1-B-13

日本オペレーションズ・リサーチ学会
2004年秋季研究発表会

One-Class Classification とその応用

02203360 筑波大学　* 鈴木 伸彦　SUZUKI Nobuhiko
01105930 筑波大学　香田 正人　KODA Masato

1 Introduction

パターン認識の分野のほとんどが（多クラス）分類問題や回帰問題を扱っているが，その他にも分類問題を拡張した 1 クラス分類問題（One-Class Classification）がある．1 クラス分類問題は，多クラス分類問題のようにオブジェクトのクラスを決定するのではなく，オブジェクト集合の記述を行う．この記述は訓練集合の入力パターンを覆うもので，理想的には入力空間上で他のパターン（例えば，はずれ値など）に対しては棄却するようなものが望ましい．この 1 クラス分類問題は，はずれ値探索やノベルティー探索などに使われている．

現在では，1 クラス分類問題の手法として SVM を拡張した One-Class Support Vector Machine と，球体を用いて記述を行う Support Vector Domain Description が知られている．両手法ともに，入力データを高次の特徴空間に非線形写像し，特徴空間上で判別関数を求める．さらに，判別関数の双対表記は内積で表せるため，特徴空間上でカーネルを利用する事が出来る．本研究では，両手法に対して，最適なカーネルや最適なパラメータについて考察を行う．

2 Notation

はじめに，両手法で共通して使用している記号と変数について以下にまとめる．

$\mathbf{x}_i$	訓練パターン　$(i=1,\cdots,n)$
$\xi_i(\geq 0)$	スラック変数
$\alpha_i(\geq 0)$	Lagrange 乗数
$\nu \in (0,1]$	ν-SVM と同様のパラメータ
$\langle\cdot,\cdot\rangle$	内積
$k(\cdot,\cdot)$	カーネル関数
$f(\cdot)$	判別関数
$\mathbf{x}$	未知の入力パターン

3 Support Vector Domain Description

1 クラス分類問題の手法として，まず，Support Vector Domain Description（SVDD）について説明する．SVDD では，入力データの全て（またはほとんど）を含むような球体のうち，その体積が最小となるものを求める [3].

球体の中心を c，半径を R とすると，最適な球体を求めるためには以下の 2 次計画問題を解けばよい．

$$\begin{array}{ll} \min\limits_{R,\xi,c} & R^2 + \dfrac{1}{\nu n}\sum\limits_{i=1}^{n}\xi_i \\ \text{s.t.} & \|\mathbf{x}_i - c\|^2 \leq R^2 + \xi_i \\ & \xi_i \geq 0 \qquad (i=1,2,\cdots,n) \end{array} \tag{1}$$

最適化問題 (1) の解 R，c に対して，判別関数は以下のようになる．

$$f(\mathbf{x}) = \mathrm{sgn}(R^2 - \|\mathbf{x}_i - c\|^2) \tag{2}$$

最適化問題 (1) の双対問題は，以下のように表される．

$$\min_{\alpha} \quad \textstyle\sum_{i,j}\alpha_i\alpha_j\langle\mathbf{x}_i,\mathbf{x}_j\rangle - \sum_i \alpha_i\langle\mathbf{x}_i,\mathbf{x}_i\rangle \tag{3}$$

$$s.t. \quad 0 \leq \alpha_i \leq \frac{1}{\nu n}, \quad \sum_i \alpha_i = 1 \tag{4}$$

一般に，データは球状には分布していないので，上述の問題を解いて得られた球体が入力データを正確に記述しているとは期待できない．そこで，入力データを高次の特徴空間に写像し，特徴空間上でのデータの記述を考える．最適な特徴空間が選択されれば，より正確なデータの記述が可能になると期待できる．双対問題 (4) の目的関数は，入力空間上での内積で表されているため，それらを特徴空間上での内積に置き換えることで，以下のような最適化問題が得られる．

$$\min_{\alpha} \quad \sum_{i,j}\alpha_i\alpha_j k(\mathbf{x}_i,\mathbf{x}_j) - \sum_i \alpha_i k(\mathbf{x}_i,\mathbf{x}_i) \tag{5}$$

$$s.t. \quad 0 \leq \alpha_i \leq \frac{1}{\nu n}, \quad \sum_i \alpha_i = 1 \tag{6}$$

このとき，判別関数 (2) は

$$f(\mathbf{x}) = \mathrm{sgn}\Bigg(R^2 - \sum_{ij}\alpha_i\alpha_j k(\mathbf{x}_i,\mathbf{x}_j) + 2\sum_i \alpha_i k(\mathbf{x}_i,\mathbf{x}) - k(\mathbf{x},\mathbf{x})\Bigg) \tag{7}$$

で表される．

対応する Lagrange 乗数 α_i が非負であるようなパターンをサポートベクター (support vector) と呼び，球体を記述するにはこれらだけが必要になる．また，特徴空間上で $\|\Phi(\mathbf{x}_i) - c\|^2 = R^2$ を満たすパターンは球体の表面上の点になり，これらのパターンに対しては，$0 < \alpha_i < 1/(\nu n)$ にな

る．球体の半径 R はこのようなパターンから，球体の中心 $c = \sum_i \alpha_i \Phi(\mathbf{x}_i)$ までの距離を計算することで求められる．さらに，$\alpha_i = 1/(\nu n)$ を満たすパターンは球体の外側にあり，これらのサポートベクターがはずれ点であると考えられる．

4 One-Class SVM

続いて，1クラス分類問題の手法として One-Class Support Vector Machine（OCSVM）について説明する．OCSVM では，入力パターンを適切なカーネル（Gaussian カーネルなど）に対応する非線形写像 Φ を用いて高次の特徴空間に写像し，それらを特徴空間上で原点から最大のマージンになるような超平面で分離する [2]．未知の入力パターン $\mathbf{x}$ に対しては，$\mathbf{x}$ が特徴空間上で超平面のどちら側にあるかによって $f(\mathbf{x})$ の値が決まる．

特徴空間上でデータを原点から分離するためには，以下の2次計画問題を解けばよい．

$$\begin{array}{lll} \min\limits_{\mathbf{w},\xi,\rho} & \frac{1}{2}\|\mathbf{w}\|^2 - \nu n \rho + \sum\limits_{i=1}^{n} \xi_i & \\ \text{s.t.} & \langle \mathbf{w}, \Phi(\mathbf{x}_i)\rangle \geq \rho - \xi_i & \\ & \xi_i \geq 0 \qquad (i = 1, 2, \cdots, n) & \end{array} \tag{8}$$

このとき，判別関数

$$f(\mathbf{x}) = \mathrm{sgn}(\langle \mathbf{w}, \Phi(\mathbf{x})\rangle - \rho) \tag{9}$$

は，SV 正則化項 $\|\mathbf{w}\|$ を小さく保ちながら，訓練集合に含まれるほとんどのパターン $\mathbf{x}_i$ に対して正の値を返す．これら二つの相反する目標は ν によってコントロールされる．

最適化問題 (8) の双対問題は，以下のように表される．

$$\min_{\alpha} \quad \tfrac{1}{2}\textstyle\sum_{ij} \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) \tag{10}$$

$$s.t. \quad 0 \leq \alpha_i \leq \frac{1}{\nu n}, \quad \sum_i \alpha_i = 1 \tag{11}$$

このとき，判別関数は

$$f(\mathbf{x}) = \mathrm{sgn}\left(\sum_i \alpha_i k(\mathbf{x}_i, \mathbf{x}) - \rho\right) \tag{12}$$

で表される．ここで，対応する Lagrange 乗数 α_i が非負であるようなパターンをサポートベクターと呼び，判別関数はこれらを用いて表される．また，$0 < \alpha_i < 1/(\nu n)$ を満たす点は判別超平面上にある点であり，このような α_i を利用することで，次式のように ρ を求めることができる．

$$\rho = \langle \mathbf{w}, \Phi(\mathbf{x}_i)\rangle = \sum_j \alpha_j k(\mathbf{x}_j, \mathbf{x}_i). \tag{13}$$

さらに，$\alpha_i = 1/(\nu n)$ を満たすパターンは判別超平面よりも原点側にあり，これらのサポートベクターがはずれ点であると考えられる．

5 SVDD and OCSVM

本節では，SVDD と OCSVM との間の関係性について説明する．

OCSVM において，次の ν-property が成り立つことが知られている [1]．

Theorem *ν-property* 最適化問題 (12) の解が $\rho \neq 0$ を満たすとき，以下の3つの性質が成り立つ．
(i) ν ははずれ点の割合の上限である．
(ii) ν はサポートベクターの割合の下限である．
(iii) カーネルが解析的で定数でなく，サンプルデータ $\mathcal{X}$ がある連続な確率分布 $P(\mathbf{x})$ から独立に生成されたとすると，n が無限大に近づくとき，はずれ点の割合とサポートベクターの割合は，ν に確率収束する．

ν-property は SVDD においても成立する．これより両手法においてパラメータ ν の値が，はずれ点の割合を表していることがわかる．

また，SVDD において Gaussian カーネルを用いた場合，式 (11) は以下のようになる．

$$\min_{\alpha} \quad \sum_{i,j} \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) - 1 \tag{14}$$

これより，Gaussian カーネルを用いた SVDD と OCSVM は同値であることがわかる．

6 Experiment

これまでに述べた SVDD，OCSVM 両手法に対して，人工的データと実データを用いて実験を行う．その際，カーネルやそのパラメーターを色々と変えることで，両手法における最適なカーネルやパラメーターについて考察する．実験結果は当日発表する．

References

[1] B.Schölkopf and A.J. Smola. *Learning with Kernels.* MIT Press, Cambridge, MA, 2002.

[2] B.Schölkopf, J.C. Platt, J. Shawe-Taylor, A.J. Smola and R.C. Williamson. *Estimating the Support of a High Dimensional Distribution.* Nural Computation,13(7):1443-1471, 2001.

[3] D.M.J. Tax and R.P.W. Duin. *Support Vector Domain Description.* Pattern Recognition Letters, 20:1991-1999, 1999.