Pattern Recognition Letters 71 (2016) 73–77

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Robust solutions to fuzzy one-class support vector machine☆

Yong Liu[a,b,*], Biling Zhang[a,b], Bin Chen[a], Yandong Yang[a]

[a] School of Network Education, Beijing University of Posts and Telecommunications, Beijing 100088, China
[b] The State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China

ARTICLE INFO

Article history:
Received 18 April 2015
Available online 6 January 2016

Keywords:
One-class SVM
Fuzzy system
Robustness

ABSTRACT

One-class SVM is used for classification which distinguishes one class of data from the rest in the feature space. For the training samples coming from different sources with different quality, in this letter, a reformulation of one-class SVM is proposed by simultaneously incorporating robustness and fuzziness to improve the classification performance. Based on the proposed model, we derive the relationship between the lower bound of fuzziness $\mu_{\min}$ and the upper bound of perturbation η in the input data. Specifically, for a given η, only when the assigned fuzziness to the input data is larger than $\mu_{\min}$, could the input data be in full use and differentiated effectively. The experiments verify the mathematical analysis and illustrate that the proposed model can achieve better classification performance.

1. Introduction

To solve one-class classification problems in many applications (David and Duin [4]), one-class SVM (Schlkopf et al. [1]) is proposed by extracting a hyperplane in a kernel feature space such that a given fraction of training objects may reside beyond the hyperplane, at the same time the hyperplane has maximal distance to the origin. It has been shown that one-class SVM provides better performance than traditional learning machines in distinguishing between a set of target objects and all other possible objects.

In the actual applications, noises will give rise to uncertain data. In such a case, how to develop a classification model with resistance to data perturbations is a critical issue. In the literature, the robustness problem has been studied by reformulating the classification model with the bounded perturbation in the input data to make the model robust to the uncertainty. In Trafalis and Alwazzi [11], considering the bounded errors in the input data, the authors investigated the stability of the linear programming SVM (LP-SVM) solution and concluded that the SVM model was stable under that case. Trafalis and Gilbert [10] proposed a new robust programming formulations for the bounded perturbation and discussed the relationship of the bounded perturbation and generalized margin. In Huang et al. [5], a robust support vector regression (RSVR) method with uncertain input and output data is studied, in which, linear formulations robust to ellipsoidal uncertainty are also considered from a geometric perspective. However, in most of existing robust models, the input training points are assumed to have the same quality and the same process is applied to all of them, which is not desirable when input points with different quality must be dealt discriminately.

In order to differentially treat the training data with different quality in the training process, the concept of fuzzy set theory is incorporated into SVM (Lin and Wang [3]), in which, a fuzzy membership is assigned to each input point of SVM. The fuzzy membership here can be regarded as the attitude of the corresponding training point toward one class in the classification problem. Specifically, one-class SVM combined with the concept of fuzziness in Hao [9], assigning each data point a membership value according to its relative importance in the class. In Filippone et al. [7], a fuzzy kernel clustering method was applied to one-class SVM. However, in these models, robustness is not involved, and the relationship between fuzziness and bounded perturbation is not well investigated.

Considering that little research has been done to incorporate both robustness and fuzziness simultaneously and discuss the relationship between them, in this letter, a model introducing fuzziness and robustness to standard one-class SVM is proposed. Based on the proposed model, the relationship between the fuzzy membership and the bounded perturbation is analyzed, and the lower bound of fuzzy membership given the bounded perturbation is derived.

The rest of this letter is organized as follows. A brief review of the theory of one-class SVM is described in Section 2. In Section 3, the one-class SVM model with fuzziness and robustness is reformulated, and the relationship between the fuzzy membership and

☆ This paper has been recommended for acceptance by A. Petrosino.

* Corresponding author at: School of Network Education, Beijing University of Posts and Telecommunications, Beijing 100088, China. Tel.: +86 13651390601.
E-mail address: liuyongamtf@aliyun.com (Y. Liu).

http://dx.doi.org/10.1016/j.patrec.2015.12.014

the perturbation bound is analyzed. Experiments are presented in Section 4, and some concluding remarks are given in Section 5.

2. Review to standard one-class SVM

Suppose that the training data is

$$x_1, x_2, \ldots, x_N \in R^n, \tag{1}$$

where N is the number of observations and R^n is the input data set. Then a feature map Φ: $R^n \to F$ is defined as

$$\langle \phi(x_i) \cdot \phi(x_j) \rangle = k(x_i, x_j), \tag{2}$$

where $\langle \cdot \rangle$ denotes the inner product, and $k(x_i, x_j)$ is the kernel of x_i and x_j (Schlkopf et al. [1]; Mller et al. [6]). For example, a Gaussian kernel can be defined as

$$k(x_i, x_j) = e^{-q\|x_i - x_j\|^2}. \tag{3}$$

After being mapped into the feature space corresponding to the kernel, the training data will be treated as belonging to one class, while the origin will be treated as the only member of the second class. The aim of one-class SVM is to separate the training data from the origin with the maximum margin defined by the pair (w^*, b^*) which is the solution of the following quadratic program

$$\min_{w \in F, \xi_i \in R^n, b \in R} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} \xi_i + b, \tag{4}$$

$$s.t. \quad \begin{aligned} &\langle w \cdot \phi(x_i) \rangle + b \geq 0 - \xi_i, \\ &\xi_i \geq 0, i = 1, \ldots, N, \end{aligned} \tag{5}$$

where $\xi_i, i = 1, \ldots, N$, are nonzero slack variables which are the measure of error in the one-class SVM and C is a predefined regularization parameter which is used to control the fraction of outliers (Mller et al. [6]).

With w^* and b^*, a function f is defined as follows which takes the value $+1$ in a small region capturing most of the data points and elsewhere

$$f(x) = sgn((w^* \cdot \phi(x)) + b^*). \tag{6}$$

3. Robust one-class SVM with fuzziness

3.1. The proposed model

In this sub-section, we extend the standard one-class SVM by introducing fuzzy and robust parameters to a robust and fuzzy one-class SVM, which is called RF-SVM in the rest of this paper.

In our proposed model, linear classification is considered for simplification. Suppose that the training points $\{x_i, i = 1, \ldots, N\}$ are subject to perturbation on each of their coordinates. In such a case, each training vector x_i is not longer a point belonging to the R^n, but a training point set χ_i which could be regarded as a sphere with x_i as center and $r_i (r_i \in R)$ as radius.

Then, a fuzzy membership $\mu_i (0 \leq \mu_i \leq 1)$ is assigned to each training point set $\chi_i, i = 1, \ldots, N$. Thus the training data could be extended as

$$(\chi_1, \mu_1), (\chi_2, \mu_2), \ldots, (\chi_N, \mu_N), \tag{7}$$

and the first set of constrains in (5) can be rewritten as

$$\langle w \cdot (x_i + r_i u_i) \rangle + b \geq 0 - \xi_i, \forall \|u_i\| \leq 1, i = 1, \ldots, N. \tag{8}$$

where $u_i \in R^n$, and $r_i u_i$ represents the perturbation vector.

Assuming that there is not a priori information about the perturbation. Instead, a bound of the norm of each perturbation vector $r_i u_i (i = 1, \ldots, N)$ with respect to the L_p norm is known. In this letter, $p = 2$ is considered, and thus $\|r_i\| \leq \eta_i$, where η_i represents perturbation bound for training point set $\chi_i, i = 1, \ldots, N$ (Trafalis and Gilbert [10]).

To find a feasible solution of (8) for every realization of the bounded perturbation η_i, which is characterized as the robust feasible solution, w should satisfy the following condition for every $i = 1, \ldots, N$ (Trafalis and Alwazzi [11])

$$\min_{\|r_i\| \leq \eta_i} \langle w \cdot (x_i + r_i u_i) \rangle + b + \xi_i \geq 0, \forall \|u_i\| \leq 1, i = 1, \ldots, N. \tag{9}$$

Further, (9) could be reformulated into the following problem as

$$\begin{aligned} &\min\nolimits_{r_i, u_i} \langle w \cdot r_i u_i \rangle, \\ &s.t. \|r_i\| \leq \eta_i, \|u_i\| \leq 1. \end{aligned} \tag{10}$$

According to Schwarz inequality we have

$$|\langle w \cdot r_i u_i \rangle| \leq \|w\| \cdot \|r_i u_i\| \leq \|w\| \cdot \|r_i\| \leq \eta_i \|w\|.$$

To consider the worst case, we replace $\langle w \cdot r_i u_i \rangle$ by its minimum value $-\eta_i \|w\|$. Then we have the linear classification problem with robustness and fuzziness as follows

$$\min_{w \in F, \xi_i \in R^n, b \in R} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} \mu_i \xi_i + b, \tag{11}$$

$$s.t. \langle w \cdot x_i \rangle + b - \eta_i \|w\| \geq 0 - \xi_i \ \forall \xi_i \geq 0, i = 1, \ldots, N. \tag{12}$$

The above optimization problem could be solved after being converted to a Second Order Cone Programming (SOCP). When primal-dual interior point methods are adopted, the computation complexity of SOCP is $O(\sqrt{\nu} ln \frac{1}{\epsilon})$ with objective function within ϵ of the optimal value, and ν being a parameter of the cone (Nesterov and Todd [13]).

Next, when consider the case that one-class SVM is implemented generally in a kernel feature space, in which $\phi(\chi_i) = \phi(x_i) + R_i$, and $\|R_i\| \leq \eta_i, i = 1, \ldots, N$, the problem in (11) and (12) can be extended to a nonlinear one-class classification as

$$\min_{w \in F, \xi_i \in R^n, b \in R} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} \mu_i \xi_i + b, \tag{13}$$

$$s.t. \quad \begin{aligned} &\langle w \cdot \phi(x_i) \rangle + b - \eta_i \|w\| \geq 0 - \xi_i, \\ &\xi_i \geq 0, i = 1, \ldots, N. \end{aligned} \tag{14}$$

Based on the objective function and the constraints described in (13) and (14), we will derive the relationship between the fuzzy membership and perturbation bound in the next sub-section.

3.2. Analysis of the relationship between fuzzy membership and perturbation bound

If we use a Gaussian kernel in (3), any data set is separable after it is mapped into the feature space according to Mller et al. [6]. A theorem is then derived based on the data set with perturbation $\{\chi_1, \chi_2, \ldots, \chi_N\}$ as follows.

Theorem 1. *If the data set is separable, given the perturbation bound η, the fuzzy membership μ has a lower bound*

$$\mu_{min} = \frac{-b}{N \cdot C \cdot (X_{\phi(a)} + \eta)^2}. \tag{15}$$

where N is the number of the training points, C is an predefined regularization parameter, b represents the intercept of the optimal superplane, $X_{\phi(a)} = max\{\|\phi(x_i)\|, i = 1, \ldots, N\}$ and $\phi(x_i)$ is a feature map of x_i.

Proof. For the convenience of discussion, the linear program in (11) and (12) is used to analyze the relationship between the fuzzy membership and perturbation bound. However, the conclusion could be extended to the feature space formulated in (13) and (14).