Chemometrics and Intelligent Laboratory Systems 126 (2013) 11–20

Contents lists available at SciVerse ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Model NOx emissions by least squares support vector machine with tuning based on ameliorated teaching–learning-based optimization

Guoqiang Li [a,b], Peifeng Niu [a,b,*], Weiping Zhang [a,b], Yongchao Liu [a,b]

[a] Key Lab of Industrial Computer Control Engineering of Hebei Province, Yanshan University, Qinhuangdao 066004, China
[b] National Engineering Research Center for Equipment and Technology of Cold Strip Rolling, Qinhuangdao 066004, China

ARTICLE INFO

Article history:
Received 14 November 2012
Received in revised form 19 April 2013
Accepted 24 April 2013
Available online 2 May 2013

Keywords:
Teaching–learning-based optimization
Least squares support vector machine
NOx emissions
Coal-fired boiler

ABSTRACT

The teaching–learning-based optimization (TLBO) is a new efficient optimization algorithm. To improve the solution quality and to quicken the convergence speed and running time of TLBO, this paper proposes an ameliorated TLBO called A-TLBO and test it by classical numerical function optimizations. Compared with other several optimization methods, A-TLBO shows better search performance. In addition, the A-TLBO is adopted to adjust the hyper-parameters of least squares support vector machine (LS-SVM) in order to build NOx emissions model of a 330MW coal-fired boiler and obtain a well-generalized model. Experimental results show that the tuned LS-SVM model by A-TLBO has well regression precision and generalization ability.

1. Introduction

With the increase of energy consumption worldwide and improved awareness of environmental protection, boiler combustion optimization problem of power plants attracts the attention of technical staffs and managers. The boiler combustion optimization technology is used to ensure boiler efficiency, and simultaneously to reduce pollutant emissions, where the NOx emissions are the main components. So the core task is to cut down NOx emissions. However, the first work of controlling NOx emissions is to set up a high precision prediction model. So building an accurate system model is very important for monitoring and optimizing the operations of power plants. In the past ten years, many research works on how to model and forecast the NOx emissions of high capacity coal-fired boilers have been published [1–5]. However, the traditional statistical analysis and forecasting methods are always based on large sample data, and many of the prediction methods, such as artificial neural networks, have theoretical assurance all just under a large sample. Due to various limits in actual circumstances, it is very difficult to gather large sample data. For this problem, the least squares support vector machine (LS-SVM) [6], which is suited for small sample data, is adopted to model and predict NOx emissions.

The LS-SVM is a reformulation to the standard support vector machine (SVM) [7–9], which simplifies the standard SVM model in a great extent by applying linear least squares criteria to the loss function to replace traditional quadratic programming method. The simplicity and inherited advantages of SVM such as excellent generalization ability and a unique solution promote the application of LS-SVM in many pattern recognition and regression problems.

The regression accuracy and generalization ability of LS-SVM are extremely dependent on two hyper-parameters: the regularization parameter γ and the kernel parameter σ^2. So choosing appropriate hyper-parameters is very important for obtaining excellent generalization ability. Parameter choosing of the LS-SVM model could be thought as an essential optimization task. This calls for the use of advanced meta-heuristic approaches, such as evolutionary or population-based methods.

The teaching–learning-based optimization algorithm [10,11] is a new and efficient meta-heuristic optimization method based on the philosophy of teaching and learning, which is proposed by Rao et al. Like other population-based optimization techniques such as particle swarm optimization (PSO) [12], evolutionary optimization (DE) [13,14], artificial bee colony (ABC) [15–19], Gravitational Search Algorithm (GSA) [20], and Coupled Simulated Annealing (CSA) [21], the TLBO is also a population-based optimization method and adopts a population of solutions to proceed to the global solution. In some researches [22–24], the performance of TLBO has already been compared with other search optimization techniques such as genetic algorithm (GA) [25,26], Bee algorithm (BA) [27], and grenade explosion method (GEM) [28]. In addition, the TLBO has been applied to some complex computational problems, such as data clustering, mechanical design, electrochemical discharge machining, and design of planar steel frames.

In this paper, in order to improve the solution quality and to quicken the convergence speed of TLBO, an ameliorated teaching–learning-based optimization algorithm called A-TLBO is proposed. In

* Corresponding author at: Key Lab of Industrial Computer Control Engineering of Hebei Province, Yanshan University, Qinhuangdao 066004, China. Tel.: +86 13230355970; fax: +86 335 8072979.
E-mail address: niupeifeng2011@163.com (P. Niu).

0169-7439/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.chemolab.2013.04.012

A-TLBO, there are three major differences: the greedy selection mechanism is not adopted but the elitist strategy, an inertia weight function and an acceleration coefficient function are introduced to quicken the processes of 'Teaching' and 'Learning'. In order to test the validity of the proposed method, it is adopted to optimize many classical numerical optimization functions and compared with other methods. Experiment results show that the A-TLBO could find better solutions and have much faster convergence speed. In addition, the A-TLBO is also used to adjust two hyper-parameters of LS-SVM in order to obtain a well-generalized model of NOx emissions for a 330MW coal-fired boiler. Results show that the tuned LS-SVM model by A-TLBO has well regression precision and generalization ability.

The rest of the paper is arranged as follows. In the next section, a brief literature review is presented. The ameliorated teaching–learning-based optimization is proposed in Section 3. In Section 4, the A-TLBO is applied to optimize some classical numerical optimization functions and compared with GSA, ABC and TLBO. In Section 5, the A-TLBO is employed to adjust the hyper-parameters of LS-SVM to model NOx emissions of a 330MW coal-fired boiler. Finally, Section 6 concludes the paper.

2. Review of related works

2.1. Teaching–learning-based optimization

The teaching–learning-based optimization (TLBO) algorithm proposed by Rao is inspired by the effect of the influence of a teacher on the output of learners in a class. In TLBO, there are two vital components, 'Teacher phase' and 'Learner phase', which indicate two different learning modes.

2.1.1. Teacher phase

In this phase, Learners learn from a teacher, who is considered as the most knowledgeable person in the society and would bring learners up to his or her level in terms of knowledge. That is to say, the teacher would put effort to move the mean of a class up to his or her level depending on his or her ability. Suppose, in the ith iteration, M_i is the mean of marks obtained by learners in a class, and T_i is the mark of the teacher. And the best learner could be mimicked as the teacher, namely:

$$T_i = X_{min\ f(X)}. \tag{1}$$

The teacher would put effort to move the mean value M_i towards itself. Namely, the new mean M_{new} will be T_i.

The learners would learn and update their knowledge according to the following form:

$$X_{new,i} = X_{old,i} + r_i(M_{new} - T_F M_i) \tag{2}$$

where $X_{old,i}$ and $X_{new,i}$ respectively denote the ith learner's mark before or after learning from the teacher; r_i is a random number between 0 and 1; and T_F is a teaching factor which controls the movement of the mean value. The value of T_F is either 1 or 2, which is a heuristic step and decided randomly by:

$$T_F = round[1 + rand(0,1)(2-1)]. \tag{3}$$

2.1.2. Learner phase

In this phase, the learners could increase their knowledge by interactions among themselves. A learner could communicate randomly with other learners in order to improve his or her knowledge with the help of group discussions, formal communications, presentations, etc. A learner learns something new from the learners who have more knowledge than him or her. The learning process could be described as follows:

In the ith iteration, for a learner X_i, randomly select another learner X_j where $i \neq j$

$$X_{new,i} = \begin{cases} X_{old,i} + r_i\left(X_i - X_j\right) & if\ f(X_i) \le f\left(X_j\right) \\ X_{old,i} + r_i\left(X_j - X_i\right) & if\ f(X_i) > f\left(X_j\right) \end{cases}. \tag{4}$$

Accept $X_{new,i}$ if it has better function value.

2.2. Least squares support vector machine

The support vector machine (SVM) is a machine learning algorithm based on the statistical learning theory [29]. The structural risk minimization (SRM) adopted by SVM could provide better generational ability than the empirical risk minimization used by traditional methods.

The ε-SVM is a commonly used SVM model. Suppose, there are a given set of samples $\{(x_1,y_1),(x_2,y_2),\ldots,(x_l,y_l)\} \subset R^n \times R$, where x_l is the lth input vector and y_l the corresponding objective value. For ε-SVM, the optimization problem can be described as:

$$\begin{aligned} \min \quad & \Psi(\omega,\xi,\xi^*) = \frac{1}{2}\omega^T\omega + \gamma\sum_{i=1}^{l}(\xi_i + \xi_i^*) \\ \text{subject to}: & \begin{cases} y_i - \left(w^T\varphi(x_i) + b\right) \le \varepsilon + \xi_i \\ \left(w^T\varphi(x_i) + b\right) - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0, \quad i = 1,2,\ldots,l \end{cases} \end{aligned} \tag{5}$$

where γ is the regularization parameter, and ξ_i and ξ_i^* are the slack variables.

Compared with SVM, the least squares support vector machine (LS-SVM) applies linear least squares criteria to the loss function to replace the inequality constraints with equality constraints. So the optimization problem can be defined as:

$$\begin{aligned} \min \quad & \Psi(\omega,e) = \frac{1}{2}\omega^T\omega + \frac{1}{2}\gamma\sum_{i=1}^{l}e_i^2 \\ \text{subject to}: & \ y_i = w^T\varphi(x_i) + b + e_i, i = 1,2,\ldots,l. \end{aligned} \tag{6}$$

The Lagrangian function can be set up by:

$$L(\omega,b,e,\alpha) = \Psi(\omega,e) - \sum_{i=1}^{l}\alpha_i\left(\omega^T\varphi(x_i) + b + e_i - y_i\right) \tag{7}$$

where α_i is the ith Lagrange multiplier.

The Karush–Kuhn–Tucker (KKT) conditions for optimality are described as follows:

$$\begin{cases} \dfrac{\partial L}{\partial \omega} = 0 \Rightarrow \omega = \sum_{i=1}^{l}\alpha_i\varphi(x_i) \\ \dfrac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{l}\alpha_i = 0 \\ \dfrac{\partial L}{\partial e_i} = 0 \Rightarrow \alpha_i = Ce_i \\ \dfrac{\partial L}{\partial \alpha_i} = 0 \Rightarrow \omega^T\varphi(x_i) + b + e_i = y_i. \end{cases} \tag{8}$$

After eliminating variables ω and e_i, the optimization problem could be transformed into the following form:

$$\begin{bmatrix} \Omega + \gamma^{-1}I_l & \vec{1}_l \\ \vec{1}_l^T & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} y \\ 0 \end{bmatrix} \tag{9}$$