Journal of Process Control 24 (2014) 1068–1075

Contents lists available at ScienceDirect

Journal of Process Control

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

Development of a variable selection method for soft sensor using artificial neural network and nonnegative garrote

Kai Sun [a], Jialin Liu [b], Jia-Lin Kang [c], Shi-Shang Jang [c,*], David Shan-Hill Wong [c,*], Ding-Sou Chen [d]

[a] Department of Automation, Qilu University of Technology, Jinan, Shandong 250353, China
[b] Center for Energy and Environmental Research, National Tsing-Hua University, Hsin-Chu 30013, Taiwan
[c] Department of Chemical Engineering, National Tsing-Hua University, Hsin-Chu 30013, Taiwan
[d] New Materials Research & Development Department, China Steel Corporation, Kaohsiung 81233, Taiwan

ARTICLE INFO

Article history:
Received 26 August 2013
Received in revised form 16 February 2014
Accepted 19 May 2014
Available online 14 June 2014

Keywords:
Variable selection
Soft sensor
Nonnegative garrote
Artificial neural network

ABSTRACT

This paper developed a new variable selection method for soft sensor applications using the nonnegative garrote (NNG) and artificial neural network (ANN). The proposed method employs the ANN to generate a well-trained network, and then uses the NNG to conduct the accurate shrinkage of input weights of the ANN. This paper took Bayesian information criterion as the model evaluation criterion, and the optimal garrote parameter s was determined by v-fold cross-validation. The performance of the proposed algorithm was compared to existing state-of-art variable selection methods. Two artificial dataset examples and a real industrial application for air separation process were applied to demonstrate the performance of the methods. The experimental results showed that the proposed method presented better model accuracy with fewer variables selected, compared to other state-of-art methods.

1. Introduction

Soft sensors are inferential models that use easily measured variables to estimate process variables that are hard to measure due to technological limitations, large measurement delays, or high investment costs [1,2]. In addition, soft sensors can give useful information in terms of fault detection by working with hardware sensors in parallel [3]. However, the complexity of a process model will increase exponentially and lead to the problem of dimensionality as the number of variables increases. It is possible to eliminate redundant variables, reduce the complexity of the model, and present more accurate model through the use of appropriate variable selection techniques. A variety of variable selection techniques for soft sensor applications and modeling methods have been studied in recent years. Wold et al. [4] developed a basic tool of chemometrics using partial least square (PLS) regression. Ahmed et al. [5] presented a recursive PLS method for soft sensors and applied it to predict the melt index during grade change in a high density polyethylene plant. Kano et al. [6] constructed the inferential models with a dynamic PLS, in which the past measurements were used, and successful application in a multicomponent distillation column showed that the dynamic PLS was effective for improving the accuracy of the prediction. Facco et al. [7] designed a soft sensor for the online estimation of product quality in a real-world industrial batch polymerization process based on the PLS method. Liu [8] developed a soft sensor for a polyethylene process with multiple production grades by integrating principal component analysis (PCA) and the fuzzy Takagi–Sugeno method. Ma et al. [9] proposed an adaptive soft sensor based on statistical identification of key variables, in which the key variables are captured using the statistical approach of stepwise linear regression. Xiang et al. [10] presented a framework of discriminative least squares regression (LSR) for multiclass classification and feature selection, in which the core idea is to enlarge the distance between different classes under the conceptual framework of LSR.

However, these linear regression methods may not have adequate accuracy in describing highly nonlinear industrial processes. Artificial neural networks (ANN) are powerful tools for nonlinear statistical data modeling or decision making and have been widely used for variable selection in recent years. Castellano and Fanelli [11] proposed a backward selection by successively removing the input nodes of a satisfactorily trained neural network with the complete set of variables as inputs, and the method showed better performance than PCA. Onnia et al. [12] presented a feature selection method using sequential forward neural networks

* Corresponding authors. Tel.: +886 3 571 3697; fax: +886 3 571 5408.
E-mail addresses: ssjang@mx.nthu.edu.tw (S.-S. Jang), dshwong@che.nthu.edu.tw (D.S.-H. Wong).

http://dx.doi.org/10.1016/j.jprocont.2014.05.010
0959-1524/

with the sum-of-squares error (SSE) as the saliency. Romero and Sopena [13] proposed a feature selection method using sequential backward multi-player perceptron (SBS-MLP), which retrained the network with every feature temporarily removed before computing its saliency. It has significant performance but its computational cost is very high because of the network retraining that occurs at every step. Maya et al. [14] developed a nonlinear variable selection ANN using partial mutual information, and demonstrated the superior performance in comparison to linear correlation-based techniques. Windeatt et al. [15] proposed an embedded feature ranking strategy based on multi-layer perceptron (MLP) weights combined with recursive feature elimination, as well as a stopping criterion based on out-of-bootstrap estimates, and the techniques worked well in removing irrelevant features. Souza et al. [16] developed a new method for input variable selection for soft sensor design using neural networks and mutual information. Kaneko and Funatsu [17] developed a region selection method by combining genetic algorithm-based wavelength selection with support vector regression, which was called GAWLS-SVR. The simulated results and application in industrial polymer process showed the GAWLS-SVR was very effective in constructing inferential models of nonlinear time delayed processes.

Recently, Breiman [18] proposed a new shrinkage method called the nonnegative garrote (NNG). The mechanism of this shrinkage method conducts variable selection by shrinking or setting some coefficients of a "greedy" model to zero. In brief, NNG is a two-step shrinkage algorithm. In the first step the initial coefficients are obtained using ordinary least squares (OLS), and in the second step magnitude shrinkage of the initial coefficients is conducted using the so-called "garrote" constraints. Yuan [19] proposed a non-parametric extension of NNG and applied it to solve the problem of component selection in a functional analysis of variance (ANOVA) model. Pan et al. [20] revised the original NNG by combining enumerative PLS with NNG, and the results showed better performance than original NNG. Wang et al. [21] developed a soft sensor modeling algorithm with an adaptive PLS nonnegative garrote by incorporating non-stationary disturbance, and the performance was demonstrated with two real industrial applications.

The motivation of this paper was to develop a robust variable selection method for the application of soft sensors that could describe complex nonlinear industrial processes. This paper is organized as follows. Section 2 reviews the theories of ANN and NNG variable selection. Section 3 presents the development of the methodology. Section 4 gives the numerical simulation results of two artificial dataset examples. In Section 5, the proposed method is applied to predict the oxygen concentration for the air separation process. Finally, some concluding remarks are given in Section 6.

2. Background theory

2.1. Artificial neural network

Artificial neural networks are powerful tools to model complex multivariable processes. Haykin [22] presented a milestone textbook about neural networks, which provided comprehensive information of neural networks from an engineering perspective. The book presented broad coverage of ANN, including the network architectures, knowledge representation, training function, etc. The paper focuses on the development of variable selection method by combining the multi-layer perceptron (MLP) neural network of [22] with NNG. Fig. 1 shows the architecture of a MLP neural network that consists of three layers: an input layer, an output layer and a hidden layer. Each layer consists of multiple neurons which are connected to neurons in adjacent layers.

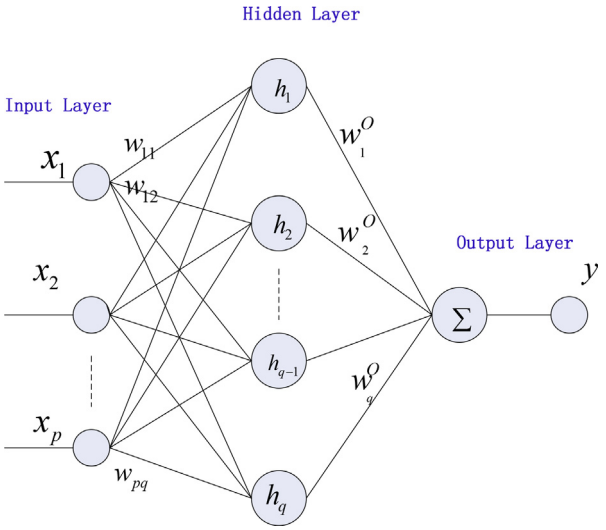

Fig. 1. Architecture of artificial neural network.

Let $x = \{x_1, x_2, \ldots, x_p\}$ denote the candidate input variables of the network. The hidden layer has q nodes represented as $h = \{h_1, h_2, \ldots, h_q\}$, and the weight w_{ij} $(i = 1, 2, \ldots, p;\ \ j = 1, 2, \ldots, q)$ denotes the input weight between input variable x_i and the jth hidden neuron h_j. The output signal of the jth neuron of the hidden layer, O_j^h, is given by:

$$O_j^h(x) = f\left(\left(\sum_{i=1}^{p} w_{ij}x_j\right) + b_j^h\right), \quad j = 1, 2, \ldots, q \tag{1}$$

where b_j^h is the bias of the jth neuron of the hidden layer, and f represents the activation function of the hidden layer.

Furthermore, the output of the network, y, which represents the response variable, is given by:

$$y = g\left(\left(\sum_{j=1}^{q} w_j^o O_j^h(x)\right) + b^o\right) \tag{2}$$

where b^o is the bias of the output layer, g represents the activation function of the output layer, and $w_j^o (j = 1, 2, \ldots q)$ represents the jth output weight between the hidden layer and the output layer. Substituting Eq. (1) into Eq. (2), y can be formulated as:

$$y = g\left(\left(\sum_{j=1}^{q} w_j^o f\left(\left(\sum_{i=1}^{p} w_{ij}x_j\right) + b_j^h\right)\right) + b^o\right) \tag{3}$$

2.2. Nonnegative garrote

Breiman [18] proposed the NNG to solve subset regression for linear problems, which can provide simple regression models with better estimation accuracy than ordinary subset selection methods. Compared to another famous shrinkage and selection method known as LASSO, which was proposed by Tibshirani [23], the most significant advantage of NNG is that it is "path consistent" [24,25], meaning that during the search path for the best estimate, there is at least one estimate that is consistent with the actual process model, in terms of both the coefficient estimate and variable selection. In original NNG, the true model has the following form:

$$y = \sum_{i=1}^{p} \beta_i x_i + \varepsilon \tag{4}$$

where $x = \{x_1, x_2, \ldots, x_p\}$ and y are input and output variables separately, $\beta = \{\beta_1, \beta_2, \ldots, \beta_p\}$ is the coefficient and ε is the random