Control Engineering Practice 72 (2018) 68–82

Contents lists available at ScienceDirect

Control Engineering Practice

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

Variable selection for nonlinear soft sensor development with enhanced Binary Differential Evolution algorithm

Le Yao, Zhiqiang Ge *

State Key Laboratory of Industrial Control Technology, Institute of Industrial Process Control, College of Control Science and Engineering, Zhejiang University, Hangzhou, 310027, PR China

ARTICLE INFO

Keywords:
Variable selection
Binary Differential Evolution algorithm
Parallel calculation
Boosting algorithm
Nonlinear process
Soft sensor

ABSTRACT

In this paper, two enhanced Binary Differential Evolution (BDE) algorithms are proposed to select variables for nonlinear process soft sensor development. Firstly, the Parallel BDE (PBDE) algorithm is presented to extract the optimal individuals of several parallel short evolution paths of basic BDE, where the spurious variables are effectively eliminated. And the most relevant variables are selected through a double-layer selection strategy with the validating Root Mean Square Error (RMSE) for evaluating criterion. Secondly, the Boosting BDE (BBDE) algorithm is proposed through applying the boosting technique to the parallel evolution paths. The performance of the previous path needs to be taken into account when conducting the current evolution path. The selected probabilities of variables are given through the weighted summation of the selection results of all paths. Also, a double-layer selection is conducted on BBDE algorithm. The feasibility and effectiveness of the proposed methods are demonstrated through a nonlinear numerical example and a real industrial process.

1. Introduction

In industrial processes, the operating condition of plants and process variables such as melt index, catalyst activity, and concentration of products should be strictly monitored (Ge, Song, & Gao, 2013; Kruger & Xie, 2012; Ge, Song, Ding, & Huang, 2017; Ge, 2017). However, measuring all variables online is difficult due to technological limitations, large signal delay and the high measuring costs (Fujiwara, Kano, & Hasebe, 2012a; Kadlec, Gabrys, & Strandt, 2009). Thus, soft sensors have been widely used to estimate the difficult-to-measure variables with the help of acquired data from easy-to-measure variables (Domlan, Huang, Xu, & et al., 2011; Yao & Ge, 2017). With the low-cost data-driven soft sensors, the value of quality variables can be predicted with a high degree of accuracy in real-time. Moreover, they can give useful information for fault detection by working together with hardware sensors (Kaneko, Arakawa, & Funatsu, 2011).

For data-driven soft sensors, classical linear regression approaches such as Principal Component Regression (PCR) (Ge, 2014; Ge, Huang, & Song, 2014), Partial Least Square (PLS) (Facco, Doplicher, Bezzo, & et al., 2009; Galicia, He, & Wang, 2012; Zheng, Ge, & Song 2016) have been applied to the plant data to produce useful models. Since the relationship between process variables and the quality variables are often nonlinear, those linear methods could not perform appropriately in nonlinear process modeling. Therefore, popular nonlinear methods based on Support Vector Machine (SVM) (Kaneko & Funatsu, 2014; Shang, Gao, Yang, & Huang, 2014) and Artificial Neural Network (ANN) (Gao & Ren, 2010; Gonzaga, Meleiro, Kiang, & Filho, 2009) have been widely utilized in soft sensor development. Besides, Gaussian Process Regression (GPR) (Ge, Chen, & Song, 2011; Grbic, Sliškovic, & Kadlec, 2013; Yu, 2012) has gained much attention recently, which provides not only predicted distribution of quality variable, but also its level of confidence. This makes GPR much more reasonable and acceptable to develop data-driven soft sensors for nonlinear industrial processes.

Nowadays, with the development of the measuring technique in industrial process, more and more online instruments have been implemented to monitor the state of the production unit (Kruger, Kumar, & Littler, 2007; Souza & Araújo, 2014; Zamprogna, Barolo, & Seborg, 2005). Thus, it is often the case that the number of process variables is very large for the regression model to estimate the quality variable. However, only the process variables which are most related and explanatory to the quality variable are meaningful for the soft sensor while the irrelevant process variables may decrease the accuracy and predictive ability of a regression model (Souza, Araujo, & Mendes, 2016; Wang, Jang, Wong, & et al., 2013). Therefore, variable selection becomes an

* Corresponding author.
E-mail address: gezhiqiang@zju.edu.cn (Z. Ge).

https://doi.org/10.1016/j.conengprac.2017.11.007
Received 22 July 2017; Received in revised form 19 September 2017; Accepted 18 November 2017

important technical procedure before developing a soft sensor, which could improve predictive ability, the model interpretation, and reduce the measurement cost. In fact, some efficient methods like least absolute shrinkage and selection operator (LASSO) (Fujiwara & Kano, 2015; Tibshirani, 2011; Zou, 2006), and a series of PLS-based variable selection methods (Fujiwara, Sawada, & Kano, 2012b; Mehmood, Liland, Snipen, & Sæbø, 2012; Wang, He, & Wang, 2015) for variable selection have already been developed. However, these methods fundamentally rely on the linear regression models, like MLS, PLS and so on, which would be low-efficient in selecting the most important variables for nonlinear process regression. Nevertheless, among the PLS-based variable selection methods, a wrapper variable selection method called Genetic Algorithm PLS (GA-PLS) has been concentrated on in recent years (Arakawa, Yamashita, & Funatsu, 2011; Kaneko & Funatsu, 2012; Leardi, Seasholtz, & Pell, 2002), which has been extended to the SVR model for nonlinear process (Kaneko & Funatsu, 2013). In this case, through using the nonlinear method for modeling to calculate the fitness score of GA, the nonlinearity of the process can be expressed more completely. Thus, the motivation of the proposed work is trying to find the most relevant variable set for nonlinear process regression modeling through a better evolutionary search algorithm.

As an evolutionary optimization algorithm, GA has been widely used in statistical and engineering areas (Weile & Michielssen, 1997). However, due to its complexity in operation and the risk of drowning into local optimization, several other simple and more efficient evolutionary algorithms have been proposed to avoid these drawbacks (Leardi, 2001; Tamaki, Kita, & Kobayashi, 1996). The Differential Evolution (DE) algorithm, a population-based stochastic optimizer firstly proposed by Storn and Price (Storn & Price, 1997), has become a new research hotspot in evolutionary computation. The standard DE algorithm, which is simple yet efficient in global optimization, has been successfully applied in scientific and engineering fields. It is capable of solving non-convex, nonlinear, non-differentiable and multimodal problems. Besides, there are only two control parameters in the updating formulas, which renders DE easier for implementation. Due to its simplicity, robustness and effectiveness, it has been demonstrated to be superior to Genetic Algorithm (GA) or Particle Swarm Optimization (PSO) in some real-world applications (Ponsich & Coello, 2011; Vesterstrom & Thomsen, 2004). However, the standard DE and most of its improved variants operate in the continuous space, which are not suitable for solving binary-coded combinational optimization problems. Therefore, several binary DE (BDE) algorithms are proposed to extend the applications of DE (Gong & Tuson, 2007; He & Han, 2007; Kong, Gao, Ouyang, & Ge, 2014). Among them, a most concise and efficient strategy for binarization is to utilize the power function, where the scaling factor is not required in the evolutionary process and binary mutation was directly carried out only on the basis of the differences between individuals, which has been certificated to be feasible and efficient in solving the 0–1 knapsack problem (Kong et al., 2014).

In order to address the variable selection issue for nonlinear process soft sensor development, the binary DE (BDE) algorithm is firstly combined with the nonlinear GPR model (Detailed descriptions are provided in Appendix) to search the optimal variable set. The variables selected or not are coded in a binary string to initialize a population. Then the mutation, crossover and selection operators are conducted as the standard DE with the fitting Root Mean Square Error (RMSE) for fitness function. Moreover, several bootstrap trials of the basic BDE algorithm is conducted to calculate and rank the selected probabilities of all variables. Then the variables are added to a selected variable set one at a time in the ranked sequence to build validating models with a validating dataset. Finally, the optimal variable set is selected with the lowest validating RMSE.

However, due to the randomness of initial population and the asymmetric fitness function, the final selected results of basic BDE are usually contaminated by spurious variables, which behaves like the "over-fitting" feature in the model training. To solve this problem, an "early-stopping" strategy which is widely used in neural network community to avoid over-fitting is applied (Zhang & Yu, 2005). In this case, the Parallel BDE (PBDE) algorithm is proposed through extracting the not fully convergent solution of the basic BDE with several parallel paths of calculations. Then the selected times of all variables are counted and ranked, and the validating RMSE is utilized to determine the number of foremostly ranked variables that should be included in the selected variable set. This is called an inside selection of the PBDE algorithm. Meanwhile, an outside selection for PBDE is the same as the basic BDE to conduct the algorithm for several trials. Through the double-layer selection, the performance of BDE would be greatly enhanced. Furthermore, the Boosting BDE (BBDE) algorithm is proposed through introducing the boosting method to the parallel paths of calculation since its potential to greatly improve the generation ability and the robustness of a single model (Mendes-Moreira, Soares, Jorge, & Sousa, 2012). In this algorithm, the performance of the previous path needs to be taken into account when conducting the current evolution. Through combining the results of all the paths with calculated weights, selected probabilities of all variables are obtained. Then the selected variable set is determined by an inside selection like the PBDE algorithm. Moreover, bootstrap trials are also conducted to form the outside selection of BBDE algorithm. The feasibility and effectiveness of the proposed enhanced BDE algorithms for variable selection will be demonstrated through a nonlinear numerical example and a real industrial process. While selecting variables for nonlinear soft sensor, the GPR algorithm is applied to build the models.

The layout of this paper is given as follows. In Section 2, the binary DE algorithm for variable selection is proposed. In the next section, two enhanced BDE algorithms, the PBDE and the BBDE, are proposed. Then the proposed variable selection methods are implemented in a numerical example and an industrial process in Section 4. Finally, conclusions are made.

2. Binary Differential Evolution algorithm for variable selection

2.1. Binary Differential Evolution algorithm

In evolutionary algorithms, the mutation operator and crossover operator are used to generate the new trail individual, and the selection operator chooses the better between the target individual and its trial alternative for the next generation by comparing their fitness values. Similar to the standard DE algorithm, the basic three evolutionary operators, i.e., the mutation operator, the crossover operator and the selection operator are also used for BDE to update the population. However, the individual elements of BDE only transform within two basic binary numbers in contrast to the standard DE with a real field. The important procedures of BDE algorithm are given as follows:

Initialization. Set the number of individuals in the population to be m and the dimension of parameters in each individual to be D. The initial random population, $\mathbf{P}(0)$ is generated by the prior probability π, which means that each individual $\boldsymbol{\omega}_i \in \mathbf{P}(0)$, $i = 1, 2, \ldots, m$ contains about πD 1's and $(1-\pi)D$ 0's. The maximum evolving generation is set to be N.

Mutation. For the standard DE, the mutation is achieved by difference of individuals. Since the individuals are real numbers, the mutation vector $\mathbf{v}_i$ for the ith individual is commonly generated by

$$\mathbf{v}_i^{t+1} = \boldsymbol{\omega}_{r_0}^t + F \cdot \left(\boldsymbol{\omega}_{r_1}^t - \boldsymbol{\omega}_{r_2}^t\right) \tag{1}$$

where t is the index of generation; F is the mutation rate; $\boldsymbol{\omega}_{r_0}, \boldsymbol{\omega}_{r_1}$ and $\boldsymbol{\omega}_{r_2}$ are three elements of the randomly chosen individuals with index $r_0 \neq r_1 \neq r_2 \neq i$. However, for binary DE, where the elements of individuals are binary numbers, the mutation vector $\mathbf{v}_i$ for the ith individual is obtained through utilizing the power function, given as the following equation:

$$v_{i,j}^{t+1} = \omega_{r_0,j}^t + (-1)^{\omega_{r_0,j}} \cdot \left|\omega_{r_1,j}^t - \omega_{r_2,j}^t\right|, j = 1, 2, \ldots, D \tag{2}$$