

SVM を用いた高累積正解率を持つ多クラスパターン識別

丸山 健一[†] 丸山 稔[†]
 宮尾 秀俊[†] 中野 康明[†]

本研究の目的は SVM を用いて高い累積正解率を持った多重仮説を構築することである。ある 1 つの候補を求める手法としては、DAGSVM が学習、識別時間の面で非常に優れている。しかしながら、累積正解率、つまり複数の候補を求めることを考慮していない。本論文では DAGSVM と Max-Win アルゴリズムを用いたハイブリッドな認識手法を提案する。具体的には、DAGSVM の結果に基づき、Max-Win アルゴリズムの候補を限定する。実験の結果、提案手法は Max-Win アルゴリズムと同等の累積正解率でありながら、DAGSVM と同等の計算量であった。

A Method to Make Multiple Hypotheses with High Cumulative Recognition Rate Using SVMs

KEN-ICHI MARUYAMA,[†] MINORU MARUYAMA,[†] HIDETOSHI MIYAO[†]
 and YASUAKI NAKANO[†]

This paper describes a method to make multiple hypotheses with high cumulative recognition rate using SVMs. To make just a single hypothesis by using SVMs, it has been shown that DAGSVM is very good with respect to recognition rate, learning time and evaluation time. However, DAGSVM is not directly applicable to make multiple hypotheses. In this paper, we propose a method of DAGSVM and Max-Win algorithm. We also provide the experimental results to show that the cumulative recognition rate of our method is as good as the Max-Win algorithm, and that the execution time is almost as fast as DAGSVM.

1. はじめに

近年注目されているパターン認識手法の 1 つに Vapnik によって提案された Support Vector Machine (SVM)¹⁾がある。この分野の研究は幅広く進められており、文字認識や数字認識にも適用され、高い認識率を持つことが報告されている²⁾。

本研究では、多クラス問題に対する SVM の適用法について述べる。SVM は 2 クラス識別器であるため、多クラス (N クラス) 問題に直接適用することはできない。2 クラス識別器を多クラス問題に適用する典型的な方法としては、あるクラスとその他のクラス間で識別を行う識別器を学習して用いる 1-v-r (one versus the rest) や、すべてのクラス対ごとに識別器を構成し、それらを組み合わせる 1-v-1 (one versus one) がある。

これらを SVM に適用した場合、1-v-r は学習にすべてのデータを用いた N 個の SVM を構築する必要がある。一方、1-v-1 の場合は学習に 2 クラス分のデータを用いた $N(N-1)/2$ 個の SVM が必要となる。ゆえに、1-v-1 は構築しなければならない SVM の個数は 1-v-r より多くなるが、それぞれの SVM を構築する際に用いるデータ数は 1-v-r より少ないため、1-v-r よりも高速な学習が可能であることが知られている³⁾。

1-v-1 タイプの SVM を多クラス問題に適用する代表的な手法として、Max-Win アルゴリズムがあげられる⁴⁾。このアルゴリズムの識別時間は $O(N^2)$ であり、1-v-r の識別時間 $O(N)$ に比べきわめて遅い。この問題を解決するために、DAGSVM (Directed Acyclic Graph Support Vector Machines) が近年提案された⁵⁾。この手法は決定木を基にした手法であり、識別時間は $O(N)$ である。

ある 1 つの候補が必要な場合は、DAGSVM が 1-v-r や Max-Win アルゴリズムよりも、学習時間、識

[†] 信州大学工学部情報工学科

Department of Information Engineering, Shinshu University

別時間の観点から考えて優れている。しかしながら、文字認識などで重要となる多重仮説が必要となる場合には、1-v-r や Max-Win アルゴリズムでは、簡単に多重仮説を与えることができるが、決定木を基にした DAGSVM は明らかに適していない。

本研究では、高速学習と高い累積正解率をともに実現するための DAGSVM と Max-Win アルゴリズムのハイブリッド法を提案する。高速な学習時間と識別時間を生かすために、基本的には DAGSVM を用いる。それぞれの入力パターンに対し、DAGSVM の結果を基にクラスを限定する。その後、多重仮説を得るために、限定したクラスに対し、Max-Win アルゴリズムを適用する。

2. 1-v-1 SVM による高速な多重仮説の計算法

2.1 ハイブリッド法

1-v-1 の学習時間の短さを生かして複数の候補を求めるためには Max-Win アルゴリズムを用いればよいが、識別時間が遅いという難点がある。Max-Win アルゴリズムが多大な識別時間を要するのは不必要な比較を行っているためであると考えられる。真のクラスを含む限定されたクラスの候補の集合を高速に求めることができれば、それらのクラスに対し Max-Win アルゴリズムを適用すればよいので、高速な識別が可能となる。そこで、本研究では高速に候補を求めるために、DAGSVM を用いる。

一般に、DAGSVM は高速であり、単一の識別結果ではあるものの、良好な結果を得ることが期待できる。このとき結果のクラスは、真のクラスではなかったとしても、それと非常に近いと考えられる。以上の考察から、ここではまず DAGSVM を適用して識別結果を求め、それに基づいて候補クラスを限定し、これに対して Max-Win アルゴリズムを適用することを考える。

2.2 クラス間の近さの決定

2.1 節で述べた手法を適用するためには、クラス間の近さを前もって決定しなければならない。クラス間の距離尺度は様々なものが考えられるが、本研究における目的は SVM が誤りの解を与えてしまうようなクラス候補を抽出することであるので、1-v-1 タイプの SVM の識別困難さを「近さ」の尺度として用いるものとする。

2つのクラスが識別困難であれば、対応する SVM の認識率は悪いはずである。つまり、それぞれの SVM の誤認識確率 (generalization error) はクラスの近さを決定するために適しているが、generalization error

を得ることは一般に困難である。代わりに学習エラーを用いることも考えられるが、学習エラーが十分小さくても、generalization error は高い場合があるため適当ではないと考えられる。そのため、本研究では、以下に示す generalization error の推定値を用いた⁶⁾。

$$\frac{\sum_{i \in SV} \alpha_i}{Cl} \quad (1)$$

ただし、 α_i はデータ i に対するラグランジュ乗数、 l は学習データ数、 C は学習エラーに対するパラメータである。この近さの尺度を用い、学習の後に、それぞれのクラスに対する近傍クラスの表を作成する。その表から、尺度の大きさに応じてある個数のクラスを選ぶ。クラスの集合が特定されれば、多重仮説を決定することは容易である。手順を以下に示す。

- (1) 入力パターン \mathbf{x} に対し DAGSVM を適用する。識別結果がクラス c だとする。
- (2) T_c を事前に特定したクラス c の近傍クラスとする。Max-Win アルゴリズムを $T = T_c \cup \{c\}$ に対し適用する。
- (3) 得票数の多い順に K 個のクラスを決定する。

2.3 パターン依存情報の利用

前節で述べたアルゴリズムは、事前に計算された近傍クラスに強く依存しているため、入力パターン \mathbf{x} への依存が弱いと考えられる。近傍クラスの集合は SVM の学習結果に基づき決定されるため、それぞれのクラスの平均的な性質を反映している。このため、2つのパターン \mathbf{x} と \mathbf{x}' が同じクラスに識別された場合は、 \mathbf{x} と \mathbf{x}' がまったく異なっても、近傍クラスは同じとなる。入力パターン自身の特性を取り入れるために、アルゴリズムを以下のように改良する。

- (1) 入力パターン \mathbf{x} に対し、DAGSVM を適用する。クラス c が認識結果であったとする。
- (2) 入力パターン \mathbf{x} に対し、 c と c_m ($m = 1, \dots, N$) 間の 1-v-1 SVM を適用する。 $f_{cc_m}(\mathbf{x})$ を SVM の出力とする。マージン $d_{cc_m}(\mathbf{x})$ を以下のように計算する。

$$d_{cc_m}(\mathbf{x}) = \frac{f_{cc_m}(\mathbf{x})}{\|\omega_{cc_m}\|} \quad (2)$$

$$\|\omega_{cc_m}\| = \sqrt{\sum_{i=1}^{\ell_c} \sum_{j=1}^{\ell_{c_m}} \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) y_i y_j}$$

ここで、 y_i は、データ i に対する教師信号である。

- (3) 以下の条件を満たすクラス c_m を検索する。

$$d_{cc_m} < \theta \quad (3)$$

ここで、 θ は閾値である。 S_c を抽出したクラ

表 1 認識結果

Table 1 Recognition results.

データベース	手法		累積正解率					計算量の比較
			1	2	3	4	5	
JEITA-HP クラス: 74 学習データ: 42,838 テストデータ: 42,828	1-v-r	学習	97.43	99.02	99.36	99.57	99.68	6.72
		テスト	93.47	97.37	98.43	98.89	99.18	
	Max-Win アルゴリズム	学習	98.73	99.84	99.93	99.93	99.96	37.0
		テスト	94.3	98.58	99.27	99.5	99.63	
	DAGSVM (original)	学習	98.74	-	-	-	-	1.0
		テスト	94.05	-	-	-	-	
	DAGSVM (hyb1, Max-Win なし)	学習	98.74	99.71	99.77	99.8	99.81	1.0
		テスト	94.05	97.28	97.98	98.4	98.64	
	DAGSVM (hyb1, T = 10)	学習	98.73	99.77	99.82	99.83	99.83	1.62
		テスト	94.17	98.3	98.91	99.09	99.17	
DAGSVM (hyb2, Max-Win なし)	学習	98.74	99.74	99.84	99.86	99.87	2.0	
	テスト	94.05	97.92	98.66	99.02	99.19		
DAGSVM (hyb2, $\theta = 0.01, N_d = 2$)	学習	98.72	99.85	99.94	99.95	99.96	2.01	
	テスト	94.21	98.61	99.26	99.49	99.59		
MNIST クラス: 10 学習データ: 60,000 テストデータ: 10,000	1-v-r	学習	97.99	99.19	99.58	99.75	99.85	6.7
		テスト	97.43	99.1	99.61	99.83	99.92	
	Max-Win アルゴリズム	学習	98.81	99.65	99.82	99.88	99.91	5.0
		テスト	97.16	99.03	99.62	99.74	99.83	
	DAGSVM (original)	学習	98.81	-	-	-	-	1.0
		テスト	97.11	-	-	-	-	
	DAGSVM (hyb1, Max-Win なし)	学習	98.81	99.13	99.37	99.6	99.73	1.0
		テスト	97.11	97.79	98.32	98.91	99.25	
	DAGSVM (hyb2, Max-Win なし)	学習	98.81	99.53	99.72	99.83	99.88	2.0
		テスト	97.11	98.73	99.36	99.67	99.83	
DAGSVM (hyb2, $\theta = 0.01, N_d = 3$)	学習	98.81	99.64	99.81	99.86	99.92	2.3	
	テスト	97.16	99.04	99.62	99.68	99.85		
USFS クラス: 7 学習データ: 11,340 テストデータ: 565,892	1-v-r	学習	99.63	99.98	99.99	100	100	2.6
		テスト	71.27	90.91	94.6	95.41	96.11	
	Max-Win アルゴリズム	学習	99.63	100	100	100	100	3.5
		テスト	72.75	95.84	99.28	99.88	99.99	
	DAGSVM (original)	学習	99.63	-	-	-	-	1.0
		テスト	72.59	-	-	-	-	
	DAGSVM (hyb1, Max-Win なし)	学習	99.63	99.9	99.96	99.96	100	1.0
		テスト	72.59	96.08	97.29	99.65	99.86	
	DAGSVM (hyb2, Max-Win なし)	学習	99.63	99.98	100	100	100	2.0
		テスト	72.59	94.86	98.56	99.36	99.69	
DAGSVM (hyb2, $\theta = 0.01, N_d = 1$)	学習	99.63	99.94	99.97	99.97	100	2.0	
	テスト	72.75	95.98	98.23	99.7	99.91		

ス, S を $S = S_c \cup \{c\}$ とする.

- (4) T_c を 2.2 節で述べた近傍クラスの集合とする. $M = |S|$ とする. S に対し Max-Win アルゴリズムを適用し, M 個の仮説を決定する.
- (5) $K > M$ の場合, 残りの $(K - M)$ 個の仮説を表 $T_c - S$ を元に決定する.

3. 認識実験および考察

表 1 に認識結果および従来の DAGSVM との計算量の比較を示す. 表 1 において, JEITA-HP は (社) 電子情報技術産業協会より提供されている手書き漢字データベースであり, その中のひらがな 74 種を使用した. 各クラスの奇数番目を学習データ, 偶数番目をテ

ストデータとして用いた. MNIST (Modified National Institute of Standards and Technology) は NIST's Special Database 1 および 3 から作成された手書き数字のデータベースである. USFS は University of California Irvine が公開している UCI Repository of Machine Learning Database 中の 1 つであり森林の植生のデータである. 学習, テストデータ数は文献 5) に従った.

JEITA-HP および MNIST においては, 手書き文字認識で有効とされている方向性パターン特徴を用いた⁷⁾. 特徴量は 256 次元である. USFS は 54 特徴 (10 連続値, 44 二値) であるが, 連続値の特徴を文献 5) に従い, ゼロ平均, 単位分散を持つように正規化した.

実験に用いたカーネル関数は以下に示すガウシアンカーネルである。

$$k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{\sigma^2}\right) \quad (4)$$

σ^2 と学習の際のパラメータ C は 1-v-r, 1-v-1 で同じ値を用いた。JEITA-HP, MNIST においては $\sigma^2 = 2 \times 256 \times 256$, $C = 1000$ であり, USFS においては, $\sigma^2 = 1.0$, $C = 10$ と設定した。これらの値は実験的に定めた。なお, 学習の高速化のため, SVM^{light}⁸⁾を用いた。

表1において, 1-v-rは1-v-rを用いた結果を示し, DAGSVM (original)は従来のDAGSVMを用いた結果を示す。DAGSVM (hyb1, $|T| = 10$)は2.2節のアルゴリズムを用いた場合の結果を示し, $|T| = 10$ の場合の結果を示す。DAGSVM (hyb1, Max-Winなし)は, 2.2節のアルゴリズムを用いているが, Max-Winアルゴリズムは適用せず, クラスの近さの尺度のみに応じて順位を付けた場合を示す。DAGSVM (hyb2, $\theta = value$, $N_d = value$)は2.3節のアルゴリズムを用いた場合であり, 閾値 $\theta = value$, M の平均が N_d の場合を示す。DAGSVM (hyb2, Max-Winなし)は2.3節のアルゴリズムを用いた場合であるが, Max-Winアルゴリズムは適用せずに, マージンの小さい順に順位を付けた場合の結果を示す。なお, 累積正解率において上位2位までの識別結果を太字で示している。

表1から, 提案手法により, 高速に高い累積正解率を持った多重仮説を得ることができていることが分かる。また, 2.3節のアルゴリズムの方が, 2.2節のアルゴリズムよりも累積正解率が優れている場合が多いことから, 入力パターンの性質を用いたほうが, それぞれのクラスの平均的な性質を用いるよりも良いことが分かる。

1-v-rを用いた場合, MNISTにおいては他の手法に比べ良い認識結果が得られているが, 学習時間, および識別時間がかかってしまうので実用的ではないと考えられる。また, クラス間の近さの尺度に応じて順位付けを行うことで, 容易に高い累積正解率を得ることができている。提案手法を用いることで性能を向上できているが, さらなる改善が必要である。

4. む す び

本研究では, DAGSVMとMax-Winアルゴリズムを用いた高い累積正解率を持つハイブリッド法を提案した。提案手法の効果を確かめるために, JEITA-HPのひらがな, MNIST, USFSのデータを用いて実験を行った。実験の結果, 累積正解率はMax-Winアルゴリズムとほぼ同等であり, 識別時間はMax-Winアルゴリズムよりもきわめて高速であった。

今後は漢字のように非常にクラス数が多い場合の対処などを調査していきたい。

参 考 文 献

- 1) Vapnik, V.: *Statistical Learning Theory*, Wiley, New York (1998).
- 2) Wang, F., Vuurpijl, L. and Schomaker, L.: Support Vector Machines for the Classification of Western Handwritten Capitals, *Proc. 7th IWFHR*, pp.167-176 (2000).
- 3) Hsu, C.-W. and Lin, C.-J.: A Comparison of Methods for Multi-class Support Vector Machines, *IEEE Trans. Neural Networks* 13, pp.415-425 (2002).
- 4) Kreßel, U.: Pairwise classification and support vector machines, *Advanced in Kernel Methods — Support Vector Learning*, Schölkopf, B., Burges, C.J.C. and Smola, A.J. (Eds.), pp.185-208, MIT Press (1999).
- 5) Platt, J.C., Cristianini, N. and Shawe-Taylor, J.: *Large Margin DAGs for Multiclass Classification*, pp.547-553, MIT Press (2000).
- 6) Maruyama, K., Maruyama, M., Miyao, H. and Nakano, Y.: Handprinted Hiragana Recognition Using Support Vector Machines, *Proc. 8th IWFHR*, pp.55-60 (2002).
- 7) Mori, S., Suen, C.Y. and Yamamoto, K.: Historical Review of OCR Research and Development, *Proc. IEEE*, Vol.80, No.7, pp.1029-1058 (1992).
- 8) Joachims, T.: Making large-scale SVM Learning Practical, *Advanced in Kernel Methods — Support Vector Learning*, Schölkopf, B., Burges, C.J.C. and Smola, A.J. (Eds.), MIT Press (1999).

(平成 15 年 3 月 19 日受付)

(平成 15 年 6 月 3 日採録)