Neurocomputing 168 (2015) 941–946

Contents lists available at ScienceDirect

Neurocomputing

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

A robust weighted least squares support vector regression based on least trimmed squares

Chuanfa Chen [a,b,*], Changqing Yan [c], Yanyan Li [d]

[a] State Key Laboratory of Mining Disaster Prevention and Control Co-founded by Shandong Province and the Ministry of Science and Technology, Shandong University of Science and Technology, 266590 Qingdao, China
[b] College of Geomatics, Shandong University of Science and Technology, 266590 Qingdao, China
[c] Department of Information Engineering, Shandong University of Science and Technology, 271019 Taian, China
[d] Shool of Geodesy and Geomatics, Wuhan University, 430072 Wuhan, China

ARTICLE INFO

Article history:
Received 14 September 2014
Received in revised form
8 December 2014
Accepted 9 May 2015
Communicated by: Wei Chiang Hong
Available online 19 May 2015

Keywords:
Outlier
Least squares support vector regression
Least trimmed squares
Robust
Effciency

ABSTRACT

In order to improve the robustness of the classcial LSSVM when dealing with sample points in the presence of outliers, we have developed a robust weighted LSSVM (reweighted LSSVM) based on the least trimmed squares technique (LTS). The procedure of the reweighted LSSVM includes two stages, respectively used to increase the robustness and statistical efficiency of the estimator. In the first stage, LTS-based LSSVM (LSSVM-LTS) with C-steps was adopted to obtain robust simulation results at the cost of losing statistical efficiency to some extent. Thus, in the second stage, the results computed in the first stage were optimized with a weighted LSSVM to improve efficiency. Two groups of examples including numerical tests and real-world benchmark examples were respectively employed to compare the robustness of the reweighted LSSVM with those of the classical LSSVM, the weighted LSSVM and LSSVM-LTS. Numerical tests indicate that the reweighted LSSVM is comparable to the weighted LSSVM, and more accurate than the classical LSSVM and LSSVM-LTS when the contaminating proportion is small (i.e. 0.1 and 0.2), whereas with the increase of contaminating proportion, the reweighted LSSVM performs much better than other methods. The real-world exmaple of regressing seven benchmark datasets demonstrates that the reweighted LSSVM is always more accurate than other versions of LSSVM. In conclusion, the newly developed method can be considered as an alternative to function estimation, especially for sample points in the presence of outliers.

1. Introduction

Support vector machines (SVM) for classification and nonlinear function estimation, as introduced by Vapnik [1–3], is based on the structural risk minimization principle which strikes a balance between the traditional empirical risk and model complexity. It is implemented by maximizing margin between the samples of two classes and simultaneously minimizing the classification errors of training samples by convex optimization without suffering from many local minima [4–7]. Due to its appealing generalization performance, SVM has been widely accepted as an important tool in the area of pattern recognition including classification, regression and function approximation [8–12].

Driven by the dream to make the approach as simple as possible, least squares versions of support vector machine (LSSVM) have been presented [13–15]. LSSVM only needs to solve a linear equation set rather than dealing with a quadratic programming problem, by using equality constraints instead of inequality ones and a least squares loss function, which greatly reduces the computational complexity. However, despite these computationally attractive features, a potential drawback of LSSVM is that it is only optimal when sample points follow a normal distribution [16–18]. Namely, it is sensitive to outliers and noises with a non-normal distribution.

Many researches have been conducted to improve the robustness of LSSVM for dataset in the presence of outliers. Some authors argued that outliers should be filtered out first by some advanced techniques, and then the non-robust LSSVM is used to train the remaining sample points. This is the well known 'two-step' method. For example, Wen et al. [19] used a criterion derived from least trimmed squares regression to recursively eliminate outliers. Chuang and Lee [20] removed those points with the slack

* Corresponding author at: State Key Laboratory of Mining Disaster Prevention and Control Co-founded by Shandong Province and the Ministry of Science and Technology, Shandong University of Science and Technology, 266590 Qingdao, China. Tel.: +86 053286057845.
E-mail address: chencf@lreis.ac.cn (C. Chen).

http://dx.doi.org/10.1016/j.neucom.2015.05.031
0925-2312/

variables bigger than the three times standard deviation. However, the detection of outliers is difficult only by the results of LSSVM when there is no prior knowledge, as LSSVM itself is sensitive to outliers.

Some scholars were prone to developing weighted LSSVMs to mitigate the influence of outliers. The core idea of this approach is how to accurately determine the weights of sample points. For example, Suykens et al. [21] pointed out that samples with large simulation residuals by the classical LSSVM should have smaller weights. Wen et al. [22] indicated that samples with large distances from others have relatively smaller weights. Brabanter et al. [23] compared four different types of weighting function including Huber, Hampel, Logistic and Myriad, and found that Logistic and Myriad weighting function are more robust than the other two functions in most cases. Motivated by the weighted LSSVM, some adaptive versions has been proposed, such as adaptive weighted LSSVM integrated with outlier detection [24] and weighted Lq adaptive LSSVM [25]. However, many authors indicated that it is unclear whether these weighting schemes are optimal with respect to the dataset subject to noises and outliers.

Some authors preferred to use robust loss functions instead of weighting schemes to reduce the effect of outliers. For example, Chen et al. [26] adopted the maximum correntropy criterion as the loss function, which is derived from information theoretic learning. Yang et al. [27] used a truncated least squares loss function. Yet, these robust loss functions are non-convex, so their corresponding objective functions are hard to be optimized without some transformations. Furthermore, many parameters in the loss function must be pre-determined by an optimization technique, which costs much time.

Generally, the procedure of the above robust methods consists of at least two stages. The first stage always employs the classical LSSVM to obtain initial simulation results. Namely, the accuracy of the robust LSSVMs depends on the results of the unweighted LSSVM to some extent [21,24,25]. Thus, if the unweighted LSSVM regression obtains unsatisfactory results, the robust versions may not perform well. Hence, to improve the prediction accuracy of LSSVMs, the initial results in the first stage should be robust.

It has been well known in statistics that the least trimmed squares (LTS) criterion is much more robust than the least squares method, and has been widely applied to the context of nonparametric regression [28–31]. In principle, LTS with a smooth objective function has a high breakdown point [30,32–34]. Hence, in this paper, we extend the LTS regression method to the context of LSSVM, and propose a robust LSSVM, termed as LSSVM-LTS. Furthermore, in order to improve the statistical efficiency of LSSVM-LTS, a weighted LSSVM is introduced to optimize simulation results.

The rest of this paper is organized as follows. In Section 2, we briefly review LSSVM and weighted LSSVM, respectively. In Section 3, we propose LSSVM-LTS and provide an algorithm to compute the LSSVM-LTS estimator. Furthermore, a weighted LSSVM is introduced to increase the statistical efficiency of LSSVM-LTS estimator. In Section 4, two groups of examples including numerical tests and real-world examples are respectively adopted to analyze the robustness of our newly developed method. Conclusions are given in Section 5.

2. Least squares support vector regression and weighted least squares support vector regression

Given a training set of n samples $\{\boldsymbol{x}_i, y_i\}_{i=1}^{n}$ with input data $\boldsymbol{x}_i \in R^d$ and output value $y_i \in R$, the objective function of LSSVM is expressed as follows:

$$\min_{\boldsymbol{w},b,e} F(\boldsymbol{w}, b, e) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \frac{\gamma}{2}\sum_{i=1}^{n} e_i^2 \tag{1}$$

subject to

$$y_i - \left[\boldsymbol{w}^T\varphi(\boldsymbol{x}_i) + b\right] = e_i, i = 1, 2, \ldots, n \tag{2}$$

where γ is a regularized parameter, controlling the tradeoff between the training error minimization and smoothness of the estimated function; $\boldsymbol{w}$ is the normal of the hyperplane; e_i is the error of the ith sample points; $\varphi(\boldsymbol{x})$ is a nonlinear function that maps $\boldsymbol{x}$ to a high-dimensional feature space and b is a bias term.

In terms of Lagrangian function, the objective function of LSSVM can be transformed into the system of linear equations as follows:

$$\begin{pmatrix} 0 & \boldsymbol{1}^T \\ \boldsymbol{1} & \boldsymbol{K} + \frac{\boldsymbol{I}}{\gamma} \end{pmatrix} \begin{pmatrix} b \\ \boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} 0 \\ \boldsymbol{y} \end{pmatrix} \tag{3}$$

where $\boldsymbol{y} = \begin{pmatrix} y_1 & \cdots & y_n \end{pmatrix}^T$; $1 = \begin{pmatrix} 1 & \cdots & 1 \end{pmatrix}^T$; $\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 & \cdots & \alpha_n \end{pmatrix}^T$; $K_{ij} = \varphi(\boldsymbol{x}_i)^T\varphi(\boldsymbol{x}_j) = K(\boldsymbol{x}_i, \boldsymbol{x}_j)$. $\boldsymbol{K}$ is a kernel matrix. In this paper, radial basis function kernel was used, expressed as $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-||\boldsymbol{x}_i - \boldsymbol{x}_j||^2/2\sigma^2\right)$, where σ is the kernel width parameter.

The objective function of the weighted LSSVM is as follows:

$$\min_{\boldsymbol{w},b,e} F(\boldsymbol{w}, b, e) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \frac{\gamma}{2}\sum_{i=1}^{n} t_i e_i^2 \tag{4}$$

subject to

$$y_i - \left[\boldsymbol{w}^T\varphi(\boldsymbol{x}_i) + b\right] = e_i, i = 1, 2, \ldots, n \tag{5}$$

where t_i is the weight of the ith sample point. It can be set as follows [21]:

$$t_i = \begin{cases} 1 & |e_i/\hat{s}| \le c_1 \\ \frac{c_2 - |e_i/\hat{s}|}{c_2 - c_1} & c_1 < |e_i/\hat{s}| \le c_2 \\ 10^{-4} & otherwise \end{cases} \tag{6}$$

where $\hat{s}$ is a robust scale estimator; c_1 and c_2 are tuning constants

The system of the weighted LSSVM linear equations is as follows:

$$\begin{pmatrix} 0 & \boldsymbol{1}^T \\ \boldsymbol{1} & \boldsymbol{K} + \frac{1}{\gamma} diag\begin{pmatrix} \frac{1}{t_1} & \cdots & \frac{1}{t_n} \end{pmatrix} \end{pmatrix} \begin{pmatrix} b \\ \boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} 0 \\ \boldsymbol{y} \end{pmatrix} \tag{7}$$

Comparing Eq. (3) with Eq. (7), it can be found that when we take $t = 1$, the weighted LSSVM is same to the classical LSSVM.

After obtaining $\boldsymbol{\alpha}$ and b by solving Eqs. (3) or (7), we finally get the following LSSVM model for function estimation,

$$f(\boldsymbol{x}) = \sum_{i=1}^{n} K(\boldsymbol{x}, \boldsymbol{x}_i)\alpha_i + b \tag{8}$$

In principle, the classical LSSVM is optimal only when the errors follow a Gaussian distribution. Thus, in the case of outliers or non-Gaussian distributions with heavy tails on data, the weighted LSSVM can be employed to further improve the simulation results of the classical LSSVM.

3. A weighted LSSVM based on LSSVM-LTS

3.1. Least trimmed squares based LSSVM (LSSVM-LTS)

LTS regression corresponds to finding the subset of h observations whose least squares fit produces the smallest sum of squared residuals. Motivated by this idea, we define the objective function