

多クラス問題におけるSVMのカーネルパラメータ最適化

木村 幸代[†], 渡邊 真也^{††}

[†] 室蘭工業大学大学院 情報工学専攻 ^{††} 室蘭工業大学 情報工学科

本研究では、多クラス Support Vector Machine(SVM) に対する多目的最適化を用いたカーネルパラメータ最適化を行った。カーネル関数としては、通常の RBF カーネルを発展させた重み付き RBF カーネルを採用し、評価基準の設定による識別性能への影響について調査した。vehicle データと glass データを用いた実験結果より、サポートベクターの数, AUC(avg), AUC(min) の 3 つを組み合わせたものが最も良い精度となることがわかった。

Kernel Parameter Optimization of Support Vector Machine in Multiclass problem

Yukiyo KIMURA[†] Shinya WATANABE^{††}

[†] Graduate School of Muroran Institute of Technology

^{††} Department of Computer Science & Systems Engineering Muroran Institute of Technology

This paper tries to apply evolutionary multi-criterion optimization (EMO) to the parameters estimation problem of multi-class support vector machines (SVMs). As the kernel function, we choose the weighted radial basis function (RBF) kernel, which the number of parameters equals the dimension of the input patterns. In this paper, we investigate which combination of objectives should be used to designing the kernel parameters of SVMs through computational experiments on vehicle and glass. Through the results of examples, it turns out that the combination of #SV and AUC(avg) and AUC(min) could be more suitable for this problem.

1 はじめに

近年, SVM(Support Vector Machine) と呼ばれるパターン認識手法が注目され, 幅広い分野からの成果が報告されている¹⁾. SVM が注目される背景には, 現在知られている多くのデータ解析手法の中で優れた識別精度を有していること, カーネルトリックにより非線形識別が可能であること, マージン最大化等により汎化能力を保つ工夫がなされていること等があげられる²⁾.

SVM は本来 2 クラスの識別器であるが, 多クラスへの拡張も可能となっている. 拡張法としては様々なものが提案されているが, 本研究では最も高い精度を持つ OAO(One Against One)³⁾ を用いて多クラス問題への拡張を行った.

この SVM においてもっとも重要となるのは, カーネル関数のパラメータの設定である. このパラメータにより, 識別精度が左右されるにも関わら

ず, パラメータ設定法は定まっていない. また, パラメータを設定することにより得られた識別器の評価方法は複数存在しており, 多クラス問題の場合はより様々な要素を考慮しなければならないので, どの評価基準を採用すればよいのかというのが一概には言えなくなってしまう.

そこで本研究では, UCI レポジトリ⁶⁾ から入手した vehicle データと glass データに対して, SVM のカーネルパラメータを多目的遺伝的アルゴリズムを用いて最適化し, 最適化過程での識別器の評価方法としてサポートベクターの数, AUC(avg), AUC(min) の 3 種類を採用して様々な組み合わせでの実験を行った.

また, 識別器の評価時に Cross Validation Method(CVM) を用いて各評価値を計算しているが, 通常の CVM だけではなく, 本研究で提案する Inverse-CVM(I-CVM) を用いての実験も行った.

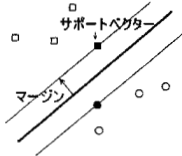


Fig. 1 分離平面とマージン

2 Support Vector Machine(SVM)

2.1 SVMとは

SVM(Support Vector Machine)とは,1960年代に Vapnik 等が考案した Optimal Separating Hyperplane を起源とし,1990年代になってカーネル学習法と組み合わせた非線形の識別手法へと拡張された²⁾. 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と考えられている.

2.2 学習方法

SVM は線形しきい素子を用いて2クラスのパターン識別器を構成する手法である²⁾.

入力特徴ベクトルに対し, (1) 式の識別関数により出力値を計算する.

$$y = \text{sign}(w^t \mathbf{x} - h) \quad (1)$$

この関数は $u > 0$ なら 1, $u \leq 0$ なら -1 を取る符号関数であり, w はシナプス荷重に対応するパラメータで, h は閾値である. このモデルは, 入力ベクトルとシナプス荷重の内積が閾値を超えれば 1 を出力し, 越えなければ -1 を出力する. これは識別平面により入力特徴空間を2つに分けることに相当する.

Fig.1に示すように, このモデルに2クラスの入力ベクトルが与えられた際, 他クラスに最も近い入力をサポートベクターとし, サポートベクター間の距離をマージンと呼ばれる量で定義する. マージンが最大となるような識別平面となるパラメータ w, h を求めることにより, SVM はクラス識別を行う.

2.3 非線形 SVM

前節の基本的な構造だけでは, 入力線形分離不可能だった場合に適用することができない. 非線形分類を可能にするために, 入力ベクトルを非線形写像 $\phi(\mathbf{x})$ によって高次元空間へ写像し, 写像先の空間において線形分離を行う.

しかしながら, $\phi(\mathbf{x})$ の計算は非常に高コストである. そして SVM の学習過程において写像先の点同士の内積を計算する必要があるため, 結果的に膨大な計算量となってしまふ.

今, 非線形空間に写像した先での二つの要素 $\phi(\mathbf{x}_1)$ と $\phi(\mathbf{x}_2)$ との内積が, $\phi(\mathbf{x}_1)^T \phi(\mathbf{x}_2) = K(\mathbf{x}_1, \mathbf{x}_2)$ のように, 入力特徴 \mathbf{x}_1 と \mathbf{x}_2 から計算できるのならば, 高次元ベクトル同士の演算を行う必要がなくなり, 計算が非常に容易になる. この K はカーネルと呼ばれており, 一般的に線形カーネル, ガウスクーネル, RBF カーネルがよく知られている²⁾.

その中でも, 本研究では (2) 式に示す Radial Basis Function Kernel(以下 RBF カーネル) に注目する.

RBF kernel

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

また本研究では, (3) 式の入力ベクトルの次元毎に重みを付ける weighted RBF kernel(以下 w-RBF カーネル) についても実験を行う⁴⁾.

weighted-RBF kernel

$$K(\mathbf{x}, \mathbf{z}) = \exp\left(-\sum_j \frac{(x_j - z_j)^2}{2\sigma_j^2}\right) \quad (3)$$

(3) 式におけるパラメータ j は入力ベクトルの次元数 (特徴数) と等しい. この次元数が 3 または 4 を超える場合には, 網羅的探索手法によるパラメータ調整が非常に困難となる.

一方, 識別器に対する評価基準は様々なものがあり, 決定的な評価軸は今だ定まっていないのが現状である. そこで, 本研究では複数の評価軸を同時に扱うことのできる多目的遺伝的アルゴリズムを用いて, SVM のパラメータ最適化を行った.

2.4 多クラス SVM

SVM とは線形識別器であり, 本来2クラスの問題にしか対応できない. これを多クラス問題に発展させるために, One Against All(OAA), One Against One(OAO), DAGSVMS 等の手法が提案されてきたが, 本研究ではその中でも最も良い精度を持つ OAO を採用した³⁾. 例として, H,E,C という3種類のクラスを持つデータが与えられた場合の OAO 概念図を Fig. 2 に示す.

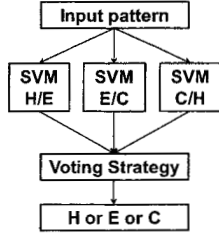


Fig. 2 one against one(OAO)

今, クラス数 k , 要素数 l のトレーニングデータが与えられたとする. この入力を $(x_1, y_1), \dots, (x_l, y_l), x_i \in R^n, i = 1, \dots, l, y_i \in 1, \dots, k$ と表すと, OAO は $k(k-1)/2$ 個の 2 クラス識別器を構成する. 入力データがクラス i か j のどちらに属するかを判断する 2 クラス識別器を構成するために, OAO は以下の式を用いる.

$$\min_{w^{ij}, b^{ij}, \xi_t^{ij}} \frac{1}{2} (w^{ij})^T w^{ij} + C \sum_t \xi_t^{ij}$$

$$(w^{ij})^T \phi(x_t) + b^{ij} \geq 1 - \xi_t^{ij}, \text{ if } y_t = i,$$

$$(w^{ij})^T \phi(x_t) + b^{ij} \leq -1 + \xi_t^{ij}, \text{ if } y_t = j,$$

$$\xi_t^{ij} \geq 0.$$

この式を全てのクラス組み合わせにおいて解くことにより $k(k-1)/2$ 個の決定関数 $\text{sign}((w^{ij})^T \phi(x) + b^{ij})$ が得られる. あるデータについてこの決定関数値を計算すると, そのデータがクラス i かクラス j のどちらに属するかが求められる. OAO は $k(k-1)/2$ 個全ての決定関数値を計算し, そのクラスに属すると判断された数を数え, 最も回数が多かったクラスであると予測を行う.

3 F-Score

本研究では, F-Score を利用して問題の難易度を評価した. F-Score は, クラス識別問題において各データの持つ特徴がクラスの差異にどの程度関わっているかを求めるための単純な手法であり, SVM に対して利用した研究も幾つか報告されている⁵⁾.

今, クラス p に対する各属性値の F-Score を考える. 対象データ n がクラス p に属する場合ポジティブラベル n^p , それ以外のクラスに属する場合ネガティブラベル $n^{\bar{p}}$ であるとする. クラスのラベルのついたデータ $\mathbf{x}_k, k = 1, \dots, m$ が与えられているとき, F-score ではデータの持つ i 番目のパラ

	vehicle	glass
クラス数	4	6
特徴数	18	9
データ数	846	214

メータのクラス p に関する決定への寄与度を $F(i)$ (F 値) として以下の式により求める.

$$F_{(i)}^p \equiv \frac{(x_i^{(p)} - x_i)^2 + (x_i^{(\bar{p})} - x_i)^2}{\frac{1}{n_+ - 1} \sum_{k=1}^{n_+} (x_{k,i}^{(p)} - x_i^{(p)})^2 + \frac{1}{n_- - 1} \sum_{k=1}^{n_-} (x_{k,i}^{(\bar{p})} - x_i^{(\bar{p})})^2}$$

上式における $x_i, x_i^{(p)}, x_i^{(\bar{p})}$ は i 番目の特徴におけるポジティブクラスやネガティブクラスの平均値を表している. また, $x_{k,i}^{(p)}$ および $x_{k,i}^{(\bar{p})}$ はそれぞれ k 個番目のポジティブおよびネガティブデータにおける特徴 i 番目の値を示している.

F-Score 関数における分子はポジティブおよびネガティブクラスの平均値との分散を表しており, 分母はポジティブおよびネガティブクラスのクラス内分散を表している. そのため, この F-Score 値の値が大きいほどクラス識別への寄与度が高いと判断することができる.

4 数値実験

web 上のデータベースサイトである UCI レポジトリ⁶⁾ から乗り物のデータ (vehicle) と, 硝子のデータ (glass) を入手し, SVM を用いた識別実験を行った. 各データの詳細を Table 1 に示す.

本研究では多目的遺伝的アルゴリズムである SPEA2+ を用いて, RBF カーネルのパラメータ σ と, 識別間違いを犯す毎に加算されるペナルティパラメータ C の値を世代数 100, 個体数 100 として最適化する.

4.1 F-score

ここで, Table 1 に示す問題の難易度を把握するため, F-Score を適用した結果を Table 2, Table 3 に示す.

結果から, 全体的に glass データの方が f-score が高く, クラス分類を行いやすいことがわかる. 対して vehicle データは, どのクラスに関しても F-Score

Table 2 F-Score(vehicle)

	1st	2st	3st
$F_{(i)}^1$	0.114(3)	0.093(7)	0.087(8)
$F_{(i)}^2$	0.109(1)	0.097(3)	0.090(14)
$F_{(i)}^3$	0.240(14)	0.195(18)	0.078(3)
$F_{(i)}^4$	0.715(8)	0.637(7)	0.628(12)

Table 3 F-Score(glass)

	1st	2st	3st
$F_{(i)}^1$	0.361(3)	0.277(4)	0.092(8)
$F_{(i)}^2$	0.098(2)	0.047(8)	0.035(9)
$F_{(i)}^3$	0.335(3)	0.159(4)	0.104(8)
$F_{(i)}^4$	1.263(3)	0.502(4)	0.269(2)
$F_{(i)}^5$	0.876(2)	0.588(3)	0.572(6)
$F_{(i)}^6$	1.796(3)	1.460(8)	1.269(4)

が1を超えることが無く、分類が難しいデータだということが読み取れる。

4.2 識別精度

今回 SPEA2+⁷⁾ によりパラメータ最適化を行う際の目的関数として以下の3つを採用した。

- AUC(avg)
- AUC(min)
- #SV(Number of Support Vector)

この中で#SVだけは最小化とし、他の値については全て最大化するよう最適化を行った。また AUC とは Fig. 3 に示す ROC 曲線の下側の面積を示す。

ROC 曲線とは、閾値を変化させ、それぞれの閾値での (4) 式による Sensitivity を縦軸に、(5) 式の 1-Specificity を横軸にプロットしたものである⁸⁾。

$$\text{Sensitivity} = \frac{TP}{P} \quad (4)$$

$$\text{Specificity} = \frac{TN}{FP + TN} \quad (5)$$

ROC 曲線では、識別能力のない識別器は A のような対角線上に曲線を描き、識別能力が向上するほど、B、C のように対角線から左上に弧を描く曲線と

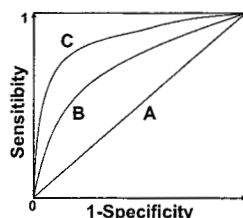


Fig. 3 ROC 曲線

		True class	
		P	N
Hypothesized class	PP	True Positive (TP)	False Positive (FP)
	NN	False Negative (FN)	True Negative (TN)

Fig. 4 各値名称

なり、100%識別できる場合は左辺-上辺上の曲線となる。よって AUC の値が1に近い程、優れた識別器であるといえる。

本研究では多クラス分類に OAO を利用しているので、学習用データが k 個のクラスラベルを持っていた場合、 $k(k-1)/2$ 個の識別器が構築される。そのため、AUC についても $k(k-1)/2$ 個の値を求めることが出来る。AUC(avg) は、全 AUC 値の平均を最大化するように目的関数を設定することとし、AUC(min) は全 AUC 値の中で最も小さな値を最大化するように目的関数の設定を行う。

4.3 評価方法

一般的に識別器を構成する際には、学習用データで識別器の学習を行い、その識別器を用いてテストデータの識別を行い、その際の識別率を結果とする。CV 法 (Cross Validation method: 交差検定法) とは、ランダムに一定数のデータを抜き取り、抜き取った残りのデータで識別モデルを求め、抜き取ったデータでモデルの識別誤差を評価するという手法であり、本研究においても採用している。CV 法のアプローチを以下に示す。

- 1) データ X を m 個のグループ X_1, X_2, \dots, X_m に分割する。
- 2) X_i を除いた $m-1$ 個のグループのデータを学習

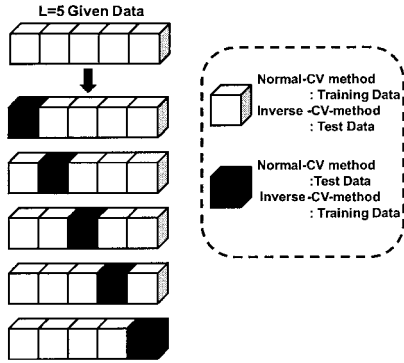


Fig. 5 Image of N-CV and I-CV

用データとする.

- 3) 学習用データで識別器の学習を行う.
- 4) X_i のデータをテストデータとし, 識別器のテストを行い, 識別器の精度を得る.
- 5) 2) から 4) の手順を m 回繰り返し, 得られた m 個の精度の平均をとったものを, この識別器全体での精度とする.

本研究では, 一般的な CV 法における学習用データとテスト用データの割合を反転させた場合についても実験を行った. ここでは, 通常の CV 法を N-CV 法 (Normal-CV 法), 反転させた場合を I-CV 法 (Inverse-CV 法) として表す. それらの概念図を Fig. 5 に示す. Fig. 5 の中で白いデータを学習用データ, 黒いデータをテスト用データとした場合が N-CV 法であり, それらを反転させたものが I-CV 法である.

何故 I-CV 法を用いるかということ, 与えられた学習用データというのは実問題におけるほんの一部にすぎない. そこで, I-CV を用いてより少ない学習用データで学習を行い, 多量のデータによる識別器の評価を行うことにより, 未知の入力に対する汎化能力を意識した学習を行うことができる.

4.4 実験結果

4.2 節で示した 3 種類の目的関数を Table 4 に示すように組み合わせ, 多目的遺伝的アルゴリズムを用いた実験を行った. 多目的遺伝的アルゴリズムで用いたパラメータを Table 4, Table 5 に示す. 各

Table 4 組み合わせパターン

パターン番号	目的関数
pattern1	AUC(avg)
pattern2	AUC(min)
pattern3	#SV
pattern4	AUC(avg), AUC(min)
pattern5	AUC(avg), #SV
pattern6	AUC(min), #SV
pattern7	AUC(avg), AUC(min), #SV

Table 5 遺伝パラメータ

パラメータ	値
世代数	100
個体数	100
突然変異率	0.05
交差率	1

目的関数組み合わせ毎の実験結果を Fig. 6 ~ Fig. 9 に示す.

4.5 考察

結果を見ると, AUC(avg) の値に大差は無いが, #SV に多少のばらつきがみえる. #SV の数というのは, 識別平面の滑らかさを表しており, SVM の汎化能力に影響を与える. #SV の数が少ない方が, より滑らかな識別平面となり, 未知のデータに対する汎化能力を保つことができる. なので, AUC(avg) の値が高く, #SV が小さい場合が優れた結果であるとする.

全ての結果に共通して, #SV の単一目的最適化

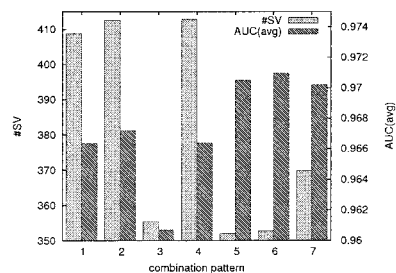


Fig. 6 vehicle, N-CV

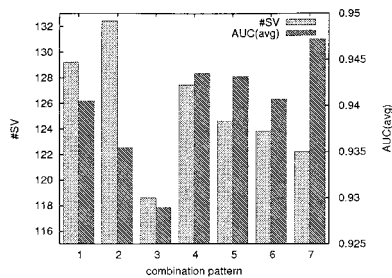


Fig. 7 vehicle, I-CV

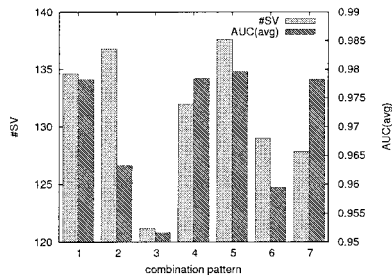


Fig. 8 glass, N-CV.

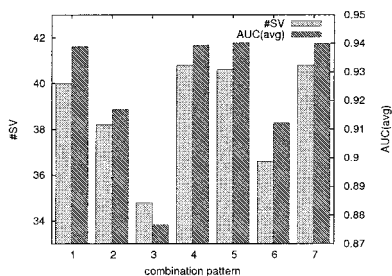


Fig. 9 glass, I-CV.

を行った pattern3 が相対的に悪くなっていることから、#SV を少なくして汎化能力を高くするだけでは識別器の精度を上げることができないといえる。

次に、目的関数を二つ組み合わせた pattern4~6 の結果を見ると、vehicle データでは pattern4 が、glass データでは pattern6 が不安定な結果となっている。

まず、vehicle データにおける pattern4 に注目すると、悪い結果となっているのは N-CV 法の場合のみである。N-CV 法での #SV が他の複数目的を組み合わせた場合に比べて多くなっていることから、vehicle データにおいて AUC(min) となるクラス組み合わせのデータ性質が他のクラス組み合わせのものとは異なっているため、AUC(avg) と AUC(min) の両方を良くしようとする、識別平面を複雑にせざるをえなかったのではないと思われる。そして、I-CV 法での結果を見ると、pattern4 は相対的に悪くない結果が出ている。これは学習用データ数を絶対的に少なくすることで、必然的に滑らかな識別平面が構成されたためだと思われる。

次に、glass データにおける pattern6 の結果に注目すると、N-CV 法と I-CV 法のどちらにおいても相対的に悪い結果となっていることがわかる。これは、#SV が他の結果に比べて少ないことから、AUC(min) と #SV を最適化することにより、AUC(min) の識別に特化してしまったことが予測される。glass データは 4.1 節で計算した F-Score が全体的に高いことより、各クラスは分類しやすいことがわかっている。このため AUC(min) の識別平面を構成する際に #SV を少なくしても AUC(min) の値は保てたのだろうが、他のクラス識別に関する能力を失ってしまったものと思われる。

pattern4~6 において二つのデータでの結果を比較すると、難易度の高い vehicle データは pattern5,6 が良い結果となり、より容易な glass データの場合は pattern4,5 が良い結果となった。

まず、vehicle データで結果が良かった pattern5, 6 の共通点は #SV である。#SV と同時に AUC 値の最大化を目的関数とすることにより、非常に良い結果を出している。これは #SV と AUC を同時に最適化することにより、汎化能力と識別精度の両方を意識した識別器が構成されたためだと思われる。

次に、glass データでの結果が良好だったのは pat-

tern4, 5で,これに共通するのはAUC(avg)であり, vehicle データと同様の傾向はみられない.これは glass データの要素数が少なく, #SV を少なくすることが精度に大きく影響を与えてしまったためだと考えられる.

最後に, Fig. 6~Fig. 9を通して見てみると, pattern7が最も安定して良い結果を出していることがわかる. このことから,今回取り上げた各目的関数を最適化することは有効であり,かつ3つ全てを目的関数に設定することで,識別精度と汎用性を保った識別器を構成できることがわかった.

5 結論

本研究では, weighted-RBF カーネルを用いた多クラス SVM に対して, 多目的最適化手法を用いてカーネルパラメータの最適化を行った. 最適化を行う際の目的関数は #SV, AUC(avg), AUC(min) の3種類を採用し, Table 4 に示す組み合わせパターンで実験を行った.

全結果において, pattern7 が最も安定して良い結果を出すことができた. 他の組み合わせでは, AUC(min) は入力データの特性に左右され, #SV は入力データの要素数によって結果にばらつきが出てしまうことがわかった.

参考文献

- 1) N.Cristianini and J. Shawe-Taylor. サポートベクターマシン入門. 共立出版. 2005. 大北剛訳.
- 2) 栗田 多喜夫. サポートベクターマシン入門. 産業技術総合研究所, 脳神経情報研究部門. <http://www.neurosci.aist.go.jp/~kurita/lecture/svm/svm.html>.
- 3) Chin-Wei Hsu Chin-Jen Lin. A Comparison of Methods for Multi-class Support Vector Machines. Department of Computer Science and Information Engineering National Taiwan University. Taipei 106, Taiwan.
- 4) Sergio A.Rojas and Delmiro Fernandez-Reyes. Adapting Multiple Kernel Parameters for Support Vector Machines using Genetic Algorithms. Division of Parasitology National Institute for Medical Research and Department

of Computer Science University College London.

- 5) Y.-W. Chen and C.-J.Lin. Feature extraction foundations and applications. Springer. 2006. Section:[Combining SVMs with various feature selection strategies].
- 6) UCI Machine Learning Repository. <http://archive.ics.uci.edu/ml/index.html>.
- 7) 多目的遺伝的アルゴリズム. http://mikilab.doshisha.ac.jp/dia/research/mop_ga/.
- 8) Tome Fawcett. ROC Graphs:Notes and Practical Considerations for Researchers. HP Laboratories. March 16,2004.