Expert Systems With Applications 91 (2018) 63–77

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Parameter optimization of support vector regression based on sine cosine algorithm

Sai Li[a], Huajing Fang[a,*], Xiaoyong Liu[b]

[a] School of Automation, National Key Laboratory of Science and Technology on Multispectral Information Processing, Huazhong University of Science and Technology, 1037 Luoyu Road, Wuhan 430074, PR China
[b] College of Engineering and Technology, Zunyi Normal College, 830 Shanghai Road, Zunyi 563003, PR China

ARTICLE INFO

Article history:
Received 22 May 2017
Revised 5 August 2017
Accepted 19 August 2017
Available online 24 August 2017

Keywords:
Support vector regression
Sine cosine algorithm
Time series prediction
Parameter optimization

ABSTRACT

Time series prediction is an important part of data-driven based prognostics which are mainly based on the massive sensory data with less requirement of knowing inherent system failure mechanisms. Support Vector Regression (SVR) has achieved good performance in forecasting problems of small samples and high dimensions. However, the SVR parameters have a significant influence on forecasting performance of SVR. In our current work, a novel SCA-SVR model has been presented where sine cosine algorithm (SCA) is used to select the penalty and kernel parameters in SVR, so that the generalization performance on unknown data can be improved. To validate the proposed model, the results of the SCA-SVR algorithm were compared with those of grid search and some other meta-heuristics optimization algorithms on common used benchmark datasets. The experimental results proved that the proposed model is capable to find the optimal values of the SVR parameters and can yield promising results.

1. Introduction

Condition-based maintenance is a decision-making strategy to enable real-time diagnosis of impending failures and prognosis of future system health. A recent approach to condition-based maintenance is the data-driven method using time series forecasting to capture the connection between measurements variables and remaining useful life of a system. Owing to the non-linearity and non-stationarity of most real-world dynamic behaviors, effective prediction of future states of a complex system from time series remains a challenge.

In real life, due to the limited effectiveness of classic easy-to-implement linear techniques, advanced nonlinear time series prediction algorithms such as neural networks and kernel-based methods in machine learning communities are introduced gradually. Considering that neural networks are risky of over-fitting, subject to the model architecture, and vulnerable to get trapped in local minima, Support Vector Machines originally introduced by Vapnik (2000), which could theoretically guarantee to achieve the global optimum by implementing the structural risk minimization principle instead of the empirical risk minimization principle used in neural networks, have emerged as an important learning technique for solving classification and regression problems in various fields. In regression field, experimental results (Mohandes, 2002; Thissen, van Brakel, de Weijer, Melssen, & Buydens, 2003) have revealed the superior generalization performance of Support Vector Regression (SVR) with respect to other non-linear techniques including neural-network. However, poor choice of penalty parameter and kernel parameters can dramatically decrease the performance of SVR.

Meta-heuristics algorithms divided into individual-based and population-based algorithms have achieved competitive results when solving optimization problems including parameter tuning problem (Gogna & Tayal, 2013). Sine Cosine Algorithm (SCA), a novel population-based optimization algorithm proposed by Mirjalili (2016), combines the random solutions in the set of solutions with a high rate of randomness to find the promising regions of the search space in the exploration phase. Moreover, there are gradual changes in the random solutions, and random variations are considerably less than those in the exploration phase. The algorithm is also featured by a small number of tuning parameters, fast convergence speed, high optimization accuracy, and strong global search ability. The performances of SCA have been well tested on different benchmark functions including unimodal, multi-modal, and composite functions in Mirjalili (2016). The results show that it outperforms other well-known optimization methods such as PSO, GA, BA, FA, FPA, and GSA.

* Corresponding author.
E-mail addresses: sli@hust.edu.cn (S. Li), hjfang@mail.hust.edu.cn (H. Fang), liuxy204@163.com (X. Liu).

http://dx.doi.org/10.1016/j.eswa.2017.08.038
0957-4174/

To improve the optimization performance, SCA with parabolic function and exponential function of conversion parameter were proposed in Liu and Ma (2017). Kommadath, Dondeti, and Kotecha (2017) showed that SCA quickly converges to a final solution, but is unable to converge to the global optima. This provides the possibility of hybridizing SCA with techniques that are efficient at exploring the search space. Bureerat and Pholdee (2017) presented an adaptive SCA integrated with differential evolution to further improve its performance. The proposed ASCA-DE algorithm is a more reliable method in the simulation experiments compared with Differential evolution (DE), Artificial bee colony algorithm (ABC), Real-code ant colony optimization (ACOR), Charged system search (ChSS), League championship algorithm (LCA), Simulated annealing (SA), Particle swarm optimization (PSO), Evolution strategies (ES), Teaching-learning-based optimization (TLBO), Adaptive differential evolution (JADE), Evolution strategy with covariance matrix adaptation (CMAES), and Sine Cosine Algorithm (SCA). Turgut (2017) proposed a hybrid optimization algorithm based on the combination of the merits of the back-tracking search (BSA) and sine cosine algorithm (SCA) to achieve the optimal design of a shell and tube evaporator. Considered as black boxes, the SCA algorithm has been successfully applied to many optimization problems such as modeling optimization (Mahdad & Srairi, 2017; Sahlol, Ewees, Hemdan, & Hassanien, 2016; Zhang, 2016), path selection (Cui, 2016), scheduling policy (Das, Bhattacharya, & Chakraborty, 2017), feature selection (Hafez, Zawbaa, Emary, & Hassanien, 2016; Sindhu, Ngadiran, Yacob, Zahri, & Hariharan, 2017), manuscript image binarization (Elfattah, Abuelenin, Hassanien, & Pan, 2016), and data clustering (Kumar & Kumar, 2017), and it shows a great potential for parameter optimization of SVR.

The main objective of this paper is to present a novel SCA-SVR model for parameters optimization of SVR with SCA, which improves the regression performance. To make an overall comparison, Flower Pollination Algorithm (FPA), Social-Spider Optimization (SSO), Grey Wolf Optimizer (GWO), Ant Lion Optimizer (ALO), and Multi-Verse Optimizer (MVO) are also extended to SVR parameter tuning.

The remainder of the paper is organized as follows: Section 2 presents the related work of SVR parameter optimization. Section 3 presents the proposed SCA-SVR forecasting model, including the presentation of support vector regression and sine cosine algorithm. In Section 4, the established model is benchmarked with other models on several datasets and the experimental results are discussed. Conclusions and future work are provided in Section 5.

2. Related work

There are many studies on SVR parameters tuning. Based on the views in the literature (Zhao, Tao, & Zio, 2015), existing methods of SVR parameter selection can be divided into trial-and-error method based on prior knowledge and optimization-based algorithms which search for the optimal values within an optimization scheme. Based on this point, there are three types of optimization-based methods. First is the traditional exhaustive search algorithms which are typical of grid search (GS) method (Hsu, Chang, & Lin, 2003). The grid search method discretizes the parameters search space with a fixed step-size and minimizes some performance metric typically measured by cross-validation error in the parameter search space, resulting in a time-consuming training process. The second type comprises the traditional numerical optimization methods represented by the gradient descent method (Chapelle, Vapnik, Bousquet, & Mukherjee, 2002; Keerthi, Sindhwani, & Chapelle, 2007). By minimizing some generalization bounds such as the leave-one-out error bounds, gradient-based numerical optimization methods are generally more efficient than GS. However, owing to the non-convexity of the generalization bounds, these methods may get stuck into local optimum, and require differential condition on the kernel functions and the bounds of generalization error. Moreover, the quality of results is largely dependent on the setting of starting points. Naturally, without the inefficient search strategies and long searching time in heuristics algorithms, meta-heuristics have been introduced as problem-independent technique to obtain an acceptable optimum in a wide range of problems (Blum & Roli, 2003; Gogna & Tayal, 2013; Talbi, 2009). As stated in No Free Lunch theorem (Igel, 2013), no one can propose an algorithm for solving all optimization problems. Thus many general-purpose meta-heuristics methods have been proposed to solve parameter optimization problem, ranging from improvement of the current techniques and new algorithms to hybrid algorithms. For example, Genetic Algorithms (GA) (Liang & Liu, 2002; Liu & Wang, 2016; Min, Lee, & Han, 2006; Shokri, Sadeghi, & Marvast, 2014; Wu, Tzeng, & Lin, 2009), Particle Swarm Optimization (PSO) (Barati & Sharifian, 2015; Li-xia, Yi-qi, & yong Liu, 2011; Wu, 2010), Artificial Bee Colony (ABC) (Hong, 2011; Kang & Li, 2015), Bat Algorithm (BA) (Tavakkoli, Rezaeenour, & Hadavandi, 2015; Tharwat, Hassanien, & Elnaghi, 2016), Krill Herd Algorithm (KHA) (Stasinakis, Sermpinis, Psaradellis, & Verousis, 2016), Moth-Flame Optimization (MFO) (Li, Li, & Liu, 2016), Grey Wolf Optimizer (GWO) (Mustaffa, Sulaiman, & Kahar, 2015) have been proposed for the selection fo SVR optimal parameters. However, meta-heuristics like Ant Colony Optimization(ACO) (Zhang, Chen, & He, 2010), Flower Pollination Algorithm (FPA) (Hoang, Bui, & Liao, 2016), Social-Spider Optimization (SSO) (Pereira et al., 2016), Ant Lion Optimizer (ALO) (Zhao, Gao, Yu, & Tu, 2016), and Multi-Verse Optimizer (MVO) (Faris, Hassonah, Al-Zoubi, Mirjalili, & Aljarah, 2017) have only been applied to parameter selection of support vector classification (SVC). Although various methods in the literature have developed over the past decades, there is no systematic well-established methodology for parameters tuning of SVR.

3. Methodology of SCA-SVR model

3.1. Support vector regression

Support vector regression, which evolved from the support vector classification for doing regression tasks by introduction of the ε-insensitive loss function, is a data-driven machine learning methodology. The detailed explanation and proofs of support vector machines can be contained in the book (Vapnik, 2000). Assume that there is a learning samples set $D = \{(x_i, y_i)\}$ where $x_i \in R$ represent the input values and $y_i \in R$ are the corresponding output values for $i = 1, 2, \ldots, N$ where N is the number of the samples. In SVR, the goal has been to find a functional relationship $f(x)$ between input data x_i and output data y_i under the assumption that the joint distribution P of (x, y) is completely unknown. The function in the linear case follows $f(x) = \langle w, x\rangle + b$, where the terms w and b represent weight coefficient and constant coefficient. In the nonlinear case, a nonlinear mapping Φ is introduced to transform the complex nonlinear problem into a simple linear problem. The regression function after this transformation reads in Eq. (1).

$$f(x) = \langle w, \Phi(x)\rangle + b \tag{1}$$

The function $f(x)$ should loosely fit the training data and be as flat as possible to avoid over-fitting problem by minimizing the norm value of w. To cope with otherwise infeasible constraints, two slack variables ξ_i and ξ_i^* are introduced. This convex optimization problem is feasible based on the assumption that a function exists which approximates every data pair (x_i, y_i) with an acceptable ε accuracy. Hence, the entire problem can be formulated as