルでの訓練サンプルの識別率 ($C = 2^{10}$ to 2^{-1} , $G = 2^2$ to $2^{1.5}$)

(この実験では 200) カーネルが必要であることを示している。

3.3 ランダム・サブセット・カーネル

単一ペアカーネルは、高い識別性能を示したが、識別率を収束させるためには約 150 個のカーネルが必要であった。そこで、複数のサンプルによるサブセット (本実験では 10 サンプルと 100 サンプル) の効果を検証した。まず、単一ペアカーネルと同様の、10 サンプルをランダムに選択し、ハイパーパラメータをランダムに設定するという手順で、実験を行った。

3.3.1 10 サンプル・サブセット・カーネル

図 8 は訓練サンプルに対する識別率を示す。訓練サンプルに対する識別率は早い収束を示し、約 30 カーネルで単一 SVM を上回った。しかし、100 カーネルを過ぎたあたりから却って性能が低下し、最終的には 92.02% となってしまった。図 9 はテストサンプルに対する識別率を示す。テストサンプルに対する識別率の推移は、訓

練サンプルに対する識別率と同様の傾向を示し、収束は早いものの 100 カーネルを越えたあたりから識別率が低下し、最終的に 91.98% となってしまった。図 10 に示される識別面も、識別率の不安定性を反映して、あまり良いものではなくなっている。

3.3.2 パラメータ選択を行う 10 サンプル・サブセット・カーネル

100 カーネル以上での識別性能の低下は、選択されたサブセットに対して適切でないパラメータが与えられたカーネルが存在したためではないかと考えられる。そこで、新たなカーネルを付加する際に適切でないパラメータを避けるため、幾つかのパラメータの組み合わせから適切なものを選択することとした。パラメータ選択の手順は、5 組のパラメータをランダムに生成し、10 サンプルのサブセットに適用する。それまでに付加してきたカーネルと組み合わせ、全体の識別率を最も向上させるものを、新たなカーネルとする。サンプルのサブセットも 5 組生成し、25 通りのサブセット・パラメータの組

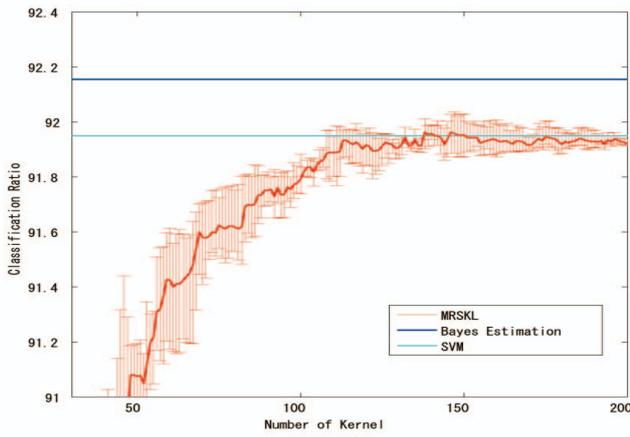


図6 単一ペアカーネルでのテストサンプルの識別率 ($C = 2^{10}$ to 2^{-1} , $G = 2^2$ to $2^{1.5}$)

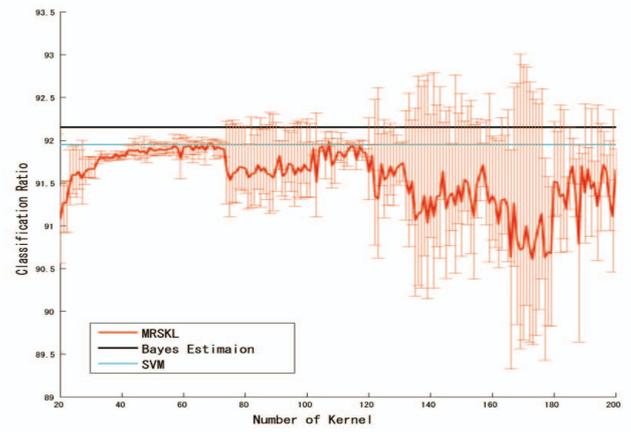


図9 10 サンプルサブセットでのテストサンプルの識別率

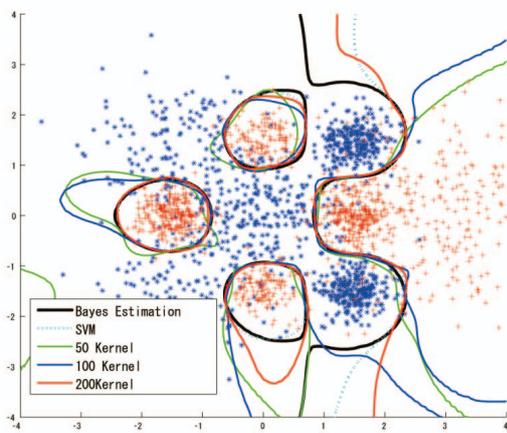


図7 単一ペアカーネルでの識別面 ($C = 2^{10}$ to 2^{-1} , $G = 2^2$ to $2^{1.5}$)

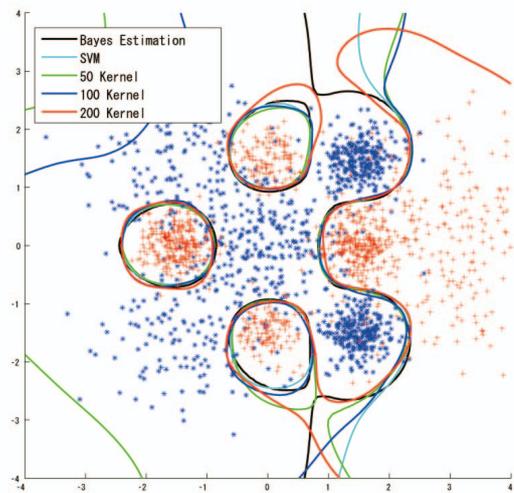


図10 10 サンプルサブセットでの識別面

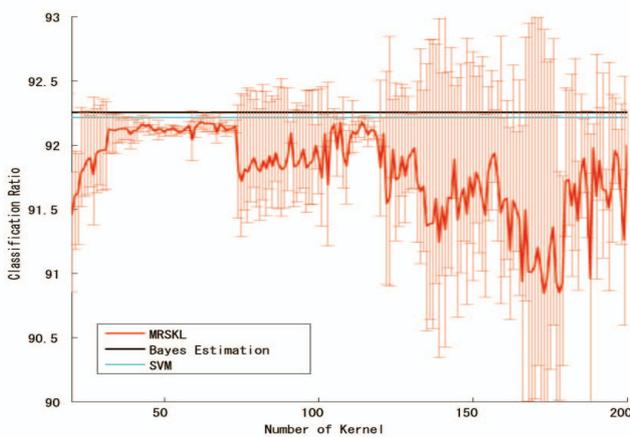


図8 10 サンプルサブセットでの訓練サンプルの識別率

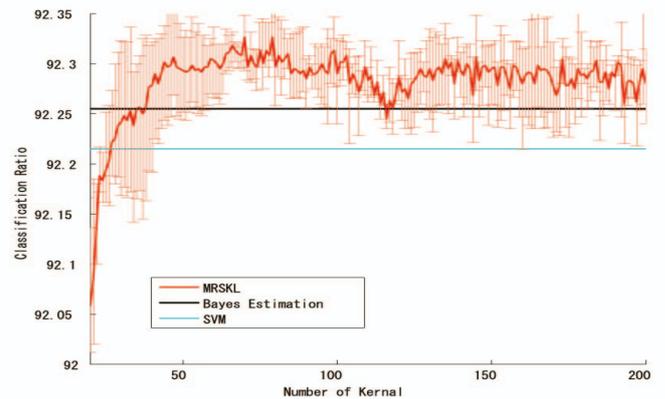


図11 パラメータ選択による10 サンプルサブセットの訓練サンプルの識別率

み合わせの中から選択することとした。

図11は訓練サンプルの識別率を示す。訓練サンプルに対する識別率は早い収束性を見せ、約40カーネルで

ベイズ推定値を越え、200カーネルで92.26%の識別率に達した。図12はテストサンプルの識別率を示す。訓練サンプルで過学習の傾向を見せたものの、テストサンプル

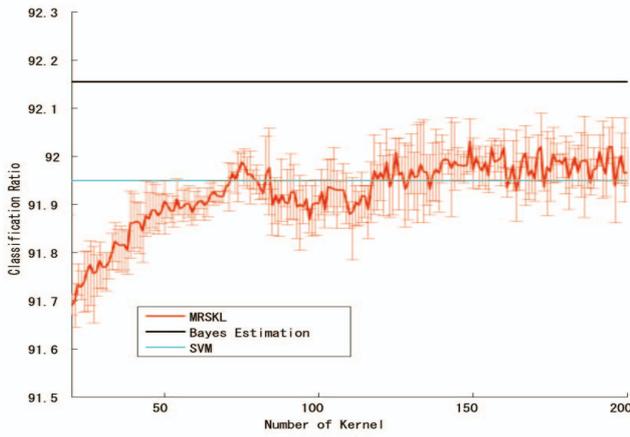


図 12 パラメータ選択による 10 サンプルサブセットのテストサンプルの識別率

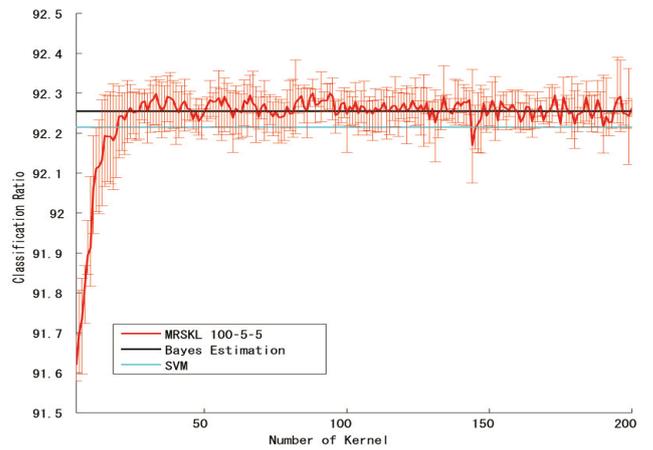


図 14 100 サンプル・サブセット・カーネルによる訓練サンプルの識別率

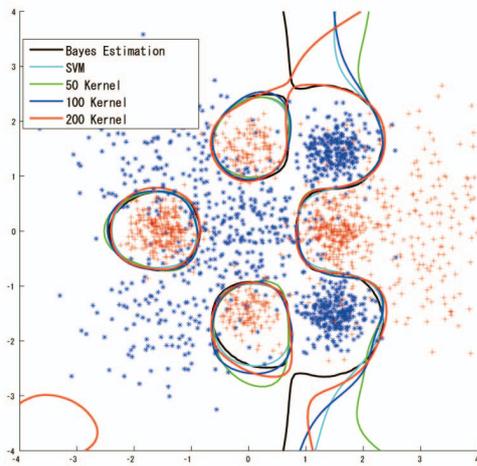


図 13 パラメータ選択による 10 サンプルサブセットでも識別面

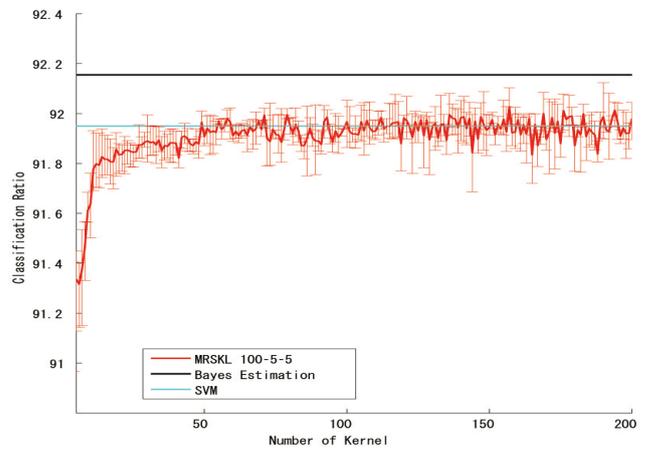


図 15 100 サンプル・サブセット・カーネルによるテストサンプルの識別率

ルに対する識別性能も高く、200 カーネルで 91.98%の性能を示した。

図 13 に示された識別面も、安定した識別性能を示していると考えられる。

3.3.3 100 サンプル・サブセット・カーネル

10 サンプル・サブセットでの結果を受け、100 サンプル・サブセットではパラメータ選択を併用した。図 14 は訓練サンプルでの識別率、図 15 はテストサンプルでの識別率を示す。実験結果は 10 サンプル・サブセットと同様の傾向を示し、若干の過学習 (200 カーネルで訓練サンプルに対して 92.26%の識別率) と単一 SVM と近いテストサンプルでの識別率 (200 カーネルで 91.98%) を示した。図 16 に示されるように、100 サンプル・サブセット・カーネルは、ベイズ推定に近い識別面を構成することができた。

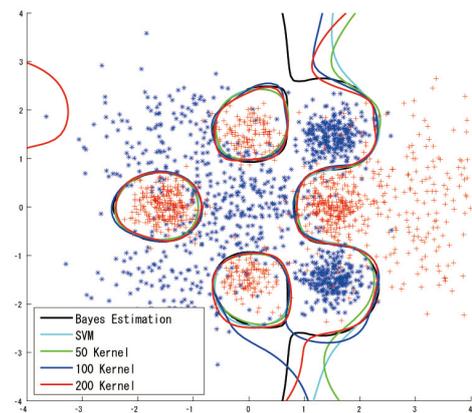


図 16 100 サンプル・サブセット・カーネルによる識別面

3.4 ベンチマークセットによる評価

LIBSVM データセットの中から、cod-rna データセッ

ト [11] による評価を行った。cod-rna データセットは、8次元の特徴ベクトルで、訓練サンプルとして 59,535 個、テストサンプルとして 271,617 個のサンプルを持った 2 クラス識別問題用データセットである。パラメータ選択を行い、500 サンプル、1,000 サンプル、5,000 サンプルのサブセットで評価を行った。表 1 に、cod-rna データセットでの実験結果を示す。1,000 サンプル・サブセット・カーネル（総サンプル数の 1.7%）と 5,000 サンプル・サブセット・カーネルでは、100 カーネルで単一 SVM の性能を越えることができた。

表 1 cod-rna データセットでの識別率

	No. of kernel	training set	test set
Single SVM (Full set)	1	95.12	96.23
MRSKL subset=500	3000	95.03	96.16
MRSKL subset=1000	2000	95.30	96.30
MRSKL subset=5000	100	94.90	96.24

4. ま と め

本稿では、ランダムサンプリングされた少数のサンプルによるカーネルを組み合わせたマルチカーネル学習法、MRSKL を提案した。最小のサブセットである 1 ペア・サンプルによる実験では、200 ペア（サンプル総数 400）で単一 SVM の性能を越えることが示された。

少数サンプル（10 あるいは 100）によるサブセット・カーネルは、識別器の収束を早くするものの、識別性能向上のためにはパラメータの選択を行う必要があった。

cod-rna ベンチマークセットによる評価では、訓練サンプルの 2% から 5% のサンプルで、単一 SVM の性能を越え、ベイズ推定値に近い識別率を示すことができた。

MRSKL では、200 から 1,000 のカーネルを統合する必要がある。しかし、SVM の計算量が $O(N^2)$ から $O(N^3)$ であることを考えると、サブセット・カーネルに対する計算量が削減されるため、全体としての計算量は大きくならないと考えることができる。

サブセット・カーネルを統合する際に線形 SVM を用いたが、カーネル数が増すにつれ、この部分での計算が全体の計算量の大半を占めることとなった。これは、訓練サンプル数と同数の非常に長い特徴ベクトルに対する学習を行わなくてはならないためと考えられる。

本稿では、全てのサブセット・カーネルの出力を上段の線形 SVM の入力とした。しかし、上段の線形 SVM に対しても、特徴選択やサンプル選択を行うことで、計算量の削減と汎化性の向上が可能になると考えられる。

謝 辞

本研究を進めるにあたり、プロトタイプと識別面学習の関係について、有用な助言をくださった和歌山大学、和田俊和教授に感謝いたします。なお、この研究は、科研費課題番号 22500172 による助成のもとに行われた。

- [1] F.R. Bach, G.R.G. Lanckriet, M.I. Jordan, "Multiple Kernel Learning, Conic Duality, and the SMO algorithm", in *Proc. International Conf. on Machine Learning*, pp.41-48, 2004.
- [2] S. Sonnenburg, G. Röttsch, C. Schäfer, B. Schölkopf, "Large Scale Multiple Kernel Learning", in *J. of Machine Learning Research*, Vol.7, pp.1531-1565, 2006.
- [3] G.R.G. Lanckriet, N. Cristianini, P. Bartlett, L.E. Ghaoui, M.I. Jordan, "Learning the kernel matrix with semidefinite programming", in *J. of Machine Learning Research*, Vol.5, pp.27-72, 2004.
- [4] A. Rakotomamonjy, F. Bach, S. Canu, Y. Grandvalet, "More Efficiency in Multiple Kernel Learning", in *Proc. of International Conf. on Machine Learning*, 2007.
- [5] M. Gönen, E. Alpaydin, "Localized Multiple Kernel Learning", in *Proc. of International Conf. on Machine Learning*, 2008.
- [6] G.Baudat, "Feature vector selection and projection using kernels", in *NeuroComputing*, Vol.55, No.1, pp.21-38, 2003.
- [7] J.Zhu, T.Hastie, "Kernel Logistic Regression and the Import Vector Machine", in *J. of Computational and Graphical Statistics*, Vol.14, No.1, pp.185-205, 2005.
- [8] Y.Oosugi, K.Uehara, "Constructing a Minimal Instance-base by Storing Prototype Instances", in *J. of Information Processing*, Vol.39, NO.11, pp.2949-2959, 1998. (in Japanese).
- [9] K.Nishida, T.Kurita, "RANSAC-SVM for Large-Scale Datasets", in *proc. ICPR2008*, Dec. 2008. (CD-ROM).
- [10] C.C.Chang, C.J.Lin, "a library for support vector machines", <http://www.csie.ntu.edu.tw/~cjlin/libsvm>, 2001.
- [11] <http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary.html#cod-rna>