Knowledge-Based Systems 88 (2015) 34–44

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

K-nearest neighbor based structural twin support vector machine

Xianli Pan, Yao Luo, Yitian Xu *

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

ARTICLE INFO

Article history:
Received 3 March 2015
Received in revised form 12 August 2015
Accepted 14 August 2015
Available online 22 August 2015

Keywords:
K-nearest neighbors
Structural information
Twin support vector machine
Weights

ABSTRACT

Structural twin support vector machine (S-TSVM) performs better than TSVM, since it incorporates the structural information of the corresponding class into the model. However, the redundant inactive constraints corresponding to non-support vectors (non-SVs) are still the burden of the solving process. Motivated by the KNN trick presented in the weighted twin support vector machines with local information (WLTSVM), we propose a novel K-nearest neighbor based structural twin support vector machine (KNN-STSVM). By applying the intra-class KNN method, different weights are given to the samples in one class to strengthen the structural information. For the other class, the superfluous constraints are deleted by the inter-class KNN method to speed up the training process. For large scale problems, a fast clipDCD algorithm is further introduced for acceleration. Comprehensive experimental results on twenty-two datasets demonstrate the efficiency of our proposed KNN-STSVM.

1. Introduction

In the field of machine learning, support vector machine (SVM) [1–3], founded on Vapnik's statistical learning theory, is a novel and effective approach for classification or regression analysises. It is proved to outperform most other systems [2,4,5] in a wide variety of applications, such as network intrusion detection, disease diagnosis, and human face detection. For classification problems, the target is to find a hyperplane that can separate different classes well. Many hyper-planes can meet this requirement, and SVM finds the one following the maximum margin principle in statistical learning theory [1]. It is a global optimal solution obtained by solving a quadratic programming problem (QPP) in the dual space. The introduction of the kernel function [6] in SVM maps training vectors into a high-dimensional space, successfully transforms the nonlinear case into linear case. On the basis of the techniques above, we can find many good properties of SVM. However, to solve the entire QPP is very time consuming, which still remains challenging.

Recently proposed twin support vector machine (TSVM) [1] improves this problem and becomes popular. It generates two non-parallel hyper-planes for the two classes, such that each hyper-plane is closest to one class and is at a distance of at least one from the other class. The label of a new sample is decided by comparing which of the two hyper-planes it lies closer to. The greatest superiority of TSVM over SVM is that it solves a pair of smaller-sized QPPs rather than a single large one, which makes TSVM faster. Many experimental results have shown that TSVM obtains faster learning speed than classical SVM. So far, scholars have developed numerous improved models based on TSVM [8–10]. Nevertheless, some problems still exists in TSVM.

One disadvantage is that the TSVM is blind to data distribution. It pays more attention on separating two classes well, thus it ignores the helpful underly structural information within classes. Based on the studies of structural SVM, such as SLMM [11], SRSVM [12], and inspired by TSVM, Qi proposed a novel structural twin support vector machine (S-TSVM) in [13]. This S-TSVM extracts the structural information by using a hierarchical clustering method, and then minimizes the tightness of each cluster with the form of covariance matrix. Experimental results reveal that S-TSVM owns excellent generalization ability and comparable learning speed. From the perspective of further improving the computational speed, we put forward the structural least square twin support vector machine (S-LSTSVM) in our previous article [14]. By introducing the least squares technique into S-TSVM, S-LSTSVM only needs to solve a pair of linear equations, which greatly accelerates the calculation progress.

Nevertheless, we find two drawbacks of the existing S-TSVM:

(1) The overall compactness of clusters is minimized in S-TSVM (or S-LSTSVM), but samples within these clusters have different importance, which is ignored. To further study the different importance of samples may strengthen the structural information of clusters, and then improve the classification accuracy.

* Corresponding author. Tel.: +86 10 62737077; fax: +86 10 62736563.
E-mail address: xytshuxue@126.com (Y. Xu).

http://dx.doi.org/10.1016/j.knosys.2015.08.009
0950-7051/

(2) Each model of S-TSVM pays attention to the data structure from its corresponding class. For the other class in constraints, all samples are required to be on one side of the corresponding support hyperplane. That is to say, the support vectors (SVs) play a decisive role in the constraints. Hence, most constraints of non-SVs are redundant and slow down the training speed of the model.

Exactly, the WLTSVM proposed by Ye et al. introduces a novel KNN method, which gives different treatments to different classes in objective function or constraints of the model. For the corresponding class in the objective function, weight is given to each sample according to its intensive degree. For the constraints, 0 weights are given to the samples regarded as non-SVs, which means to discard the corresponding constraints. This simple trick not only improves the classification accuracy but also reduces the time complexity by only keeping the possible SVs in the constraints. We have successfully applied it to settle the regression problem in our previous work [16]. And for the disadvantages of S-TSVM mentioned above, we can find that this KNN method is a good solution.

Inspired by the studies above, we propose a K-nearest neighbor based structural twin support vector machine (KNN-STSVM). With the KNN trick applied in S-TSVM, we not only strengthen the structural information by giving different weights to samples among clusters, but also remove the redundant constraints to speed up the computation process. The whole process can be divided into three steps: (1) Get proper clusters with the Ward's agglomerative hierarchical clustering method [17], and calculate the covariance matrix of each cluster. (2) Use the KNN method [15] to get the weights of samples from one class and select the possible SVs from the other class. W (the weighted matrix) and f (the vector with only 0 and 1) are obtained before the next step. (3) Solve the KNN-STSVM by two smaller-sized QPPs which is similar to the TSVM.

However, like the SVM, the long training time is still one of the main challenges of our KNN-STSVM, especially for large scale problems. So far, many fast optimization algorithms have been presented to accelerate the training speed, including the decomposition method [18], sequential minimal optimization (SMO) [19], the geometric algorithms [20,21], the clipping dual coordinate descent algorithm (clipDCD) [22–24] and so on. The recently proposed clipDCD algorithm can directly solve the dual QPPs of SVM by orderly updating one variable with a single-variable sub-problem. The updated variable will derive the largest decrease on the objective value. This clipDCD method shows a fast learning speed and great efficiency in experiments. In this paper, we introduce it into our KNN-STSVM for the large scale datasets. Comparisons of the KNN-STSVM with some other algorithms in terms of classification accuracy and learning time have been made on several UCI datasets, indicating the superiority of our KNN-STSVM.

The rest of this paper is organized as follows. Section 2 outlines the S-TSVM and WLTSVM. Details of KNN-STSVM are given in Section 3, both the linear and nonlinear cases are included. Section 4 analyzes the characteristics of KNN-STSVM. Section 5 discusses the experimental results on twenty-four benchmark datasets to investigate the feasibility and validity of our proposed algorithm. In Section 6, the clipDCD is introduced into KNN-STSVM to accelerate the learning speed. Conclusions drown from this study are in the last section.

2. Related works

In this section, we review the basics of the S-TSVM and WLTSVM. For the binary classification problems, the training samples are denoted by a set $T=\{(x_1,y_1),\ldots,(x_l,y_l)\}$, where $x_i\in R^n,\ y_i\in\{1,-1\},\ i=1,\ldots,l$. For convenience, matrix A in $R^{l_1\times n}$ represents the points of class +1 and matrix B in $R^{l_2\times n}$ represents the points of class -1.

2.1. Structural twin support vector machine

S-TSVM extracts the data distribution information of each class through the Ward's linkage clustering method. Then it minimizes these clusters with the form of the covariance matrices in the objective functions of TBSVM [25]. The S-TSVM characterizes each class more precisely than TSVM as it combines the structural information of samples into the model. Suppose there are C_P and C_N clusters in the positive class P and negative class N, respectively, i.e., $P=P_1\cup\cdots P_i\cup\cdots P_{C_P},\ N=N_1\cup\cdots N_j\cup\cdots N_{C_N}$. The subsequent optimization problem of S-TSVM can be formulated as follows [13]:

$$\begin{aligned}\min_{w_+,b_+,\xi}\quad & \frac{1}{2}\|Aw_++e_+b_+\|_2^2+c_1e_-^T\xi+\frac{c_2}{2}(\|w_+\|_2^2+b_+^2)\\ & +\frac{c_3}{2}w_+^T\Sigma_+w_+,\\ \text{s.t.}\quad & -(Bw_++e_-b_+)+\xi\geqslant e_-,\\ & \xi\geqslant 0,\end{aligned}\tag{1}$$

and

$$\begin{aligned}\min_{w_-,b_-,\eta}\quad & \frac{1}{2}\|Bw_-+e_-b_-\|_2^2+c_4e_+^T\eta+\frac{c_5}{2}(\|w_-\|_2^2+b_-^2)\\ & +\frac{c_6}{2}w_-^T\Sigma_-w_-,\\ \text{s.t.}\quad & (Aw_-+e_+b_-)+\eta\geqslant e_+,\\ & \eta\geqslant 0,\end{aligned}\tag{2}$$

where ξ and η are slack vectors, both e_+ and e_- are vectors of ones of appropriate dimensions. $\Sigma_+=\Sigma_{P_1}+\cdots+\Sigma_{P_{C_P}}$, $\Sigma_-=\Sigma_{N_1}+\cdots+\Sigma_{N_{C_N}}$, Σ_{P_i} and Σ_{N_j} are the corresponding covariance matrices of clusters in the two classes. $c_i\ (i=1,\ldots,6)$ is a pre-specified factor. S-TSVM solves two smaller-sized QPPs as follows:

$$\begin{aligned}\max_{\alpha}\quad & -\frac{1}{2}\alpha^TG(H^TH+c_2I+c_3J)^{-1}G^T\alpha+e_-^T\alpha,\\ \text{s.t.}\quad & 0\leqslant\alpha\leqslant c_1e_-,\end{aligned}\tag{3}$$

and

$$\begin{aligned}\max_{\beta}\quad & -\frac{1}{2}\beta^TP(Q^TQ+c_5I+c_6F)^{-1}P^T\beta+e_+^T\beta,\\ \text{s.t.}\quad & 0\leqslant\beta\leqslant c_4e_+,\end{aligned}\tag{4}$$

where $G=[B\ e_-],\ H=[A\ e_+],\ J=\begin{bmatrix}\Sigma_+ & 0\\ 0 & 0\end{bmatrix},\ P=[A\ e_-],\ Q=[B\ e_+]$, and $F=\begin{bmatrix}\Sigma_- & 0\\ 0 & 0\end{bmatrix}$.

A new data point is classified as class $+1$ or class -1 depends on which of the two hyper-planes it lies closet to.

2.2. Weighted twin support vector machine with local information

The main idea of WLTSVM [15] is to discover the intrinsic similarity within samples from the same class and extract the possible SVs residing in the constraints. Inspired by the graph-based dimensionality reduction method [26], WLTSVM constructs two graphs, a intra-class graph W_s and a inter-class graph W_d. Then, weights W_s are given to samples in the objective functions, and the constraints will be more succinct via W_d. Similar to TSVM, the WLTSVM also tries to find two nonparallel hyper-planes, each of which fits the points of the corresponding class. Considering the analogous geometric interpretation of TSVM, WLTSVM solves a pair of smaller sized QPPS as follows [15]: