Neurocomputing 205 (2016) 430–438

Contents lists available at ScienceDirect

Neurocomputing

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

K-nearest neighbor-based weighted multi-class twin support vector machine

Yitian Xu [1]

College of Science, China Agricultural University, Beijing 100083, China

ARTICLE INFO

Article history:
Received 1 November 2015
Received in revised form
12 March 2016
Accepted 12 April 2016
Communicated by Jiayu Zhou
Available online 12 May 2016

Keywords:
TSVM
K-nearest neighbor
Weights
Multi-class classification

ABSTRACT

Twin-KSVC, as a novel multi-class classification algorithm, aims at finding two nonparallel hyper-planes for the two focused classes of samples by solving a pair of smaller-sized quadratic programming problems (QPPs), which makes the learning speed faster than other multi-class classification algorithms. However, the local information of samples is ignored, and then each sample shares the same weight when constructing the separating hyper-planes. In fact, they have different influences on the separating hyper-planes. Inspired by the studies above, we propose a K-nearest neighbor (KNN)-based weighted multi-class twin support vector machine (KWMTSVM) in this paper. Weight matrix W is employed in the objective function to exploit the local information of intra-class. Meanwhile, both weight vectors f and h are introduced into the constraints to exploit the information of inter-class. When component $f_j = 0$ or $h_k = 0$, it implies that the j-th or k-th constraint is redundant. Removing these redundant constraints can effectively improve the computational speed of the classifier. Experimental results on eleven benchmark datasets and ABCD dataset demonstrate the validity of our proposed algorithm.

1. Introduction

The support vector machine (SVM) [1] is a popular machine learning technique and it has many advantages. First, SVM solves a QPP, assuring that once an optimal solution is obtained, it is the unique (global) solution. Second, SVM derives a sparse and robust solution by maximizing the margin between two classes. Third, SVM implements the structural risk minimization principle rather than the empirical risk minimization principle, which minimizes the upper bound of the generalization error.

One of main challenges for the classical SVM is the high computational complexity. The training cost of $O(l^3)$, where l is the total size of training data, is expensive. To improve the computational speed, Jayadeva et al. proposed a twin support vector machine (TSVM) classifier [2] for the binary classification in the spirit of proximal SVM [3–5]. TSVM determines two nonparallel hyper-planes by solving two smaller-sized and related SVM-type problems, where each hyper-plane is closer to one class and as far as possible from the other. The strategy of solving two smaller-sized QPPs rather than a single larger-sized one, makes the learning speed of TSVM approximately 4 times faster than that of the classical SVM. At present, TSVM has become one of the popular methods because of its low computational complexity. Many variants of TSVM have been proposed in recent years [6–15], and they enrich the related theory of TSVM.

Multi-class classification problem is often met in our real life. At present, we usually resolve it by two approaches: *one-versus-one*, and *one-versus-rest* [16–19]. It needs us to construct $k(k-1)/2$ possible binary classifiers and only two kinds of samples are involved for each classifier in the first approach, and no information is given for the rest samples, therefore we receive unfavorable outputs. The second approach easily leads to the class imbalance problem and produces a bad performance [20,21].

Although a multi-class problem can be transformed into a series of binary classification problems and some effective methods in two-class learning can be used, recent studies [22] have shown that some binary classification techniques are often not so useful when being applied to the multi-class problem directly. A new multi-class classification algorithm, called K-SVCR, was proposed in [23] for the k-class classification purpose, which produces better forecasting results as it evaluates all the training points into the 1-*vs*-1-*vs*-*rest* structure. Therefore, it has received much attention.

By integrating both the structural advantage of K-SVCR and the speed's advantage of TSVM, Xu [24] proposed a novel multi-class classification algorithm, called Twin-KSVC. It first constructs two nonparallel hyper-planes for the two focused classes of samples from k classes, and then maps the remaining samples into a region between the two nonparallel hyper-planes. Two nonparallel hyper-planes are obtained by solving two smaller-sized QPPs

E-mail address: xytshuxue@126.com
[1] Tel.: +86 10 62737077; fax: +86 10 62736563.

http://dx.doi.org/10.1016/j.neucom.2016.04.024
0925-2312/

rather than a single large one as in the K-SVCR, which implies that its computational speed is at least 4 times faster than that of the classical K-SVCR in theory. However, the local information is not exploited in Twin-KSVC, and samples share the same effects on the classification hyper-plane. In fact, they have different effects.

KNN method is an effective tool to exploit the local information in classification and regression problems [25–28]. But the present research on the classification problem is only limited to the binary classification problem. In binary classification problem, it not only can exploit the intra-class information but also the inter-class information, and its value embodies in the objective function W and constraint condition f. However, in the regression problem, it only embodies in the objective function W since there is no label for a point. Inspired by the studies above, we introduce the KNN method to the multi-class classification problem, and propose a KNN-based weighted multi-class twin support vector machine in this paper. Different weights W_s are given to samples from the same class, meanwhile functions f and h are introduced into the constraints to exploit the information of inter-class. When component $f_j=0$ or $h_k=0$, it implies that the j-th or k-th constraint is redundant. Removing these redundant constraints can effectively improve the computational speed of the classifier. Experimental results on eleven benchmark datasets and ABCD dataset demonstrate the validity of our proposed algorithm.

The paper is organized as follows. Section 2 outlines the TSVM and Twin-KSVC. A KNN-based weighted twin support vector machine for multi-class classification problem is proposed in Section 3, which includes both the linear and nonlinear cases. Algorithm analysis is shown in Section 4. Numerical experiments on eleven real-world benchmark datasets and ABCD dataset are conducted to investigate the effectiveness of our proposed algorithm in Section 5. The last section contains the conclusions.

2. Related works

In this section, we review the basics of TSVM and Twin-KSVC.

2.1. Twin support vector machine

TSVM generates two nonparallel hyper-planes instead of a single one as in the conventional SVMs. The two nonparallel hyper-planes are obtained by solving two smaller sized QPPs as opposed to a single large one in the standard SVMs. Consider a binary classification problem with l_1 samples belonging to class $+1$ and l_2 samples belonging to class -1 in the n-dimensional real space R^n. Let matrix $A\in R^{l_1\times n}$ represent the positive samples and matrix $B\in R^{l_2\times n}$ represent the negative samples. The linear TSVM seeks two nonparallel hyper-planes

$$x^Tw_1+b_1=0 \quad \text{and} \quad x^Tw_2+b_2=0 \tag{1}$$

such that each hyper-plane is closer to one class and as far as possible from the other. A new sample is assigned to class $+1$ or -1 depending upon its proximity to the two nonparallel hyper-planes.

TSVM is obtained by solving the following pair of QPPs:

$$\begin{aligned}\min_{w_1,b_1,\xi}\ & \frac{1}{2}\|Aw_1+e_1b_1\|^2+c_1e_2^T\xi\\ \text{s.t.}\ & -(Bw_1+e_2b_1)\ge e_2-\xi,\\ & \xi\ge 0e_2,\end{aligned} \tag{2}$$

and

$$\begin{aligned}\min_{w_2,b_2,\eta}\ & \frac{1}{2}\|Bw_2+e_2b_2\|^2+c_2e_1^T\eta\\ \text{s.t.}\ & (Aw_2+e_1b_2)\ge e_1-\eta,\\ & \eta\ge 0e_1.\end{aligned} \tag{3}$$

By introducing the Lagrangian multipliers α and β, we can derive their dual problems as follows:

$$\begin{aligned}\max_{\alpha}\ & e_2^T\alpha-\frac{1}{2}\alpha^TG(H^TH)^{-1}G^T\alpha\\ \text{s.t.}\ & 0e_2\le\alpha\le c_1e_2,\end{aligned} \tag{4}$$

and

$$\begin{aligned}\max_{\beta}\ & e_1^T\beta-\frac{1}{2}\beta^TH(G^TG)^{-1}H^T\beta\\ \text{s.t.}\ & 0e_1\le\beta\le c_2e_1.\end{aligned} \tag{5}$$

Where $H=[Ae_1]$ and $G=[Be_2]$. Once QPPs (4) and (5) are solved, we can get vectors

$$\begin{bmatrix} w_1\\ b_1\end{bmatrix}=-(H^TH)^{-1}G^T\alpha, \tag{6}$$

and

$$\begin{bmatrix} w_2\\ b_2\end{bmatrix}=(G^TG)^{-1}H^T\beta. \tag{7}$$

A new testing sample $x\in R^n$ is assigned to class $i(i=+1,-1)$ by

$$\text{Class } i=\arg\min_{k=1,2}\frac{|x^Tw_k+b_k|}{\|w_k\|}. \tag{8}$$

In TSVM, if the number of samples in two classes is approximately equal to $l/2$, the computational complexity of TSVM is $O(2\times(l/2)^3)$. Thus the ratio of run-times between SVM and TSVM is $l^3/(2\times(l/2)^3)=4$, which implies that TSVM works approximately four times faster than SVM in theory [2].

2.2. Multi-class twin support vector machine

Twin-KSVC generates two nonparallel hyper-planes for the two focused classes of samples such that each hyper-plane is closer to one class and as far as possible from the other. The remaining samples are mapped into a region and satisfy constraints both $-(x^Tw_1+b_1)\ge 1-\epsilon$ and $x^Tw_2+b_2\ge 1-\epsilon$, where ϵ is a positive parameter chosen a priori. The two nonparallel hyper-planes

$$x^Tw_1+b_1=0 \quad \textit{and} \quad x^Tw_2+b_2=0 \tag{9}$$

can be obtained by solving the following pair of QPPs,

$$\begin{aligned}\min_{w_1,b_1,\xi,\eta}\ & \frac{1}{2}\|Aw_1+e_1b_1\|^2+c_1e_2^T\xi+c_2e_3^T\eta\\ \text{s.t.}\ & -(Bw_1+e_2b_1)+\xi\ge e_2,\\ & -(Cw_1+e_3b_1)+\eta\ge e_3(1-\epsilon),\\ & \xi\ge 0e_2,\quad \eta\ge 0e_3,\end{aligned} \tag{10}$$

and

$$\begin{aligned}\min_{w_2,b_2,\xi^*,\eta^*}\ & \frac{1}{2}\|Bw_2+e_2b_2\|^2+c_3e_1^T\xi^*+c_4e_3^T\eta^*\\ \text{s.t.}\ & (Aw_2+e_1b_2)+\xi^*\ge e_1,\\ & (Cw_2+e_3b_2)+\eta^*\ge e_3(1-\epsilon),\\ & \xi^*\ge 0e_1,\quad \eta^*\ge 0e_3,\end{aligned} \tag{11}$$

where $w_1\in R^{n\times 1}$, $w_2\in R^{n\times 1}$, $b_1\in R^{1\times 1}$, $b_2\in R^{1\times 1}$, $\xi\in R^{l_2\times 1}$, $\eta\in R^{l_3\times 1}$, $\xi^*\in R^{l_1\times 1}$, $\eta^*\in R^{l_3\times 1}$, $e_1\in R^{l_1\times 1}$, $e_2\in R^{l_2\times 1}$, and $e_3\in R^{l_3\times 1}$.

By introducing the Lagrangian multipliers α and β, we can derive their dual problems as follows:

$$\begin{aligned}\max_{\alpha}\ & e_4^T\alpha-\frac{1}{2}\alpha^TN(H^TH)^{-1}N^T\alpha\\ \text{s.t.}\ & 0\le\alpha\le F,\end{aligned} \tag{12}$$