

Confident Margin を用いた SVM のための特徴選択手法

青木 一真[†]黒柳 奨^{†a)}Mauricio Kugler^{†b)}Anto Satriyo Nugroho^{††c)}岩田 彰[†]

Feature Selection using Confident Margin for SVM

Kazuma AOKI[†], Susumu KUROYANAGI^{†a)}, Mauricio KUGLER^{†b)},
Anto SATRIYO NUGROHO^{††c)}, and Akira IWATA[†]

あらまし 本論文では SVM におけるマージンをベースとした特徴選択手法を提案する。しかし一般的なマージン (Normal Margin と呼ぶことにする) の場合、マージンの大きさと SVM の学習により得られる識別関数の良さが適切に対応しない場合があることが明らかとなった。すなわち、Normal Margin を評価値とした特徴選択を行って得られた特徴セットが、必ずしも最良の識別関数を与えるとは限らない。そこでこの問題を解決するために Confident Margin(CM) という新しい評価基準を導入し、それを用いた特徴選択アルゴリズム SBS-CM を提案する。SBS-CM による人工データや実データを用いた実験を行った結果、従来手法よりも良い結果を得ることができた。また評価値 CM の値の変化から最適な特徴セットをおおよそ求めることが可能であった。

キーワード SVM, マージン, 特徴選択, SBS, Confident Margin

1. まえがき

近年、パターン認識の分野で V.Vapnik [1] によって提案された Support Vector Machine(SVM) が注目されている。SVM はニューラルネットワークや nearest neighbor など現在知られている多くの手法の中でも、最も認識性能の優れた学習モデルの一つである。SVM は 1995 年に発表された手法であるにもかかわらず、画像認識や手書き文字認識、バイオインフォマティクス (遺伝子発現解析, タンパク質構造解析) などのような多くの分野に適用されている [2] [3]。

パターン認識で扱うデータは、高次元ベクトルである場合が多い。高次元ベクトルデータをそのまま識別器械で識別した場合、計算コストが高くなる、あるい

は、識別に関してあまり意味を持たない不要な特徴 (ノイズ) により、最良の識別を行うことができないというような問題が生じる。

また overfitting に関して、SVM を用いた場合には特徴次元数を削減しなくても overfitting を避けることができると言われている [4]。しかしこの場合でも、特徴選択を行うことにより識別精度を向上させることができるという報告がされており、SVM においても特徴選択は有効な手段といえる [5]。

そこで本論文では、SVM による SVM のための特徴選択を提案する。SVM の最大の特徴は、マージンを最大化することにより最適な分離超平面を引くことにある。こうすることで SVM はより汎化能力が高い識別器械となる。そこで、SVM の分離超平面を決定するマージンを評価値とした特徴選択を提案することで、SVM で識別をするために重要な特徴を取り出すことができると考えられる。

しかし一般的な Geometric Margin を計算した場合、識別関数の良さとそのマージンの大きさが必ずしも適切に対応しないことがある。そのため本論文では、マージンの信頼性を高めた新たな手法”Confident Margin(CM)”を提案し、それを評価値とする特徴選択を提案する。

[†] 名古屋工業大学大学院工学研究科情報工学専攻

〒 466-8555 名古屋市昭和区御器所町

Department of Computer Engineering, Nagoya Institute of Technology, Gokiso-cho, Showa-ku, NAGOYA, 466-8555, JAPAN

^{††} 中京大学生命システム工学部身体システム工学科

〒 470-0393 豊田市貝津町床立 101

School of Life System Science & Technology, Chukyo University, 101 Tokodachi, Kaizu-cho, Toyota, 470-0393, JAPAN

a) E-mail: bw@nitech.ac.jp

b) E-mail: mauricio@kugler.com

c) E-mail: nugroho@life.chukyo-u.ac.jp

2. SVM(Support Vector Machine)

SVMは線形SVMと非線形SVMに大別できるが、以降は非線形SVMに関して説明する。

SVMは入力サンプルを高次元空間に写像し、その空間において最適な分離超平面を求める2クラス分類のための識別器械である[1]。SVMにおいては求められる分離超平面と各サンプルとの距離をマージンと定義し、これを最大とするような識別関数を求めることを目的とする。

各々の学習データは、特徴ベクトル $\mathbf{x}_i \in \mathbf{R}^n$ ($i = 1, \dots, l$)、それに割り当てられたクラス $y_i \in \{-1, 1\}$ の組からなる。ここで写像

$$\Phi: \mathbf{R}^n \mapsto \mathbf{R}^q$$

を用いて学習データをより高次元の空間 \mathbf{R}^q に写して、その空間で線形識別を行うことを考える。

一般に、このような非線形写像によって変換した特徴空間の次元は非常に大きくなり、膨大な計算量が必要となる。しかしSVMでは内積が計算できれば最適な識別関数を求めることができる。つまり、非線形に写像した空間で二つの要素 $\Phi(\mathbf{x}_1)$ 、 $\Phi(\mathbf{x}_2)$ の内積が

$$\Phi(\mathbf{x}_1) \cdot \Phi(\mathbf{x}_2) = K(\mathbf{x}_1, \mathbf{x}_2) \quad (1)$$

のように、入力特徴 \mathbf{x}_1 、 \mathbf{x}_2 のみから計算できるなら、非線形写像によって変換された特徴空間での特徴 $\Phi(\mathbf{x}_1)$ 、 $\Phi(\mathbf{x}_2)$ を実際に計算する代わりに、カーネル $K(\mathbf{x}_1, \mathbf{x}_2)$ から最適な非線形写像を求められる。本論文では次に示す Gauss カーネルを用いた。

$$K(\mathbf{x}_1, \mathbf{x}_2) = \exp\left(-\frac{\|\mathbf{x}_1 - \mathbf{x}_2\|^2}{2\sigma^2}\right) \quad (2)$$

ここで識別関数は次式のように表せる。

$$f(\mathbf{x}) = \mathbf{w} \cdot \Phi(\mathbf{x}) + b \quad (3)$$

ここでマージンは \mathbf{w} を用いて $\frac{1}{\|\mathbf{w}\|}$ と表すことができる[6]ため、 $\|\mathbf{w}\|$ を最小化することによりマージンを最大化することができる。

この問題を解くためにラグランジュの乗数法（ラグランジュ乗数を α とする）を用いると、ウェイトベクトル \mathbf{w} は、

$$\mathbf{w} = \sum_{i=1}^l \alpha_i y_i \Phi(\mathbf{x}_i) \quad (4)$$

となる。したがって識別関数は次のようになる。

$$f(\mathbf{x}) = \sum_{i=1}^l \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b \quad (5)$$

SVMは次の制約付き最適化問題に定式化できる。

目的関数：

$$\sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \rightarrow \text{最大化} \quad (6)$$

制約条件：

$$0 \leq \alpha_i \leq C \quad (i = 1, \dots, l), \quad \sum_{i=1}^l \alpha_i y_i = 0 \quad (7)$$

ここで C は多少の識別誤りは許すように制約を緩める（ソフトマージン）ためのパラメータである。

3. 特徴選択

3.1 特徴選択の評価基準

n 個の要素を持つオリジナルの特徴集合を Y 、所望の選択された部分集合 $X (X \subseteq Y)$ の要素数を d とする。また、集合 X に対する評価基準を与える関数を $J(X)$ と表す。評価値 J の値が高いほど良い特徴の集合であるといえる。

特徴選択問題は、集合 $X \subseteq Y$ を求めることで、 $|X| = d$ と、

$$J(X) = \max_{Z \subseteq Y, |Z|=d} J(Z) \quad (8)$$

を満たすものである。

特徴選択は評価基準の種類によって、ラッパーモデル(wrapper model) とフィルターモデル(filter model) に分けられる[7]。

ラッパーモデルは、評価基準関数に認識率 $1 - p_e$ (p_e : 誤り確率) を用いる特徴選択である。したがって、直接的にパフォーマンスが最大となる特徴セットを求めることができる。

一方フィルターモデルは、直接パフォーマンスを最適化する特徴セットを求めるのではなく、前処理としての特徴選択である。この場合、認識率以外の何らかの評価関数を用いて特徴セットの良さを表す必要がある。

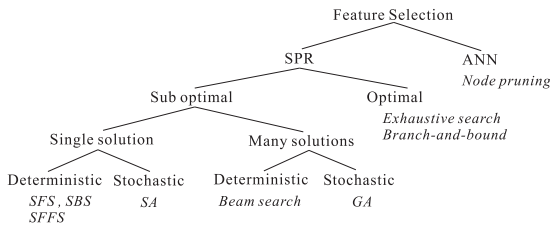


図 1 特徴選択アルゴリズムの分類

Fig. 1 Category of feature selection algorithms

3.2 特徴選択の種類

Jain [8] により特徴選択アルゴリズムは図 1 のように分類された。

その中の”Deterministic, Single-Solution Methods”カテゴリに分けられるアルゴリズムは、与えられた問題の解がいつも同じになる決定論的 (deterministic) なアルゴリズムであり、単一の解 (特徴集合) を持ち、ターゲットとする基準を満たすまで繰り返し特徴を増やしたり減らしたりするものである。

このカテゴリに属する手法は 3 つのグループに分けることができる。一つは、特徴数が 0 の状態からスタートし特徴を増やしていく方法であり、forward 型と呼ばれるものである。二つ目は、全ての特徴を持つ状態からスタートし特徴を削除していく方法で、backward 型と呼ばれるものである。forward 型アルゴリズムの代表例として、Whitney [9] が提案した SFS (Sequential Forward Selection) があり、backward 型アルゴリズムの代表例として、Marill [10] が提案した SBS (Sequential Backward Selection) がある。これらの方法は、可能な全ての特徴の組合せを試せるわけではないので、最適な解を与えるという保証はないということに注意しなければならない。

これらは 1 方向だけの探索アルゴリズムであり、最良の特徴の組合せを求めることは困難である。そこで forward 型と backward 型を組み合わせた floating 型アルゴリズムが研究された。このアルゴリズムの代表例として、Pudil [11] により提案された SFFS (Sequential Floating Forward Search) がある。

本論文では、SVM の最大の特徴であるマージンを評価関数とした特徴選択を提案する。SVM はマージンを最大化することで最適な分離超平面を求める。マージンを評価値とする特徴選択を行うことで、SVM で識別をするために重要な特徴を取り出すことができると考えられる。マージンについて次章で詳しく説明

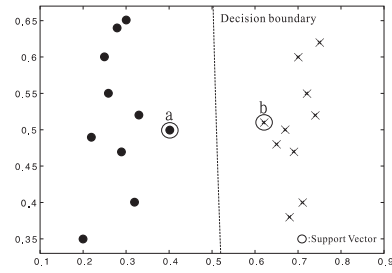


図 2 ノイズのないデータでの識別境界とサポートベクター

Fig. 2 Decision function and support vector in the non-noise data

する。

4. マージン

4.1 Normal Margin (NM)

2. で述べた SVM のマージン (Geometric Margin) を、本論文では Normal Margin (NM) と呼ぶことにする。Normal Margin はウェイトベクトル \mathbf{w} を用いて以下のように求めることができる。

$$NM = \frac{1}{\|\mathbf{w}\|} \quad (9)$$

また $\|\mathbf{w}\|$ は次のようになる。

$$\|\mathbf{w}\|^2 = \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \quad (10)$$

したがって Normal Margin は次の式によって計算することができる。

$$NM = \left(\sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \right)^{-\frac{1}{2}} \quad (11)$$

ここで例として、Normal Margin を用いて図 2,3 それぞれのマージンを求めることとする。

図 2 のサンプルは各クラスに 10 個のサンプルが与えられている完全な線形分離問題である。図 3 は、図 2 のサンプルにそれぞれのクラスに一つずつのサンプルを追加したものである。ただし実際のクラスとは反対側のグループにそれらサンプルが入り込んでいる。この場合、図 3 のような識別境界を引くことができるが、追加した 2 つのサンプルは異なるクラスのグループに入り込んでいるため完全に分類することはできない。

このとき、図 2 の Normal Margin は 0.233 で、図 3 の Normal Margin は 0.308 であった。

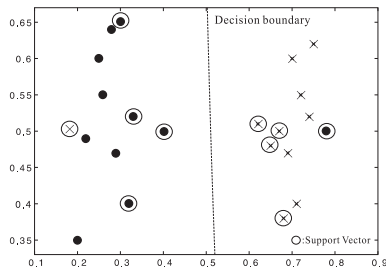


図3 ノイズのあるデータでの識別境界とサポートベクター
Fig.3 Decision function and support vector in the noise data

このように Normal Margin を用いると、図3のような完全に分離ができていない状況の時のマージンの方が、図2のように完璧な分離超平面が引ける時のマージンより大きくなってしまふことがある。

Normal Margin はサポートベクターとなるサンプルから求めることができる。図2において2個のサポートベクターを \mathbf{a}, \mathbf{b} と名付け、Normal Margin を求める過程を説明する。Normal Margin は式(11)より $(\alpha_a \alpha_a y_a y_a K(\mathbf{x}_a, \mathbf{x}_a) + \alpha_b \alpha_b y_b y_b K(\mathbf{x}_b, \mathbf{x}_b) + 2\alpha_a \alpha_b y_a y_b K(\mathbf{x}_a, \mathbf{x}_b))^{-\frac{1}{2}}$ である。ここで $y_a y_a = 1, y_b y_b = 1, y_a y_b = -1$ であり、本論文ではカーネルに式(2)で示す Gauss カーネルを用いたため $K(\mathbf{x}_a, \mathbf{x}_a) = 1, K(\mathbf{x}_b, \mathbf{x}_b) = 1$ である。したがって $(\alpha_a^2 + \alpha_b^2 - 2\alpha_a \alpha_b K(\mathbf{x}_a, \mathbf{x}_b))^{-\frac{1}{2}}$ となる。

図3における Normal Margin も同様に求めることができるが、反対側に入り込んだノイズサンプルがサポートベクターになるためそれを考慮しなければならない。このようなサポートベクターを bad SV(b-SV)、正しく分類されたサンプルでサポートベクターになっているものを good SV(g-SV) と呼ぶことにする。式(11)で Normal Margin を求める際、異なるクラスラベルを持つサンプルにおける b-SV と g-SV の項の値は負となる。またそれらはグルーピングにおいて同一のカテゴリに入っているため、サンプル間の距離が図2における \mathbf{a}, \mathbf{b} の距離よりも小さくなりカーネル K の値が大きくなる。したがって負の項の値が大きくなり、全体として Normal Margin の値が大きくなる。

Normal Margin の値は、ノイズがサンプル全体のどの位置に入るかにより大きく変わってしまう。

本論文で用いた例の場合 Normal Margin を用いると、図3のような完全に分離ができていない状況の時の場合の方が、図2のように完璧な分離超平面が引け

る時のマージンより大きくなった。すなわち、識別関数の良さとマージンの大きさが適切に対応しない場合があることが明らかとなった。

特徴選択の評価関数としてマージンを用いる場合、このような問題を解決しなければならない。なぜなら識別に関して意味を持たない特徴の場合、反対側に入り込んでしまったサンプルが多いため、それらによって評価値であるマージンの大きさが適切に表現されなくなる場合があるからである。

そこで識別関数の良さを考慮したマージン、Confident Margin(CM) を提案する。

4.2 Confident Margin(CM)

識別関数の良さを表すために、正しく分類されなかったサンプルに対してペナルティを与えることが Confident Margin の特徴である。このペナルティを Confidence と呼ぶことにし、Confidence(c) を次式によって定義する。

$$c = \frac{1}{l} \sum_{i=1}^l y_i f(\mathbf{x}_i) \quad (12)$$

ここで $f(\mathbf{x}_i)$ は式(5)で示した識別関数で、 y_i はサンプル i の属するクラス、 l はサンプル数である。この式(12)より、正しく分類されないサンプルが多いほど識別関数の信頼性(confidence)が小さくなり、識別関数の良さを表すことができる。

この Confidence を用いて Confident Margin を次式のように定義する。

$$CM = c \cdot NM \quad (13)$$

Normal Margin と Confidence の積をとることで、識別誤りが多い場合のマージンの大きさは、識別誤りが少ない、あるいは無い場合に比べて小さくすることができる。すなわち Confident Margin は識別関数の良さとマージンの大きさを適切に対応させることができる。また識別誤りがないような場合においても、識別関数の出力値とサンプルのクラスの積により confidence を求めているので、マージンの相対的な大きさへの影響はないと言える。

なお、SVM における識別境界の最適化においては、正しく分類されかつ識別境界から十分に離れたサンプルについては評価の対象とはされない。このため、上記 Confidence c についても c が十分に大きい正の値になった場合にはこれを飽和させることで、評価値を制限する方法が考えられる。

しかし、予備実験の結果、このような制限を与えた場合には、CM 値を用いた特徴選択において良好な特徴の選択が行えないことがわかった。これは評価値を制限することにより、選択された特徴セットごとの評価値の差が小さくなり、特徴選択の精度が落ちるためだと考えられる。そこで本論文ではこのような評価値の制限は行わず、すべてのサンプルを用いて c を計算することとした。

実際に図 2,3 の Confident Margin を求めた結果を表 1 に示す。

表 1 Confident Margin と Normal Margin の比較
Table 1 Comparison between Confident Margin and Normal Margin

	Fig.2	Fig.3
Confiden Margin	0.364	0.266
Normal Margin	0.233	0.308

表 1 より Confident Margin を用いた場合、識別関数の良さとマージンの関係が正しく表現されていることが分かる。

4.3 Confident Margin を用いた SBS

本論文では、backward 型アルゴリズムの一つである SBS をベースとし、評価関数として Confident Margin を用いた特徴選択を提案する。Confident Margin を評価値とすることで、マージンの値を最大にする特徴セット、すなわち最良の分離超平面を与える特徴セットを求めることができる。提案する特徴選択アルゴリズムを以下 SBS-CM と呼ぶことにし、SBS-CM のアルゴリズムを図 4 に示す。

```

1.Initialize:
  Subset of surviving features  $\mathbf{s} = [1, 2, \dots, n]$ 
2.repeat
  for all surviving features
  (a)do Train the SVM classifier with all the training
    examples without  $i$ th feature.
  (b)do compute  $CM^{(i)}$ 
    /* $CM^{(i)}$  is the Confident Margin
      without  $i$ th feature.*/
  (c)Find the worst feature
     $worst = \mathit{argmax}(CM^{(i)})$ 
  (d)Remove the feature that maximizes  $CM$ 
     $[1, \dots, worst-1, worst+1, \dots, n]$ 
3.until  $\mathbf{s}$  is empty

```

図 4 SBS-CM アルゴリズム
Fig.4 SBS-CM algorithm

SBS-CM は、 i 番目の特徴を削除して SVM で学

習した後、 i 番目の特徴を削除したときの Confident Margin を求める。そして Confident Margin の値が最大になったときの特徴を特徴ベクトルから削除する。これを繰り返すことにより、マージンを最大とする特徴セットを求めることができる。

5. 評価実験

本論文で提案する SBS-CM の性能を評価するため、まず人工データを用いた実験を行うことで提案手法が正しく動作することを確認し、次に実データを用いた実験により従来手法との比較を行う。比較対照として用いた特徴選択アルゴリズムは、次に示す SVM-RFE と、SBS アルゴリズムで SVM による識別率を評価値としたラッパーモデルの特徴選択である。

5.1 SVM-RFE

SVM を用いた特徴選択の一つとして、Guyon [5] によって提案された backward 型アルゴリズムがベースの SVM-RFE アルゴリズムがある。SVM-RFE では SVM におけるウェイトベクトルのノルム $\|\mathbf{w}\|^2$ を用いた、次に示す評価関数を用いる。

$$\begin{aligned} & \left| \|\mathbf{w}\|^2 - \|\mathbf{w}^{(i)}\|^2 \right| \\ &= \frac{1}{2} \left| \sum_{j,k} \alpha_j \alpha_k y_j y_k K(\mathbf{x}_j, \mathbf{x}_k) \right. \\ & \quad \left. - \sum_{j,k} \alpha_j^{(i)} \alpha_k^{(i)} y_j y_k K^{(i)}(\mathbf{x}_j, \mathbf{x}_k) \right| \quad (14) \end{aligned}$$

ここで $K^{(i)}$, $\alpha_j^{(i)}$ は i 番目の特徴を削除したときのカーネルマトリックスと、式 (6) で計算される α である。本論文で式 (14) で表された評価基準を SR と呼ぶことにする。

SVM-RFE はウェイトベクトルのノルムを評価基準として用いているため、式 (11) で示す Normal Margin を評価基準にすると考えることができる。

5.2 人工データを用いた実験

ここでは、人工データを用いた実験を行うことで SBS-CM の有効性の検証を行う。人工データとして、次に示す範囲の一樣乱数で構成されたデータを用いた。class1 に属するサンプルは、次元 1 が 0~0.2、次元 2 も 0~0.2 の範囲の乱数 50 サンプルと、次元 1 が 0.8~1.0、次元 2 も 0.8~1.0 の範囲の乱数 50 サンプルの計 100 サンプルである。class2 に属するサンプルは、次元 1 が 0~0.2、次元 2 が 0.8~1.0 の範囲の乱数 50 サンプルと、次元 1 が 0.8~1.0、次元 2 が 0~

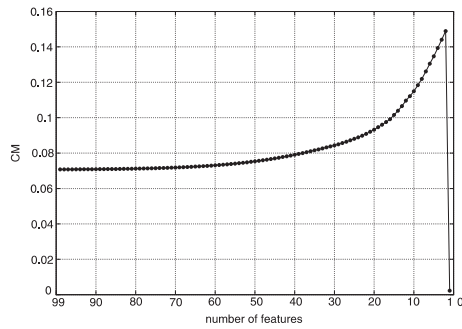


図 5 CM の変化 (XOR データ)
Fig. 5 Variation of CM(XOR data)

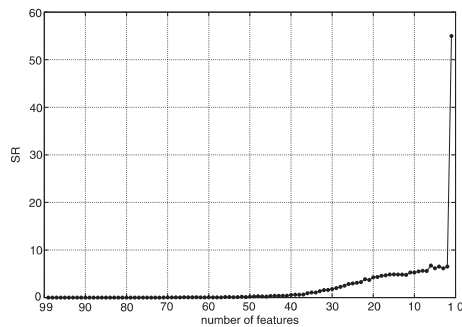


図 6 SR の変化 (XOR データ)
Fig. 6 Variation of SR(XOR data)

0.2 の範囲の乱数 50 サンプルの計 100 サンプルである。これら 2 次元で構成されるデータは XOR 問題を表している。

このデータにノイズとして、0~1 の値を持つ一様乱数で構成された特徴を 98 次元追加した。すなわちデータ数 200(各クラス 100 サンプル)、次元数 100 のデータである。なお XOR を構成する有効な次元は 1 次元目と 2 次元目である。

このデータから XOR を構成する有効な 2 次元を取り出すことを目標とする特徴選択を SBS-CM を用いて行った。SVM の学習におけるパラメータは、 $\sigma = 1.0$ 、 $C = 100$ とした。実験結果を図 5 に示す。図 5 は評価値 CM の値の変化を示しており、横軸は次元数、縦軸は CM の値である。

実験結果はノイズから順に削除されていき XOR を構成する有効な 1,2 次元目が最後まで残っていた。また有効な 2 次元の特徴のみの時に評価値が最大となった。

一方 SVM-RFE を用いて同様の実験を行った結果を図 6 に示す。

SVM-RFE を用いた場合も SBS-CM の場合と同様

に、ノイズから順に削除されていき XOR を構成する有効な 1,2 次元目が最後まで残った。しかし有効な 2 次元の時に評価値が最大にはならなかった。このように SVM-RFE では特徴セットの良さと評価値の大きさが対応しないことがある。

ここで用いたデータのように重要な特徴が既知の場合は、削除された特徴を見ることで、特徴選択の結果の善し悪しを判別することができるが、一般的にどの特徴が重要かが分からない問題の場合の方が多い。その場合、評価値の大きさでその特徴セットの良さをはかることができれば効率的である。しかし SVM-RFE ではこのようなことができないことがデメリットであると考えられる。

人工データを用いた実験により、このような単純なデータで構成されている問題について SBS-CM は正しく動作することを確認できた。次に実データを用いた実験を行うことで、複雑な問題に対する SBS-CM の有効性について検証する。

5.3 実データを用いた実験

実データを用いた実験として、UCI データベース [12] の sonar データと、ionosphere データを用いた。sonar データはサンプル数 208(class1:97,class2:111) 次元数 60 で、ionosphere データはサンプル数 351(class1:225,class2:126) 次元数 33 である。ただしオリジナルの ionosphere データは 34 次元であるが 2 次元目の全ての値が 0 であったため、その次元を削除し全体を 33 次元とした。これらのデータについて SBS-CM により特徴選択を行うことで、どの程度パフォーマンスが向上するかを調べることで SBS-CM の有効性を検証する。

実験方法は SBS-CM で特徴選択を行い、特徴が 1 個削除されるごとに SVM による識別を行う。いずれのデータについてもサンプル数が充分でないため学習データとテストデータを分けることは困難である。予備実験として全データを学習データとして用いたところ、特徴選択を行わない状態で認識率がほぼ 100% になってしまい、特徴選択による認識率の向上を確認することができなかった。これは学習データに対して識別機械が過学習の状態になっているためであり、実験手法としては不相当である。このようにサンプル数が充分でないために学習データに対して過学習が起る状況はパターン識別課題において一般的にみられる。そこで本論文では 1 サンプルをテストデータ、残りのサンプルを学習データとしてテストデータの識別を行う

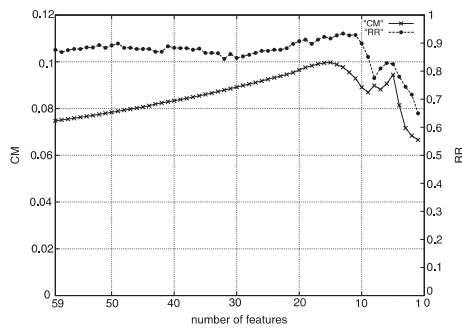


図 7 CM と RR の変化 (sonar データ)
Fig. 7 Variation of CM and RR(sonar data)

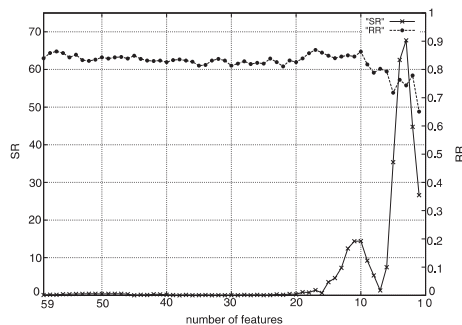


図 8 SR と RR の変化 (sonar データ)
Fig. 8 Variation of SR and RR(sonar data)

Leave One Out 法を用いた。これを全てのサンプルがテストデータになるように繰り返すことで識別率を求めた。学習データ数が十分に多い場合には、学習データに対する SVM の認識率により特徴選択の指標を得られる可能性もあり、この場合における SBS-CM 手法の検証については今後の検討課題としたい。

sonar データを用いて、パラメータを $\sigma = 1.8$, $C = 10$ としたときの SBS-CM の実験結果を図 7 に示す。図 7 の横軸は次元数、縦軸の左が Confident Margin(CM), 右が Recognition Rate(RR) である。また SVM-RFE による実験結果を図 8 に示す。

SBS-CM による特徴選択の結果、評価値 CM の値が最大となったのは 15 個の特徴の時であり、そのときの識別率は 92% であった。また識別率のみに注目したとき、識別率のピークは 13 個の特徴の時に 93% であった。図 7 から分かるように、CM の値と RR の値がほぼ追従するように変化している。そのため評価値のピークと識別率のピークが非常に近接するという結果であった。すなわち SBS-CM を特徴選択に用いた場合、評価値のみから最適な特徴セットをおおよそ

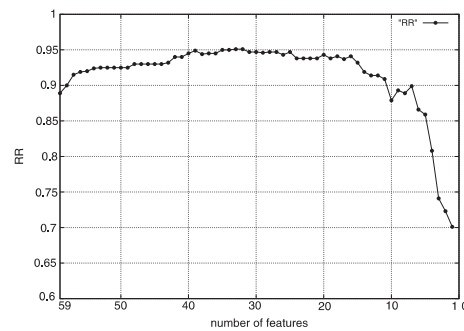


図 9 SBS-SVM での RR の変化 (sonar データ)
Fig. 9 Variation of RR using SBS-SVM(sonar data)

求めることが可能である。

一方、SVM-RFE を用いた特徴選択の結果、評価値 SR のピークは 3 個の特徴の時で、そのときの識別率は 74% であった。また識別率のピークは 17 個の特徴の時に 87% であった。図 8 で示したように SVM-RFE での評価値 SR と RR の関係は、SBS-CM における CM と RR のような関係性があるとは言えない。したがって評価値のみから最適な特徴セットを求めることは困難である。また識別率のピークが 87% であり、SBS-CM における識別率のピークが 93% であったことより、SVM-RFE は SBS-CM に比べて良い特徴選択ができなかったと言える。

また SVM による識別率を評価値とする SBS を用いた特徴選択 (以下 SBS-SVM と呼ぶ) の結果を図 9 に示す。

その結果 32,33 個の特徴で 95% の識別率であった。ここで表 2 に、それぞれの特徴選択における評価値が最大となった時の識別率と最良の識別率とそのときの次元数をまとめる。

表 2 各特徴選択における評価値最大の時の識別率と最良の識別率とそのときの次元数 (sonar データ)

Table 2 Recognition rate and the number of features at the peak of criterion score and its maximum of each method(sonar data)

Feature Selection	Max. criterion value's RR		Best RR	
	#DIM	RR[%]	#DIM	RR[%]
SBS-CM	15	92	13	93
SVM-RFE	3	74	17	87
SBS-SVM	-	-	32,33	95

識別率のみに注目すれば SBS-SVM が最も良い結果を与えるが、このアルゴリズムは削除する特徴を決定するために Leave One Out による SVM での学習と

認識を行わなければならないラッパーモデルであるため、莫大な計算コストがかかり実用的ではない。一方フィルター型の SBS-CM は SVM の学習のみで削除する特徴を決定できるので計算コストは少ない。また同じフィルター型の SVM-RFE と比較しても、より良い特徴セットを見つけることができたため識別率のピークも高く、評価値のみからおおよそ特徴セットの良さを判断することも利点である。

実際の計算コストの例を示す。Xeon 2.6GHz メモリ 1GByte の計算機を用いて演算を行った結果、sonar データを用いて特徴選択を行った場合、SBS-CM を用いた場合に約 10 分、SVM-RFE では約 60 分で特徴選択が終了するのに対し、SBS-SVM では約 20 時間が必要であった。このことから SBS-CM, SBS-RFE に対する SBS-SVM の計算コストの大きさが確認できる。

CM と Confidence の関係を調べるため、図 10 に SBS-CM の実験結果 (図 7) に Confidence c および Normal Margin (NM) の変化を追加したグラフを示す。この図より、次元数が少なくなった場合に、NM が急激に変動しており、これと相反する向きに c の値が変化している様子が確認できる。これは 4.1 で述べた NM における b-SV の問題が次元数が少なくなったために顕著に現れ、これを補正するように c の値が変化していると考えられる。また、補正後のマージン CM の値の変化は Recognition Rate (RR) の変化によく追従しており、提案した Confidence を用いて MN を補正することで効果的な特徴選択の指標を得ることができたといえる。

本論文では特徴選択における評価基準として CM を提案したが、Confidence c を評価基準として使用することも考えられる。そこで、確認のため図 11 に Confidence を評価基準としたときの特徴選択に対する Confidence の変化と RR の変化を示した。この結果より、Confidence の変化のピークは RR に追従しておらず、CM を評価基準として用いることの有効性が確認できた。

次に ionosphere データを用いて、パラメータを $\sigma = 5.0$, $C = 10$ としたときの SBS-CM の実験結果を図 12 に、SVM-RFE による実験結果を図 13 に示す。

SBS-CM による特徴選択の結果、評価値 CM のピークは 15 個の特徴の時に、そのときの識別率は 93% であった。またこの問題において CM のピークと RR のピークが一致した。

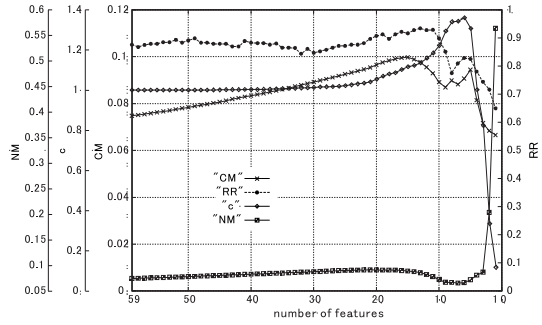


図 10 CM, NM, Confidence ならびに RR の変化 (sonar データ)

Fig. 10 Variation of SR criterion, RR, NM and Confidence (sonar data)

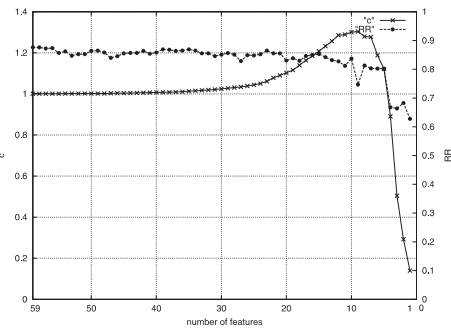


図 11 Confidence を評価基準とした場合の Confidence と RR の変化 (sonar データ)

Fig. 11 Variation of Confidence and RR evaluated by Confidence (sonar data)

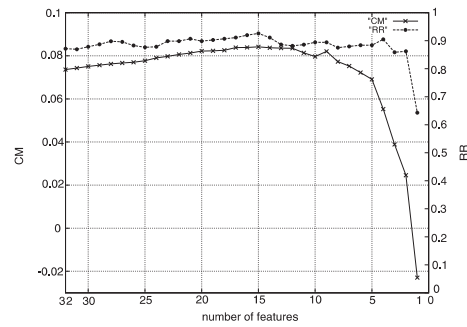


図 12 CM と RR の変化 (ionosphere データ)

Fig. 12 Variation of CM and RR (ionosphere data)

SVM-RFE を用いた場合、評価値 SR のピークは 3 個の特徴の時に、そのときの識別率は 72% であった。また識別率のピークは 14 個の特徴で 90% であった。図 13 に示したようにこの問題においても sonar データの時と同様に、SR と RR の関係性は全くなかった。また SBS-SVM による結果を図 14 に示す。

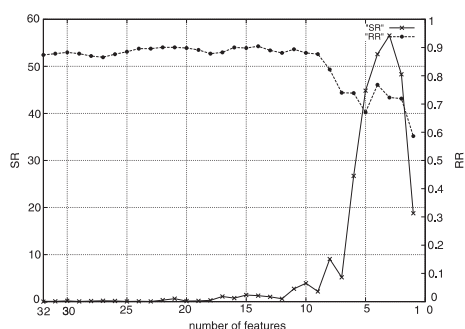
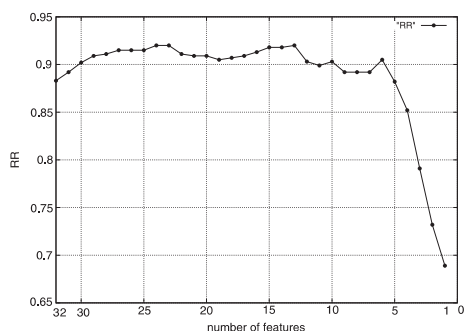


図 13 SR と RR の変化 (ionosphere データ)

Fig. 13 Variation of SR criterion and RR (ionosphere data)

図 14 SBS-SVM での RR の変化 (ionosphere データ)
Fig. 14 Variation of RR using SBS-SVM (ionosphere data)

その結果 23,24 個の特徴で 92% の識別率であった。表 3 に各特徴選択における評価値が最大となった時の識別率と最良の識別率とそのときの次元数をまとめる。

表 3 各特徴選択における評価値最大の時の識別率と最良の識別率とそのときの次元数 (ionosphere データ)
Table 3 Recognition rate and the number of features at the peak of criterion score and its maximum of each method (ionosphere data)

Feature Selection	Max. criterion value's RR		Best RR	
	#DIM	RR[%]	#DIM	RR[%]
SBS-CM	15	93	15	93
SVM-RFE	3	72	14	90
SBS-SVM	-	-	23,24	92

ionosphere データを用いた場合も, SBS-CM は SVM-RFE と比較してより良い特徴セットを見つけることができ, 評価値のみから特徴セットの良さを判断することができた。またラッパー型の SBS-SVM と同等の識別率を得ることができた。

以上の実験より, sonar データ ionosphere データのいずれを用いた場合においても SBS-CM における評

価値 CM と識別率 RR の追従性を確認することができた。すなわち評価値のみで特徴セットの良さをおよそ判断することが可能である。一方で SVM-RFE では評価値 SR と RR の関係性は全くなく, 評価値のピークから最適な特徴セットを求めることは困難であることがわかった。また識別率のみに注目した場合, SBS-CM は SVM-RFE よりも高い識別率を与える特徴を選択することができた。さらにラッパー型の SBS-SVM とほぼ同程度の識別率を与える特徴セットを見つけることができたと言える。

6. まとめ

本論文では Confident Margin という新しい評価値を定義し, それを用いた特徴選択 SBS-CM を提案した。Confident Margin を評価値とすることで, SVM-RFE のような Normal Margin をベースとする特徴選択よりも評価値と識別関数の良さが正しく対応するため, より良い特徴選択が行えることを示した。また SVM-RFE の場合, 評価値 SR の大きさと識別率の対応が全くなく評価値のみから良い特徴セットを見つけることは困難であった。一方 SBS-CM は, 評価値 CM の大きさと識別率の関係が強いため評価値のみから良い特徴セットを判別することがおよそ可能である。さらに SBS-SVM のようなラッパーモデルでは計算コストが大きくなるが, SBS-CM の計算コストは小さく大規模問題に対しても応用できる。

今回の実験においては特徴選択を行う際の SVM のパラメータは一定の値を用いているが, 選択された各特徴量に対して同一のパラメータが最適であるという保証はない。特徴選択時における SVM パラメータの最適値の変化については更なる検証が必要であろう。また今後は特徴選択アルゴリズムを floating 型アルゴリズムに拡張することで, さらに精度の高い特徴選択アルゴリズムを提案する予定である。

文 献

- [1] Cortes, C, Vapnik, V, "Support-Vector Networks," Machine Learning, vol.20, pp.273-297, 1995.
- [2] Michael P.S.Brown, William Noble Grundy, David Lin, Nello Cristianini, Charles Walsh Sugnet, Terrence S.Furey, Manuel Ares, Jr., David Haussler, "Knowledge-based analysis of microarray gene expression data by using support vector machines," PNAS, vol.97, no.1, pp.262-267, January, 2000.
- [3] M.P.S.Brown, W.N.Grunsky, D.Lin, N.Cristianini, C.Sugnet, M.Ares, D.Haussler, "Support vector ma-

- chine classification of microarray gene expression data," Technical Report UCSC-CRL-99-09, University of California, Santa Cruz, 1999.
- [4] V.N.Vapnik, Statistical Learning Theory, Wiley Interscience, 1998.
- [5] Isabelle Guyon,"Gene Selection for Cancer Classification using Support Vector Machines," Machine Learning, vol.46, pp.389-422, 2002.
- [6] Nello Cristianini,John Shawe-Taylor, An Introduction to Support Vector Machines, Cambridge University Press, 2000.
- [7] Kohavi,R,John,G,"Wrapper for feature subset selection" Artificial Intelligence, 97, pp.273-324, 1997.
- [8] Anil Jain,Douglas Zongker,"Feature Selection : Evaluation,Application,and Small Sample Performance," IEEE Trans. Pattern Analysis and Machine Intelligence, vol.19, no.2, pp.153-158, 1997.
- [9] Whitney,A.W,"A direct method of nonparametric measurement selection," IEEE Trans. Comput.20 , pp.1100-1103, 1997.
- [10] Marill,T,D.M.Green,"On the effectiveness of receptors in recognition system," IEEE Trans. Inform. Theory 9, pp.11-17, 1963.
- [11] P.Pudil, J.Novovičová, J.Kittler, "Floating Search Methods in Feature Selection," Pattern Recognition Letters, vol.15, no.11, pp.279-283, 1994.
- [12] C.L.Blake,C.J.Merz,"UCI Repository of machine learning databases," University of California, Irvine, Dept.of Information and Computer Sciences, 1998, <http://www.ics.uci.edu/~mllearn/MLRepository.html>

(平成 xx 年 xx 月 xx 日受付)

青木 一真

平 15 名工大・電気情報工卒. 平 17 同大大学院工学研究科情報工学専攻博士前期課程了. 同年 中部電力(株)入社.

黒柳 奨 (正員)

平 3 名工大・電気情報卒. 平成 5 同大大学院博士前期課程了. 平 8 同大大学院博士後期課程了. 同年名工大・電気情報・助手. 平 15 同大大学院・助手, 現在に至る. ニューラルネットワーク, 聴覚情報処理に関する研究に従事. 博士(工学). 日本音響学会, 日本神経回路学会, 日本エムイー学会会員.

Mauricio Kugler (学生員)

平 13 Electrical Engineering, Federal Center of Technological Education, Brazil 卒. 平 15 MSc in Biomedical Engineering, Federal Center of Technological Education, Brazil 了. 現在, 名工大大学院工学研究科情報工学専攻博士後期課程に在学中. マシンラーニング, 大規模分類問題, 特徴選択, バイオインフォマティクス研究に従事.

Anto Satriyo Nugroho (正員)

平 1 インドネシア協和国 Agency for the Assessment and Application of Technology に採用. 平 7 名工大・電気情報工卒. 平 15 同大大学院工学研究科電気情報工学専攻博士後期課程了. 博士(工学). 同年 中京大学・情報科学部メディア科学科客員教授. 平 16 同大学生命システム工学部身体システム工学科客員教授. ニューラルネットワーク, パターン認識, バイオインフォマティクス研究に従事. IEEE, IPSJ, ISCB, JSBI 学会各会員.

岩田 彰 (正員)

昭 48 名大・工・電気卒. 昭 50 同大大学院修士課程了. 同年名工大・情報・助手. 昭 57 年 4 月より昭 58 年 10 月まで, ドイツ連邦共和国ギーセン大学医学部医用情報研究所客員研究員. 昭 59 名工大・情報・助教授. 平 5 同大・電気情報・教授. 平 14 同大・副学長, 平 16 同大大学院・教授, 現在に至る. ニューラルネットワーク, 情報セキュリティに関する研究に従事. 工博. H5 年度本会論文賞受賞, H10 年度情報処理学会 Best Author 賞受賞. 情報処理学会, 日本ME学会, 日本心電図学会, 日本神経回路学会, 日本医療情報学会各会員, IEEE Senior Member

Abstract The aim of this study is to develop a feature subset selection method based on the concept of margin in SVM. The problem of this approach is that the margin of SVM does not always provide clear relationship between its value and the SVM's performance. The subset obtained by the selection algorithm does not correspond to the best performance subset. In this paper, we describe a new solution by introducing Confident Margin as the feature selection criterion. The experimental results show the possibility to achieve a result closer to the best recognition rate by monitoring the peak of CM curve.

Key words SVM, Margin, Feature Selection, SBS, Confident Margin