Chemical Engineering Science 172 (2017) 571–584

Contents lists available at ScienceDirect

Chemical Engineering Science

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

Adaptive just-in-time and relevant vector machine based soft-sensors with adaptive differential evolution algorithms for parameter optimization

Yiqi Liu

School of Automation Science & Engineering, South China University of Technology, Wushang Road, Guang Zhou 510640, China

HIGHLIGHTS

- A Just-in-time (JIT) and Relevant-vector-machine (RVM) to perform perditions.
- Nature-inspired optimized algorithm that ensures optimal parameters selection for JIT and RVM.
- A JADE evolution algorithm to optimize parameters for JIT and RVM without hyper-parameter setting.
- A moving window methodology for improvement of the JIT and RVM model.
- Proposed method is efficiently attractive in a wastewater plant monitoring.

ARTICLE INFO

Article history:
Received 6 December 2016
Received in revised form 28 April 2017
Accepted 3 July 2017
Available online 12 July 2017

Keywords:
Soft-sensors
Just-in-time
Relevant vector machine
Differential evolution
Parameters

ABSTRACT

Just-in-time (JIT) and Relevant vector machine (RVM) are two of commonly used models for soft-sensors modeling, the efficiency of which is governed by few critical parameters and hyper-parameters significantly. These parameters are routinely selected by trial and error or experience, thus leading to over- or under-fitting for the prediction. Adaptive differential evolution with optional external archive (JADE) has been used to optimize the parameters of JIT and RVM in this paper. The resulted JADE-JIT and JADE-RVM based soft-sensors are further enhanced into an adaptive format by the moving window (WM) technique. The proposed methodologies are applied to prediction of hard-to-measured variables in the wastewater treatment plants (WWTPs) and successful results are obtained.

1. Introduction

Soft-sensors are widely used to estimate variables that are difficult to measure online due to technical difficulty, large measurement delays, high investment cost, and so on (Kadlec et al., 2009; Kano and Nakagawa, 2008; Liu et al., 2013). To describe the relationship between easy-to-measure variables and hard-to-measure responses, machine learning methods including Neural networks (Wold et al., 2001), Relevant vector machine (RVM) (Salahshoor and Komari Alaei, 2010), Bayesian networks (Cai et al., 2014, 2016) and Support vector regression (SVR) (Yan et al., 2004) are typically researched as the soft sensor models.

One of the main bottlenecks limiting widely use of soft-sensors modeling is the proper calibration of their parameters. Parameter calibration is essentially an optimization problem, which requires higher-level optimization methods to tackle. A recent study showed that a framework for self-tuning algorithms can be indeed established with promising results (Fister et al., 2013). The task of fine-tuning parameters in machine learning aims at finding suitable values for those parameters in order to maximize some fitness function, such as prediction accuracy, when dealing with supervised problems. However, some parameters, such as regularization item and kernel function in SVMs, concerning model structure configuration play an important role to the algorithm performance. Typically, this is performed by crossing validation or trial and error manually, which is laborious, particularly, as the dimension of parameters increases significantly. Also, this could leave many users to select algorithms based on reputation or intuitive appeal, and/or to leave parameters set to default values. This suggests a natural challenge for machine learning: given a dataset, automatically and simultaneously choosing a learning algorithm and setting its parameters to optimize empirical performance. The existing techniques for adjusting the parameters can be summarized into

E-mail address: aulyq@scut.edu.cn

http://dx.doi.org/10.1016/j.ces.2017.07.006
0009-2509/

two kinds: one is based on analytical techniques; the other is based on heuristic searches. The first kind of techniques determines the parameters with gradients of some generalized error measures (SS, 2002). And the second kind of techniques determines the parameters with modern heuristic algorithms including Genetic Algorithms (GA), simulated annealing algorithms and other evolutionary strategies. Iterative gradient-based algorithms rely on smoothed approximations of a function. So, it does not ensure that the search direction points exactly to an optimum of the generalization performance measure which is often discontinuous. Also, this procedure needs to locate the interval of feasible solution and a suitable sampling step. Moreover, when there are more than two parameters, the manual model selection may become intractable. On the other hand, nature-inspired optimized methodologies are among the most used ones for optimization problems, since they provide simple and elegant solutions in a wide range of applications. Since the early 1970s, various nature-inspired optimization algorithms have emerged starting with the Genetic Algorithm (GA) (Yang, 2015). The motive behind the GA is to mimic the nature to obtain suitable working solutions by using computational techniques, while create future generations in biological organisms. During optimized stage, nature inspired operators, like crossover and reproduction, are used to identify good working solutions. GA often fails to address very complex high dimensional, multi-modal problems where fitness function evaluation becomes computationally very complex. Along with the GA, Particle Swarm Optimization (PSO) (Guo et al., 2008) is based on the collective group behavior of organisms, such as fish schooling, insect swarming or birds flocking, whereby the group attempts to meet the collective objective of the group based on the feedback from the other members Particle swarm optimization is used for problems where the function to be optimized is discontinuous, non-differentiable with too many non-linearly related parameters. Wind Driven Optimization (WDO) technique is a population based iterative heuristic global optimization algorithm for multi-dimensional and multi-modal problems with the potential to implement constraints on the search domain. Compared to similar particle based algorithms, WDO employs additional terms in the velocity update equation (e.g., gravitation and Coriolis forces), providing robustness and extra degrees of freedom to tune properly. In aforementioned nature-inspired algorithms, the best solutions explored in the history are used to direct the movement of the current population, thus diminishing population diversity significantly. Instead of the best solutions previously explored, a new proposed algorithm, JADE, is able to explore a set of recently inferior solutions and consider their difference from the current population as a promising direction toward the optimum (Zhang and Sanderson, 2009). In spite of greedy property of JADE, this strategy is able to diversify the population so that the problems such as premature convergence can be alleviated.

Due to the highly varying operating conditions, Moving window (MV) method is commonly used to collect the most recent and fairly long-term data for soft-sensor modeling during each step when new data points are coming (Dayal and MacGregor, 1997; Shao et al., 2015). During the adaption, parameters can be tuned properly for the local model at each window rolling through nature-inspired optimized algorithms. However, standard nature-inspired optimized algorithms, such as GA, PSO, and WDO, have to tune hyper-parameters for itself. To make sure full-scale parameter adaptation for adaptive soft-sensors, nature-inspired optimized algorithms should be able to have self-adaptive hyper-parameter control. According to the classification scheme introduced by Angeline (1995) and Eiben et al. (1999), three classes of hyper-parameter control mechanisms are defined, deterministic hyper-parameter control, adaptive hyper-parameter control and self-adaptive hyper-parameter control. Deterministic hyper-parameter control: The control hyper-parameter is altered by some deterministic rules without taking into account any feedback from the evolutionary search. On the contrary, adaptive hyper-parameter control take full use of feedback from the evolutionary search to change the control hyper-parameters dynamically (Valdez et al., 2014). Self-adaptive hyper-parameter control is a method of "the evolution of evolution" aiming to conduct the self-adaptation of control hyper-parameters (Abbass, 2002). Adaptive or self-adaptive hyper-parameter control, if well designed, can enhance the robustness of an algorithm by dynamically adapting the hyper-parameters to the characteristic of different fitness landscapes. It is thus applicable to various optimization problems without trial and error. In addition, the convergence rate can be improved if the control hyper-parameters are adapted to appropriate values at different evolution stages of a specific problem. JADE is new proposed nature-inspired algorithm with self-adaptive hyper-parameter control. Due to the self-adaptive hyper-parameter control of JADE, it can achieve faster and more reliable convergence performance than the classic algorithms without hyper-parameter control for many benchmark problems. The high efficiency of JADE also makes it possible to apply them to a wide range of problems in diverse applications. The reliability of the algorithm is further improved by the adaptive hyper-parameter control (Zhang and Sanderson, 2009). In view of the above considerations, we introduce JADE for hyper-parameters optimization for Moving window during soft-sensor model adaptation.

Both of RVM and Just-in-time (JIT) are two powerful and typical machine learning methods for prediction (Liu et al., 2011, 2014). However, the existence of few non-optimized parameters always compromises their performance significantly ('Kernel' and 'width' for RVM, 'Maximum number of neighbors' and 'Combination parameters' for JIT). This paper intends to introduce four nature-inspired optimized algorithm, GA, PSO, WDO and JADE, to optimize the non-optimized parameters of the RVM and JIT firstly. Due to assimilation of nature-inspired optimized algorithm, tedious cross-validation for parameter control can be avoided properly and optimal parameters can be obtained. Considering the highly varying operating conditions, the JIT and RVM are enhanced by the WM technique to build adaptive soft-sensors. Additionally, to deal with the problem of parameters optimization online, a hyper-parameter-free algorithm, JADE, is proposed to search for the best parameters for JIT and RVM on-line, thus resulting in the variable structure adaptive soft-sensors secondly.

In Section 2 the preliminary knowledge of JIT model, RVM model and JADE algorithm is introduced. Section 3 proposes JADE to derive the optimal parameters of JIT and RVM. The resulted JIT and RVM models are further improved by WM to be able to adapt to the dynamic processes. The proposed soft-sensors are validated through the data sets from a real wastewater treatment plants (WWTPs) in Section 4. Finally Section 5 concludes.

2. Predicted models and adaptive differential evolution with optional external archive (JADE)

2.1. Just-in-time model

A global linear model does not function well when a process has strong nonlinearity in its operation range. Division of a process operation region into small multiple regions and configuration of a local model in each small region provide an alternative to deal with this problem. JIT learning, also called Lazy learning, is a local learning technique which postpones all the computation until an explicit request for a prediction is received. The request is fulfilled by interpolating locally the samples considered relevant according to a distance measure. Each prediction requires therefore a local