Information Sciences 465 (2018) 174–190

Contents lists available at ScienceDirect

Information Sciences

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

Scaling up twin support vector regression with safe screening rule

Hongmei Wang [a], Yitian Xu [b,*]

[a] College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China
[b] College of Science, China Agricultural University, Beijing 100083, China

ARTICLE INFO

Article history:
Received 14 December 2017
Revised 4 July 2018
Accepted 6 July 2018
Available online 9 July 2018

Keywords:
Twin support vector regression
Large-scale problem
Safe screening rule
Sparse
Variational inequality

ABSTRACT

Twin support vector regression (TSVR) is a popular and efficient regression method, since it solves a pair of smaller-sized quadratic programming problems (QPPs) rather than a single large one as in the traditional SVR. However, it is time-consuming to deal with the large-scale problems, especially for the multi-parameter case. Inspired by the sparsity of TSVR, we propose an efficient safe screening rule based on variational inequality (VI) to accelerate TSVR, termed as SSR-TSVR. Through this rule, most 0 and 1 components in dual solution can be identified before actually training TSVR. Then the scale of the model will be extremely reduced by preassigning the identified components. In this way, the computational time of TSVR can be sharply shortened. There are two main advantages of our method: (1) it is safe in the sense that it guarantees to achieve the exactly same solution as solving original problem; (2) it is efficient for both the linear and nonlinear cases. Another contribution is that the dual coordinate descent method (DCDM) is employed to further accelerate the computational speed. Experimental results on twelve benchmark datasets demonstrate the efficiency and safety of our proposed method.

1. Introduction

Support vector machine (SVM) [1,4–6,8,17,45], motivated by the Vapnik Chervonenkies (VC) dimensional theory and statistical learning theory [33], is a promising technique in machine learning. The classical SVM [6] is to maximize the margin between two classes by minimizing the regularization term and to minimize classification errors by applying hinge loss function. It has evolved into a multitude of diverse formulations with different properties, such as ν-SVM [6] and LS-SVM [2,31]. So far, SVM has been successfully applied in various fields, such as text categorization [12], speech recognition [7], information security [20], biomedicine [10], and so on.

Although SVM owns a good classification ability, it is time-consuming to train it. In order to improve the computational speed, twin support vector machine (TSVM) was proposed by Jayadeva et al. [12], which is motivated by generalized eigenvalue proximal support vector machine (GEPSVM) [15]. It aims to construct two nonparallel hyper-planes such that each plane is closer to one class and as far as possible from the other. Only two smaller-sized QPPs need to be solved, which makes the computational speed faster. Many improved models have been proposed in [14,25,27,29].

* Corresponding author.
E-mail address: xytshuxue@126.com (Y. Xu).

https://doi.org/10.1016/j.ins.2018.07.008

As for support vector regression (SVR) [31,36–39], it adopts ϵ-insensitive loss function and has good generalization capability in some applications. However, it is expensive to train it. Inspired by TSVM, an efficient twin support vector regression (TSVR) was proposed by Peng et al. [24]. It aims at generating nonparallel ϵ-insensitive up-bound and down-bound functions by solving two smaller-sized QPPs rather than a single large one in SVR. Therefore, the computational speed of TSVR is approximately four times faster than SVR. Currently, TSVR has become one of the most popular regression methods because of its low computational complexity. Many modified TSVR has been proposed in [3,32,42–44,49]. However, the applications of TSVR to large-scale problems still pose significant challenges.

At present there exists many approaches to accelerate TSVR. On one hand, the preprocessing method [50] based on statistical learning theory removes most training samples before solving the problem. Although the scale of the problem is reduced, it may mistakenly delete some related training samples. The performance of TSVR will be affected more or less, i.e., it is not safe. On the other hand, many fast algorithms have been proposed to speed up TSVR, including modified Newton method (MNM) [13], geometric algorithms (GA) [18,19], successive over-relaxation algorithms (SOR) [28], sequential minimal optimization (SMO) [30], genetic algorithm [40] and dual coordinate descent method (DCDM) [11,26]. However, the complexity of these algorithms grows fast with the number of samples because they can not indeed reduce the scale of the problem.

Recently, a safe screening rule (SSR) was first proposed by El. Ghaoui et al. [9] in the context of l_1 sparse regression. This rule can safely discard many irrelevant features, which has 0 coefficients in the optimal solution before actually solving it. In this way, the speedup can be several orders of magnitude. Due to its significant performance on sparse model, some extensions [16,35,41] have been reported for other models, including fast lasso, group lasso, and so on. Safe screening rule was first introduced into SVM by Ogawa et al. [21]. Unlike l_1 sparse regression problem, it aims to identify the non-support vectors (non-SVs) in SVM and preassign them prior to training phase, which leads to substantial reduction in the training cost. Later, Wang et al. [34] proposed a novel safe screening rule to discard non-support vectors by analyzing the dual problem of SVM via variational inequalities (VI). It focuses on the connection between the two optimization problems with adjacent parameters. For example, suppose that two dual optimization problems are $D(c_j)$ and $D(c_{j+1})$, if the optimal solution $\alpha^*(c_j)$ is available, most 0 and 1 components in $\alpha^*(c_{j+1})$ can be screened out based on it. Then, only one smaller-sized problem needs to be solved to obtain the remained components. In this way, the computational time can be greatly saved. Now SSR has been widely applied to other SVM models [21,46–48] and TSVM model [23].

For TSVR, the regressor is determined only by a small part of samples called support vectors (SVs) and independent of the remained samples (non-SVs). Motivated by the above studies, in this paper we propose a safe screening rule based on VI for TSVR to accelerate its training speed. It can identify most non-SVs and preassign them before the training phase, which reduces the scale of the problem. Then the computational time will be greatly saved. It's worth mentioning that we put forward the corresponding screening rule for each parameter. Thus, it is efficient during all parameters tuning process. There are two differences between our reduction method and the existing preprocessing methods [50]: (i) Our SSR can guarantee that the identified samples must be non-SVs. However, the preprocessing method may mistakenly discard SVs, i.e., it is not safe; (ii) Our method is embedded in all parameters tuning process. It can discard different scale of samples with different parameters. But the preprocessing method discards same samples with different parameters.

To sum up, our method is very efficient to accelerate TSVR. The main contributions and advantages of our SSR-TSVR are as follows:

1) We construct the SSR for accelerating TSVR with multiple parameters. Therefore, our method is an extension of the existing methods.
2) One significant superiority of our SSR is that we can obtain the exactly same solution as the original problem, i.e., our SSR is safe.
3) Most existing safe screening methods are only available for the linear SVMs. However, our SSR-TSVR is further extended to the nonlinear case.
4) SSR is independent of the solvers. Therefore, any existing efficient solvers can be combined with it. To further improve the computational speed, an efficient DCDM is chosen as the solver.

Numerical experiments on twelve benchmark datasets are conducted. The experimental results show the safety and efficiency of our proposed SSR-TSVR.

This paper is designed as follows: Section 2 briefly describes the basic knowledge of SVR and TSVR. We analyze the property of dual solution in Section 3. Section 4 presents the details of the safe screening rule for TSVR, gives the reduced TSVR (RTSVR), and introduces the DCDM to solve the TSVR and RTSVR. Section 5 analyzes the complete algorithm process. In Section 6, we perform experiments on twelve benchmark datasets to verify the efficiency of our method. The last section gives the conclusions of this paper.

2. Related work

In this section, we briefly review the classical SVR and TSVR. Given a training set $T = \{(x_1, y_1), (x_2, y_2), \cdots, (x_l, y_l)\}$, where $x_i \in R^n$, $y_i \in R$, $i = 1, 2, \cdots, l$. For notational convenience, let matrix $A = [x_1, x_2, \cdots, x_l]^T \in R^{l\times n}$ and matrix $Y = [y_1; y_2; \cdots; y_l] \in R^l$. $\|\cdot\|$ refers to the vector norm. e denotes the vector of ones of appropriate dimensions.