Neurocomputing 287 (2018) 163–172

Contents lists available at ScienceDirect

Neurocomputing

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

# A safe screening based framework for support vector regression

Xianli Pan[a], Xinying Pang[a], Hongmei Wang[b], Yitian Xu[a,*]

[a] College of Science, China Agricultural University, Beijing 100083, China
[b] College of information and electrical engineering, China Agricultural University, Beijing 100083, China

ARTICLE INFO

*Article history:*
Received 29 March 2017
Revised 18 January 2018
Accepted 31 January 2018
Available online 6 February 2018

Communicated by Dr. Kaizhu Huang

*Keywords:*
Support vector regression
Safe screening
Grid search
Sample selection

ABSTRACT

Support vector regression (SVR) is popular and efficient for regression problems. However, it is time-consuming to solve it, especially for large datasets. Inspired by the sparse solutions of SVR, a safe screening based framework for SVR (SVR-SBF), including both linear and nonlinear cases, is proposed in this paper to improve its training speed. This SBF has two steps: First, the constant solutions of SVR along the regularization path of parameter $C$ are deleted before training; Second, a safe screening rule via variational inequalities (SSR-VI) is embedded into the grid search method to further discard the inactive solutions of SVR. This SBF can efficiently accelerate the training speed of SVR without affecting its solutions. Compared to existing safe rules, our SVR-SBF can identify more inactive solutions by finding constant solutions beforehand. In addition, our SBF is further expanded to more situations and models. To be specific, a modified SSR-VI is proposed to be adapted to other parameter selection methods, and models including variants of SVR and classical SVM are analyzed. Experiments on both synthetic and real datasets are conducted to demonstrate the superiority of SVR-SBF.

© 2018 Elsevier B.V. All rights reserved.

## 1. Introduction

Support vector regression (SVR) [1–3] is an efficient model for regression problems [4–6]. It has many attractive properties, including sparsity, robustness and excellent generalization ability. Therefore, a lot of improved methods [7–11] are developed based on it. So far, these models have been widely applied to various areas, including pattern recognition [12,13], bioinformatics [14], image understanding [15,16] and document summarization [17].

In SVR, two terms are minimized in the objective function. The first term is training error, which is represented as the $\epsilon$-insensitive loss function. Sparse solutions will be obtained due to this specific loss function. The second term is the regularization term, which controls the complexity of the model. Similar to the classical support vector machine (SVM), SVR enjoys solid theoretical foundation and good statistical properties. However, it is expensive to solve SVR for the reason that it has $2l$ constraints ($l$ is the number of training samples), which is two times of SVM.

To accelerate the training process of SVR, a variety of techniques have been proposed. Here, we roughly divide them into two categories:

**Sample selection methods.** To reduce the scale of SVR itself, many efficient sample selection methods have been proposed. They select samples that are possible to be support vectors (SVs) with some priori information. For example, methods in references [18,19] search the potential SVs with $k$ nearest neighbors. However, real SVs may be mistakenly deleted by these methods and the performance of SVR will be affected more or less.

**Fast solving algorithms.** This kind of methods includes modified Newton methods (MNM) [20], geometric algorithms (GA) [21,22], successive overrelaxation algorithms (SOR) [23] and so on. Among them, the dual coordinate descent method (DCDM) [24,25] is popular and efficient, especially for problems with sparse solutions. However, fast solving algorithms can not indeed reduce the scale of SVR.

Recently, a kind of safe sample screening methods has been proposed for SVM [26–31]. They can enormously reduce the scale of SVM by finding and eliminating the real non-SVs before training. Most importantly, it can be proved that this kind of methods gives identical solutions with the original SVM. However, existing safe screening rules still have some shortcomings. To the best of our knowledge, existing safe rules are only embedded with the grid search (GS) [32] method. They have to repeatedly search the same non-SVs in SVM for all parameters, which is a waste of time.

Inspired by the studies above, in this paper, we propose an efficient safe screening based framework (SBF) for SVR (SVR-SBF) to accelerate its training speed. SVR-SBF has two steps. In the first step, the constant solutions are found and deleted before training according to the regularization path of parameter $C$ [33]. Constant solutions represent the solutions whose values keep as a constant

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

https://doi.org/10.1016/j.neucom.2018.01.083
0925-2312/© 2018 Elsevier B.V. All rights reserved.

when parameter $C$ changes. In this way, the scale of SVR can be substantially reduced for all $C$. We call the simplified optimization problem SVR-1. In the second step, a safe screening rule via variational inequalities (SSR-VI) is proposed for SVR-1 with each parameter $C_i$. To be specific, if the solution of SVR-1($C_{i-1}$) is known, the inactive components (inactive components mean the components with "0" and "1" values) in solution of SVR-1($C_i$) will be safely identified. In this way, the scale of SVR-1($C_i$) is further reduced. Applying SSR-VI sequentially with the GS method, the whole parameter selection process will be accelerated. In summary, SVR-SBF achieves a two-layer acceleration and is more efficient than existing safe rules.

In addition, considering that besides the GS method, many other efficient parameter selection methods have been proposed [34–39]. Therefore, to extend the application scope of our SVR-SBF, a modified SSR-VI is further proposed to be adapted to other parameter selection methods, such as the genetic algorithm (GA) [37,38] the particle swarm optimization (PSO) [39] and so on.

In general, the main innovations of this paper are as follows:

- This is the first time to construct safe screening methods for SVR. Since the number of constraints in SVR is almost two times of SVM, it is more meaningful to develop safe sample screening methods for it.
- SVR-SBF is more efficient than existing methods. Existing safe rules have to search the same inactive solutions iteratively. To avoid doing this, the first step of SVR-SBF identifies constant solutions beforehand. This technique saves a lot of searching and calculation time.
- Most existing safe screening methods are applied to linear SVMs. In this paper, SBF is further extended to kernel SVR which needs much more training time.
- SSR-VI is independent of parameter selection. To the best of our knowledge, existing safe rules are all done with grid-searching. In this paper, a modified SSR-VI is proposed to be adapted to other efficient parameter selection methods. This expands the application scope of our method.
- SBF is independent of the solving algorithm. Therefore, any efficient algorithms can be combined with it.

Properties of SVR-SBF, including computational complexity and scalability, are further analyzed. Discussions on extending SBF to variants of SVR and classical SVM are also given to expand our method to other models. Experiments on synthetic and real datasets are conducted. The experimental results verify the efficiency and safety of SVR-SBF.

This paper is structured as follows: Section 2 reviews the basic concepts and property of SVR. SVR-SBF is introduced in Section 3. Properties of SVR-SBF are analyzed in Section 4. Section 5 gives experiments on fifteen real-word datasets to assess the effectiveness and safety of SVR-SBF for both linear and nonlinear cases. Conclusions drown from this study are in the last section.

## 2. Support vector regression

In this section, the basics of SVR and related properties are introduced.

### 2.1. The formulation of SVR

Given training set $T = \{(x_1, y_1), (x_2, y_2), \ldots, (x_l, y_l)\}$ where $x_i \in R^d$ and $y_i \in R$, the goal of SVR is to find the prediction function $f(x) = w^T\phi(x)$ (the bias term is dropped by preprocessing the data with zero mean). The formulation of SVR can be expressed as

$$\min_{w,\xi^{(*)}} \quad \frac{1}{2}||w||_2^2 + Ce^T(\xi + \xi^*) \tag{1}$$
$$\text{s.t.} \quad \langle\phi(x_i), w\rangle - y_i \le \epsilon + \xi_i,$$
$$y_i - \langle\phi(x_i), w\rangle \le \epsilon + \xi_i^*,$$
$$\xi_i, \xi_i^* \ge 0, i = 1, 2, \ldots, l.$$

In most cases, problem (1) can be solved easily in the following dual formulation.

$$\min_{\hat{\alpha}} \quad \frac{C}{2}\hat{\alpha}^T H\hat{\alpha} + g^T\hat{\alpha} \tag{2}$$
$$\text{s.t.} \quad 0 \le \hat{\alpha} \le e.$$

Here, $H = \begin{bmatrix} K & -K \\ -K & K \end{bmatrix}, K \in R^{l\times l}$ is the kernel function, $K_{ij} = K(x_i, x_j), i, j \in \{1, 2, \ldots, l\}$, and $g = \begin{bmatrix} \epsilon e - Y \\ \epsilon e + Y \end{bmatrix}$. $\hat{\alpha} = \frac{1}{C}\begin{bmatrix} \alpha^* \\ \alpha \end{bmatrix}$, where $\alpha^*$ and $\alpha$ are Lagrange multipliers [40].

Many efficient algorithms have been proposed to solve problem (2). In this paper, DCDM is chosen as the solving algorithm.

### 2.2. The grid search method for SVR

There are three parameters in problem (2), i.e., $C$, $\epsilon$ and the kernel parameter $p$. In real applications, they need to be determined with efficient parameter selection methods. The GS method is a commonly used technique whose framework is given in the following algorithm.

From Algorithm 1, we can see that SVR needs to be trained with every combination of the parameters ($C$, $\epsilon$, $p$) with the GS method. The whole process is time-consuming, especially for large scale problems. Our SVR-SBF is developed based on this SVR-GS.

**Algorithm 1** SVR-GS.

**Input:** $X_{train}, Y_{train}, X_{test}, Y_{test}; C \in R^{m_1}, \epsilon \in R^{m_2}, p \in R^{m_3}$;
**Output:** Optimal parameters $(C^*, \epsilon^*, p^*)$;

```
GS-SVR:
s = 0;
for k = 1 to m3 do
    for i = 1 to m2 do
        for j = 1 to m1 do
            s ←s+1;
            α̂[s] ← SVR(C[j], ϵ[i], p[k]);
            MSE[s] ← predict(α̂[s], X_test, Y_test);
        end for
    end for
end for
(C*, ϵ*, p*) ← parameters corresponding to min(MSE).
```

### 2.3. Basic property of SVR

For SVR, the following proposition is its important basic property.

**Proposition 1.** *[41] Suppose that $\hat{\alpha} = \frac{1}{C}[\alpha_1^*, \alpha_1, \ldots, \alpha_l^*, \alpha_l]$ is the solution of problem (2), and $f(x) = w^T\phi(x)$ is the corresponding regression function. Then*

(i) *if $\hat{\alpha}_i = 1, \hat{\alpha}_i^* = 0$ or $\hat{\alpha}_i^* = 1, \hat{\alpha}_i = 0$, then the corresponding point $(x_i, y_i)$ lies outside or on the boundary of the $\epsilon$-tube of hyperplane $f(x) = w^T\phi(x)$;*
(ii) *if $0 < \hat{\alpha}_i < 1, \hat{\alpha}_i^* = 0$ or $0 < \hat{\alpha}_i^* < 1, \hat{\alpha}_i = 0$, then the corresponding point $(x_i, y_i)$ lies on the boundary of the $\epsilon$-tube of hyperplane $f(x) = w^T\phi(x)$;*
(iii) *if $\hat{\alpha}_i = \hat{\alpha}_i^* = 0$, then the corresponding point $(x_i, y_i)$ lies inside or on the boundary of the $\epsilon$-tube of hyperplane $f(x) = w^T\phi(x)$.*

Proof of proposition 1 can be found in [41].

دریافت فوری ← متن کامل مقاله

ISIArticles
مرجع مقالات تخصصی ایران

✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
✓ امکان دانلود نسخه ترجمه شده مقالات
✓ پذیرش سفارش ترجمه تخصصی
✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
✓ دانلود فوری مقاله پس از پرداخت آنلاین
✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات