Advanced Engineering Informatics 36 (2018) 112–119

Contents lists available at ScienceDirect

Advanced Engineering Informatics

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

Full length article

Soft sensor based on stacked auto-encoder deep neural network for air preheater rotor deformation prediction

Xiao Wang, Han Liu*

School of Automation and Information Engineering, Xi'an University of Technology, China
Shaanxi Key Lab of Complex System Control and Intelligent Information Processing, China

ARTICLE INFO

Keywords:
Soft sensor
Deep learning
Process industry
Nonlinear regression

ABSTRACT

Soft sensors have been widely used in industrial processes over the past two decades because they use easy-to-measure process variables to predict difficult-to-measure ones. Some success has been achieved by the dominant traditional methods of modeling soft sensors based on statistics, such as principal components analysis (PCA) and partial least square (PLS), but such sensors usually become inaccurate and inefficient when processing strong nonlinear data. In this paper, a new soft sensor modeling approach is proposed based on a deep learning network. First, stacked auto-encoders (SAEs) are employed to extract high-level feature representations of the input data. In the process of training each layer of a SAE, the Limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm (L-BFGS) is adopted to optimize the weights parameters. Then, a support vector regression (SVR) is added to predict the target value on the basis of the features obtained from the SAE. To improve the model performance, Genetic Algorithm (GA) is used to obtain the optimal parameters of the SVR. To evaluate the proposed method, a soft sensor model for estimating the rotor deformation of air preheaters in a thermal power plant boiler is studied. The experimental results demonstrate that the soft sensor based on the SAE-SVR algorithm is more effective than the existing methods are.

1. Introduction

There are many key variables in the industrial process that are closely related to product quality or production efficiency. We cannot achieve high product efficiency and consistent product quality if these variables are not monitored and controlled effectively. However, they are usually difficult to measure directly by hardware sensors because such measurement instruments can be unavailable and costly [1].

Fortunately, the aforementioned problems may be solved by the soft sensing method. Soft sensing is primarily used to choose a set of measurable variables that are relevant to the estimated variables and to construct a mathematical model that uses measurable variables as input and estimated variables as output [2].Soft sensors can deliver real-time continuous estimates of those difficult-to-measure target process variables [3].

At a very general level, there are two classes of soft sensors: model-driven and data-driven [1]. Model-driven methods have long occupied a dominant position, and they achieved good performance. These methods include multivariate statistics [4,5], Kalman filters [6,7], and clustering [8]. However, these methods may lead to lower model precision than the data-driven class because their modeling processes are always based on knowledge of the visible mechanism but not on real data from the industrial process. Therefore, it is difficult to model the whole process while considering every aspect when using model-driven methods due to the intricate physical background and harsh environments of the industrial locations. Recently, data-driven soft sensors have gained increasing popularity because they can save time and have a low cost, especially in the complicated process industry. It has been demonstrated that data-driven models have higher accuracy because they are based on the data measured in the processing plants, so they can describe real process conditions more reliably.

In this study, we propose a data-driven soft sensor model integrating stacked auto-encoders (SAEs) with a support vector regression (SVR) to estimate the rotor deformation of air preheaters in a thermal power plant boiler. An air preheater is the gas heat exchanger in the power plant boiler, in which the heat carried by the exhaust gas is transferred to the combustion-supporting air. In this process, a mushroom-shaped deformation occurs to the rotor of the air preheater due to the uneven heat, which can lead to air leakage and greatly reduce the thermal efficiency of the boiler greatly. Therefore, it is critical to monitor the deformation of the rotor and minimize the air leakage in a power plant. However, the air preheater is located in a harsh environment with high-

* Corresponding author at: School of Automation and Information Engineering, Xi'an University of Technology, China.
E-mail address: Liuhan@xaut.edu.cn (H. Liu).

https://doi.org/10.1016/j.aei.2018.03.003
Received 11 September 2016; Received in revised form 27 February 2018; Accepted 13 March 2018
1474-0346/

temperatures and dust, which often put the hardware sensors out of service. Therefore, we introduce a soft sensor technique to solve this problem.

Our contributions in this paper are as follows:

- We propose a SAE-SVR soft sensor that integrates stacked auto-encoders with SVR. The reconstruction loss function of each auto-encoder and the optimal algorithm are improved to pursue convergent speed and prediction accuracy.
- We introduce the SAE-SVR to estimate the rotor deformation of an air preheater in a thermal power plant boiler. This is the first such application of the soft sensor technique according to the currently available literature.

The rest of this paper is structured as follows: The related literature is reviewed in Section 2. In Section 3, the proposed method is detailed and consists of feature learning by stacked auto-encoders and support vector regression. The soft sensor modeling procedure based on SAE-SVR is detailed in Section 4. An industrial case study about the rotor deformation prediction of air preheaters in a thermal power plant boiler is shown in Section 5. Finally, Section 6 provides some conclusions.

2. Related work

For a traditional data-driven soft senor, the most widely used modeling methods are principal components analysis (PCA) [9], partial least square (PLS) [10], artificial neural network [11,12] and their nonlinear extensions, such as neural network PLS (NNPLS) [13]. These methods have had some successful applications in industrial process control. However, they can only utilize labeled data containing both input and target values, which is available for only a limited number of industrial process control systems due to cost, time, or limitations of other resources [14]. Large amounts of unlabeled data containing only input samples are wasted, and there are no guarantees of achieving high convergence speed or avoiding local minima in general neural networks [15]. Therefore, these methods are often insufficient to handle the highly correlative data [16]. To process industrial data that reflect complex nonlinear underlying relationships, a more suitable method should be developed. In recent years, layer-by-layer unsupervised feature learning with deep networks has been studied and applied successfully in many fields [17–19]. Deep learning performs well in those complex problems that cannot be perfectly solved by traditional methods, due to its remarkable representation ability [20,21]. It can successfully build more complex features, thus making supervised learning easier; meanwhile, it can avoid problems such as gradient diffusion when training deep architectures directly [22].It has also been verified that deep learning is especially suitable for soft sensor modeling because it has a greater ability to yield informative representation than traditional data-driven models do [23]. Compared to traditional soft sensors, deep learning captures the underlying intricate relationships of industrial data accurately and incorporates all of the available process data, including the unlabeled data wasted by traditional methods.

An auto-encoder is an effective method for the unsupervised pretraining of deep neural networks [24]. There are a number of successful applications of auto-encoders and their variants in many fields, including facial identification and acoustic feature extracting [25,26]. Despite the powerful capacity of feature extracting, the plain auto-encoder is inappropriate for regression because an additional operation is requested on the extracted features. Support vector regression (SVR) is the extended application of the famous support vector machine (SVM), and it has a strong ability to approximate nonlinear functions and good generalization performance [27–29]. However, SVR is often inefficient when the data set becomes too large [30,31]. In [32], a denoise auto-encoder (DAE) and SVR are integrated to estimate the oxygen-content of flue gasses in ultra-supercritical units, and the combination performs well.

In the present study, we propose a novel soft-sensing modeling approach integrating the stacked auto-encoders (SAEs) with SVR. We firstly construct a deep network by stacking several auto-encoders to extract features of the input data. Then, in view of the need for regression prediction in industrial process control, we add a SVR predictor to the top layer. Unlike [32], in our approach, there is not only a single hidden layer but there are two to provide better data handing capability. Moreover, the reconstruction loss function and optimization algorithm are improved in this study. In short, we use SAE as a feature extractor of input data and SVR as the predictor of continuous target variable values in our soft sensor model.

3. The proposed SAE-SVR soft sensor

3.1. Data features extraction by stacked auto-encoders

An auto-encoder is a three-layer feed forward network that attempts to reproduce its input with minimum reconstruction loss; i.e., the target data are the same as the input data. Therefore, the output layer size is always the same as the input layer. An auto-encoder consists of an encoder and a decoder. The encoder maps the input data x to a hidden representation y and can be seen as a function:

$$y = s(W^{(1)}x + b^{(1)}) \tag{1}$$

where W is the weight matrix and b is the bias vector. In this paper, W and b are collectively called weight for simple representation. s(.) is the sigmoid function and is defined as:

$$s(x) = \frac{1}{1 + \exp(-x)} \tag{2}$$

The decoder maps y to $\hat{x}$ such that $x \approx \hat{x}$, and is considered a function:

$$\hat{x} = s(W^{(2)}y + b^{(2)}) \tag{3}$$

The reliability of the auto-encoder is estimated by its reconstruction capability. To recover the input data from the output layer, we need to obtain the optimal parameter set θ by minimizing the reconstruction loss, where the parameter set is $\theta = \{W^{(1)},b^{(1)},W^{(2)},b^{(2)}\}$. In this study, which is different from the generally accepted in auto-encoder, the cost function J_{cost} is defined as follows:

$$J_{cost}(\theta) = J_{recon}(\theta) + J_{weight}(\theta) \tag{4}$$

in which the reconstruction loss function is defined as

$$J_{recon}(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}\log(\hat{x}_{ij}) + (1-x_{ij})\log(1-\hat{x}_{ij})) \tag{5}$$

where m is the number of training samples and n is the input dimension. Cross-entropy is employed as the reconstruction loss function, as suggested in [33]. It is insensitive to outliers, unlike the mean square error (MSE), and has been successfully utilized for cost function design in non-Gaussian signal processing [27,31].

Then, the weight decay term is defined as

$$J_{weight}(\theta) = \frac{\lambda}{2}\sum_{l=1}^{2}\sum_{i=1}^{k_l}\sum_{j=1}^{k_{l+1}}(w_{ji}^{(l)})^2 \tag{6}$$

where λ is the decay coefficient, k_l is the number of neurons in layer l and $w_{ji}^{(l)}$ represents an element in $W^{(l)}$. By adding the weight decay term, the cost function is turned into a strictly convex function, ensuring that it obtains the global optimal solution while simultaneously preventing overfitting.

After initializing the weights, an optimization algorithm should be selected to complete the training process, aiming for the optimal weight values. At present, there are a variety of optimization algorithms, most of which are based on iteration. An advanced Limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm (L-BFGS) [34] is chosen in this paper and is discussed later.