Improving iterations for SVDD-based pattern denoising

Yong Tian [a], Xiaoming Wang [b] and Xiaohuan Yang [c]

School of Computer and Software Engineering, Xihua University, Chengdu 610039, China

[a]xhuty1990@163.com, [b]mistertian1990@163.com, [c]misswu0814@163.com

Abstract

Support vector data description is one of the support vector learning methods, in which one tries to find a ball to distinguish the normal data points from all other possible abnormal or outlier data points. The ball should be as small as possible and include as more normal data points as possible. On the basis of the principle, it can be applied for pattern denoising and achieves a promising performance by comparing it with other pattern denoising methods such as PCA, KPCA and wavelet denoising. Since each denoising will be approaching the exact denoised point, it spends much time on iteration. Therefore, in the article we propose an improving algorithm that can directly accomplish the denoising and avoid multiple iterations besides saving the cost of time.

Keywords

Support vector data description; pattern denoising; multiple iterations; improving algorithm.

1. Introduction

Kernel methods[1-2] is widely used and learned as available tools in the area of machine learning over past decade. Support vector data description [3] (SVDD) is one of the support vector learning methods for descripting the normal data point. The main idea is to build up the SVDD ball that should be as small as possible and include as much data normal points as possible. The points inside or on the SVDD ball are regarded as the normal and the rest are abnormal.

Based on the character of SVDD, it can be applied for pattern denoising. The purpose of SVDD-based pattern denoising is that mapping the original data to high-dimensional feature space by using kernel function to find out the SVDD ball and move the noise point along with the surface of Gaussian ball to the cross point and finish the pattern denoising[4]. Although the mapping from original space to feature space does not exist, the inverse mapping cannot be obtained. Therefore, finding an approximate point instead seems necessary [5]. Since the position of each point in data space is determined by its several neighborhood points, we look for a few points near the noise and use the multi-dimensional analysis[6](MDS) to find the denoised point in original space.

However, the result of each pattern denosing of SVDD is closer to the final point and it need many times of iteration. Every iteration spends much time on loop and matrix calculations. Each pattern denoising is moving the noise point forward some direction and the last time is on the SVDD ball becoming the normal one. So, we propose an improving iterations algorithm for SVDD-based pattern denoising.

The remaining parts of the paper are organized as follows. In section 2, the idea of SVDD will be claimed. The problem existing in SVDD-based pattern denoising will be expressed in section 3. The new method will be discussed in section 4. The experimental results of the proposed method on a toy dataset and several real-world datasets will be presented in section 5. Finally, conclusions are drawn in last section.

2. The idea of SVDD

In the paper, we assume a dataset consisting of N data points, denoted as $\mathbf{X}=\{\mathbf{x}_i \mid \mathbf{x}_i \in \mathrm{R}^d, i=1,\cdots,N\}$, where d is the dimension of the data space. In this section, we briefly review the SVDD and its primal optimization problem is denoted as

$$\begin{aligned} &\min \quad R^2 + C\sum_{i=1}^{N}\xi_i \\ &s.t. \quad \|\phi(\mathbf{x}_i) - \mathbf{a}\|^2 \le R^2 + \xi_i, \xi_i \ge 0, i = 1, \cdots, N \end{aligned} \tag{1}$$

Here, $R \in \mathrm{R}$ is the radius of the SVDD ball in feature space, $\mathbf{a} \in \mathrm{R}^d$ is the vector of the center of SVDD ball, C is predefined and controls the fraction of outliers in $\mathbf{X}$ and ξ is the slack variable. In the case of Gaussian kernel, its dual optimization problem can be presented as

$$\begin{aligned} &\min \quad \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j\kappa(\mathbf{x}_i, \mathbf{x}_j) \\ &s.t. \quad \sum_{i=1}^{N}\alpha_i = 1, \alpha_i \in [0, C], i = 1, \cdots, N \end{aligned} \tag{2}$$

If $\boldsymbol{\alpha}^* = [\alpha_1^*, \cdots, \alpha_N^*]^T$ is the solution to (1), the $\mathbf{a}$ can be formulated as

$$\mathbf{a} = \sum_{i=1}^{N}\alpha_i\phi(\mathbf{x}_i) \tag{3}$$

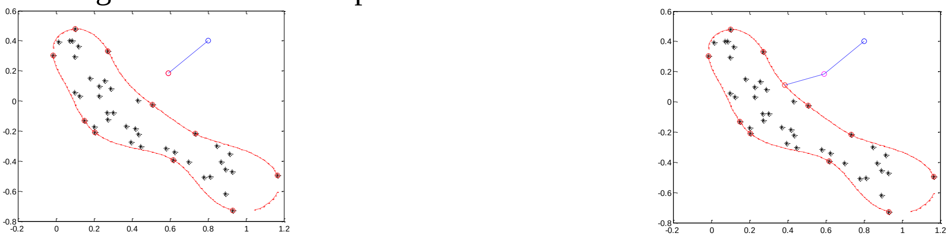
The formulation of R also can be acquired

$$R^2 = \kappa(\mathbf{x}, \mathbf{x}) - 2\sum_{i=1}^{N}\alpha_i\kappa(\mathbf{x}, \mathbf{x}_i) + \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j\kappa(\mathbf{x}_i, \mathbf{x}_j) \tag{4}$$

3. The problem of SVDD-based pattern denoising

In this section, we will discuss the problem existing in SVDD-based pattern denoising. And aiming at the shortcoming, we propose a new method to improve the iterations in pattern denoising with SVDD.

Based on above idea, the mapping from original space to feature space is unknown and the calculation in feature space refers to the product of two points. So, kernel method is the best choice in solving the high-dimensional problem. The results of experiments demonstrate that it need many iterations in obtaining the final denoised point.



(a) (b)

Fig. 1 The photograph of pattern denoising. (a)One time denoising. (b) 100 times denoising.

According to Figure1, it's not hard to see that one time denoising cannot acquire the final denoised point, but 100 times denoising. Based on the result, we propose a new method to improve the iterations in next section.

4. The improving iteration method for SVDD-based pattern denoising

In the basis of shortcoming of SVDD-based pattern denoising, we proposed the improving iteration method. Its main idea is that each denoising toward to SVDD ball is closer to the final point. Connecting the noise point with the first denoising point and expanding the line to intersect with the SVDD ball, the intersection is the denoised point in feature space and embed the idea of MDS back to the original space.

Suppose the noise point $\mathbf{x} \in R^d$, the first denoising point $\hat{\mathbf{x}} \in R^d$, and the final denoising point $\mathbf{z} \in R^d$ and with respect to the idea of proposed method, we can get the following formula.

$$\mathbf{z} = \mathbf{x} + \lambda(\hat{\mathbf{x}} - \mathbf{x}), \lambda > 0 \tag{5}$$

So, the $\mathbf{x}$ and $\hat{\mathbf{x}}$ are easy to get and $\lambda \in R$ is hard to discover. If we want to work out the final denoised point $\mathbf{z}$, perhaps we solve the value of λ or a certain range of values.

Gaussian kernel is used in the article to carry out the algorithm. The final denoised point must be the normal and if mapping it from original space to feature space becoming the $\phi(\mathbf{z})$, it should be in or on the SVDD ball. Hence, there is a formula to express the idea.

$$\|\phi(\mathbf{z})-\mathbf{a}\|^2 \le R^2 \tag{6}$$

Due to the Gaussian kernel, $\|\phi(\mathbf{z})\|^2=1$ is set up and $\mathbf{a}$, R can be obtained in formula (3), (4). The formula (6) can be presented in following form.

$$\frac{1+\mathbf{a}_F^2-R^2}{2} \le \sum_{i=1}^{N}\alpha_i \exp\left(-\frac{\|\mathbf{z}-\mathbf{x}_i\|^2}{2\sigma^2}\right) \tag{7}$$

Assuming $M=\frac{1+\mathbf{a}_F^2-R^2}{2}$, take it into formula (7) and combining with formula (5) to rewrite it.

$$M \le \sum_{i=1}^{N}\alpha_i \exp\left(-\frac{\|\lambda(\hat{\mathbf{x}}-\mathbf{x})+(\mathbf{x}-\mathbf{x}_i)\|^2}{2\sigma^2}\right) \tag{8}$$

So, the next work is to find out the value of λ or a certain range of values.

According to the property of convex functions, we can get the following formula.

$$\sum_{i=1}^{N}\alpha_i \exp\left(-\frac{\|\lambda(\hat{\mathbf{x}}-\mathbf{x})+(\mathbf{x}-\mathbf{x}_i)\|^2}{2\sigma^2}\right) \ge \exp\left(-\frac{\sum_{i=1}^{N}\alpha_i\|\lambda(\hat{\mathbf{x}}-\mathbf{x})+(\mathbf{x}-\mathbf{x}_i)\|^2}{2\sigma^2}\right) \tag{9}$$

If the λ does exist, there have

$$\exp\left(-\frac{\sum_{i=1}^{N}\alpha_i\|\lambda(\hat{\mathbf{x}}-\mathbf{x})+(\mathbf{x}-\mathbf{x}_i)\|^2}{2\sigma^2}\right) \ge M \tag{10}$$

established and the formula

$$\sum_{i=1}^{N}\alpha_i \exp\left(-\frac{\|\lambda(\hat{\mathbf{x}}-\mathbf{x})+(\mathbf{x}-\mathbf{x}_i)\|^2}{2\sigma^2}\right) \ge M \tag{11}$$

must be established. In another word, we cannot get the value of λ directly, but there we find an intermediate quantity to work out indirectly and satisfy with the conditions above. Comparing with formula (11), formula (10) seems easier to figure out.

Due to $\sum_{i=1}^{n}\alpha_i=1$, formula (11) can be transformed another form

$$\lambda^2\|\hat{\mathbf{x}}-\mathbf{x}\|^2+2\lambda(\hat{\mathbf{x}}-\mathbf{x})\cdot(\mathbf{x}-\sum_{i=1}^{n}\alpha_i\mathbf{x}_i)+\sum_{i=1}^{n}\alpha_i\|\mathbf{x}-\mathbf{x}_i\|^2+2\sigma^2\ln(M)\le 0 \tag{12}$$

Suppose $a=\|\hat{\mathbf{x}}-\mathbf{x}\|^2$, $b=2(\hat{\mathbf{x}}-\mathbf{x})\cdot(\mathbf{x}-\sum_{i=1}^{n}\alpha_i\mathbf{x}_i)$, $c=\sum_{i=1}^{n}\alpha_i\|\mathbf{x}-\mathbf{x}_i\|^2+2\sigma^2\ln(M)$, formula (12) can be presented as

$$a\lambda^2+b\lambda+c\le 0 \tag{13}$$

So, the problem about solving λ transforms into a quadratic function for the root problem. Obviously, $a>0$ determines the parabola opening up. Assuming λ_1,λ_2 are the two roots for equation $a\lambda^2+b\lambda+c=0$, there should be $\lambda_1=(-b+\sqrt{b^2-4ac})/2a$, $\lambda_2=(-b-\sqrt{b^2-4ac})/2a$ .When existing $\lambda_2\le\lambda\le\lambda_1$, $a\lambda^2+b\lambda+c\le 0$ must be established. Generally, we can set $\lambda=-b/2a=(\hat{\mathbf{x}}-\mathbf{x})(\mathbf{x}-\sum_{i=1}^{n}\alpha_i\mathbf{x}_i)\Big/\|\hat{\mathbf{x}}-\mathbf{x}\|^2$. Taking the value of λ into formula (5), we can get the approximate final denoised data point.

5. Experiments

In this section, we carry out the experiments on toy datasets and several real-world datasets and compare the time performance of the proposed method with the former iterations.

5.1 Experiments on toy datasets.

Dataset 1: Firstly, we randomly generate some data to describe SVDD. Set the width of Gaussian kernel $\sigma_0=0.3$, $C=0.3$ and the noise $noise=[0.8,0.4]$.

(a) time: T_{former}=6.364841 (b) time: $T_{latter} \cong 0$

Figure2.The denoising photograph on toy dataset1. (a)The iterations SVDD. (b)The proposed method

Comparing Figure2 (a) with Figure2 (b), in the case of same dataset and parameters, we can see that when the effect of pattern denoising is extremely similar, the proposed method have a better time performance on toy dataset.

Dataset 2: Secondly, we change another toy dataset to carry out the experiment. Set the width of Gaussian kernel $\sigma_0 = 0.2$, $C = 0.1$ and the noise $noise = [1.8, -0.7]$.

(a) time: T_{former}=7.831250 (b)time: $T_{latter} \cong 0$

Figure3.The photograph on toy dataset2. (a)The iterations SVDD. (b)The proposed method.

When the dataset has changed, in the case of same dataset and parameters, according to the Figure3 (a), (b), it's easy to figure out that the proposed method is better on time performance.

5.2 Experiments on real-world datasets.

In this section, we use the real-word dataset called Japanese Vowels downloaded from the website: http://archive.ics.uci.edu/ml/datasets/Japanese+Vowels.Japanese Vowels consists of several training sets and test sets and each set includes 20 samples and each sample consists of 12 attributes. We randomly choose a few training samples for training to build up the model. Adding the noise to the testing samples is to carry out the denoising. Set the decisive function:

$$f(\mathbf{x}) = \|\phi(\mathbf{x}) - \mathbf{a}\|^2 - R^2 \tag{14}$$

Here, $\mathbf{x}$ is the final denoised point, $\mathbf{a}$ is the vector of the center of SVDD ball and R is the radius of SVDD ball.

There have the conditions of judgment:

(1)If existing $f(\mathbf{x}) \le 0$, the point $\mathbf{x}$ is in or on the curve in original space and $\phi(\mathbf{x})$ is in or on the SVDD ball in feature space. The point $\mathbf{x}$ can be regarded as the normal point;

(2)If existing $f(\mathbf{x}) > 0$, the point $\mathbf{x}$ is out of the curve in original space and $\phi(\mathbf{x})$ is out of SVDD ball in feature space. It can be regarded as the abnormal point.

In the factual experiments, when $f(\mathbf{x}) \le 1.0 \times 10^{-5}$ is established, the point is the normal point.

$f(\mathbf{x})$: The value of decisive function with the noised point.

$f_1(\mathbf{x})$: The value of decisive function with the denoised point in SVDD-based pattern denoising.

$f_2(\mathbf{x})$: The value of decisive function with the denoised point in the proposed method.

Table. 1 The value of decisive function in different cases

	$f(\mathbf{x})$	$f_1(\mathbf{x})$	$f_2(\mathbf{x})$
Data1	$f(\mathbf{x}_1) =$ 0.6513	$f_1(\mathbf{x}_1) =$ 7.5776×10^{-6}	$f_2(\mathbf{x}_1) =$ -0.0020
	abnormal	$t_1 = 1.045207\text{s}$	$t_2 = 0.171601\text{s}$

		normal	normal
Data2	$f(\mathbf{x}_2) = 0.5865$	$f_1(\mathbf{x}_2) = 9.0347 \times 10^{-6}$	$f_2(\mathbf{x}_1) = -0.0278$
	abnormal	$t_1 = 0.904806\text{s}$	$t_2 = 0.202801\text{s}$
		normal	normal
Data3	$f(\mathbf{x}_3) = 0.3586$	$f_1(\mathbf{x}_3) = 9.8415 \times 10^{-6}$	$f_2(\mathbf{x}_3) = -0.0127$
	abnormal	$t_1 = 0.904806\text{s}$	$t_2 = 0.140401\text{s}$
		normal	normal
Data4	$f(\mathbf{x}_4) = 0.2529$	$f_1(\mathbf{x}_4) = 9.5279 \times 10^{-6}$	$f_2(\mathbf{x}_4) = -0.0106$
	abnormal	$t_1 = 0.967206\text{s}$	$t_2 = 0.124801\text{s}$
		normal	normal
Data5	$f(\mathbf{x}_5) = 0.2745$	$f_1(\mathbf{x}_2) = 6.8502 \times 10^{-6}$	$f_2(\mathbf{x}_1) = -0.0245$
	abnormal	$t_1 = 0.826805\text{s}$	$t_2 = 0.187201\text{s}$
		normal	normal

According to Table1, we can get that the proposed method have a better time performance than the original SVDD-based pattern denoising on real-world datasets

6. Conclusion

As we know when carrying out the denoising with SVDD, it need many iterations and spend much time on loop and matrix calculations. So, aiming at this weakness in the algorithm, in the paper, we propose a novel method to improve the iterations in SVDD-based pattern denoising. Connecting the noise point with the first denoised point and expanding the line to the SVDD ball, the intersection is the final denoised point. The mathematical derivation is presented in the article. Finally, we calculate the formula of the final denoised point.

We carry out the experiments on toy datasets and real-world dataset. The results of experiments prove that the proposed method have a better time performance than the original SVDD-based pattern denoising. However, the parameter in the algorithm is not the best choice and we will pay attention to the optimization of parameters.

References

[1] Scholköpf B, Smola A: Learning with Kernels [M].Cambridge, MA: MITPress, 2002,p305-310.

[2] Muller, K.-R., Mika, S., Ratsch, G., Tsuda, K., & Scholkopf, B.: An introduction to kernel-based learning algorithms. IEEE Transactions on Neural Networks, 2001, 12(2),181–201.

[3] Tax, D., & Duin, R.: Support vector data description. Pattern Recognition Letters,1999, 20(11–13), 1191–1199.

[4] Burges, C. J. C. : Geometry and invariance in kernel based methods. In A. J. Smola, P. L. Bartlett, B. Schölkopf, & D. Schuurmons (Eds.), Advances in kernel methods—support vector learning. Cambridge, MA:MITPress,1999.

[5] Schölkopf, B., Mika, S., Burges, C., Knirsch, P., Muller, K.-R., Ratsch, G., & Smola, A. J.: Input space vs feature space in kernel-based methods.IEEE Transactionson Neural Networks, 1999, 10(5), 1000–1017.

[6] Cox, T. F., & Cox, M. A. A.: Multidimensional scaling (2nd ed.). London: Chapman & Hall,2001, p. 208-215.