Control Engineering Practice 46 (2016) 94–104

Contents lists available at ScienceDirect

Control Engineering Practice

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

Use of a quantile regression based echo state network ensemble for construction of prediction Intervals of gas flow in a blast furnace

Zheng Lv, Jun Zhao*, Ying Liu, Wei Wang

School of Control Science and Engineering, Dalian University of Technology, Dalian, China

ARTICLE INFO

Article history:
Received 7 February 2015
Received in revised form
8 October 2015
Accepted 8 October 2015

Keywords:
BFG generation
High level noises and outliers
Echo state networks
Quantile regression
Prediction intervals

ABSTRACT

The usual huge fluctuations in the blast furnace gas (BFG) generation make the scheduling of the gas system become a difficult problem. Considering that there are high level noises and outliers mixed in original industrial data, a quantile regression-based echo state network ensemble (QR-ESNE) is modeled to construct the prediction intervals (PIs) of the BFG generation. In the process of network training, a linear regression model of the output matrix is reported by the proposed quantile regression to improve the generalization ability. Then, in view of the practical demands on reliability and further improving the prediction accuracy, a bootstrap strategy based on QR-ESN is designed to construct the confidence intervals and the prediction ones via combining with the regression models of various quantiles. To verify the performance of the proposed method, the practical data coming from a steel plant are employed, and the results indicate that the proposed method exhibits high accuracy and reliability for the industrial data. Furthermore, an application software system based on the proposed method is developed and applied to the practice of this plant.

1. Introduction

Blast furnace gas (BFG) is a very significant byproducts energy resource generated during steel making process. In the production procedure, the generation of BFG usually fluctuates with high magnitude along with the changing of equipment status, and such a sudden fluctuation can even lead to the production accidents (Zhao, Wang, Liu, et al., 2011). Given a large amount of historic operational data coming from the practical production process, a class of data-driven based methodology could be an effective attempt for resolving the forecasting problem of the gas flow (Liu, Zhao, Lv, et al., 2012; Liu, Zhao, & Wang, 2010).

In the literature, a number of studies related to time series based prediction were reported in (Box, Jenkins, & Reinsel, 2013). Echo state network (ESN), a class of recurrent neural network concentrated on its specific dynamic reservoir (DR) structure, has a remarkable potential of short-term memory and exhibits good performance for the dynamic prediction for the nonlinear or non-Gaussian systems (Jaeger & Haas, 2004). And, such a network model had been applied to time series prediction (Wyffels & Schrauwen, 2010), short-term load prediction (Showkati, Hejazi, & Elyasi, 2010), automatic control (Plöger, Arghir, Günther, et al., 2004), etc. However, it should be noticeable that the current ESN application focused on those chaotic systems without noises, and the performance could be largely impacted if it is adopted to the industrial data with high level noises or anomalies (Liu, Zhao, & Wang, 2013). On the other hand, the most important issue of ESN modeling lies in its calculation of its output weight, and the generic approach for such a problem is to employ the least square-based algorithm. However, such a method usually leads to ill-conditioned solutions when there are anomalies data involved in the training samples. With respect to such an issue, an additive noise-based method for the state variables were mentioned in Jaeger (2005). Nevertheless, it was very difficult to determine the magnitude of the noise, which made it hard to be applied to the practice. Besides, a ridge regression learning-based method was then reported in Shi and Han (2007a) to obtain the output weight, which exhibited a remarkable result only when the ill-condition appeared and it was similarly hard to select the ridge parameters. In addition, although a denoising method based on empirical mode decomposition (EMD) was designed before modeling an ESN for an industrial application, see the details in Liu, Zhao, Wang, et al. (2009), the denoising process took a rather large period of computing time. And, a class of kernel-based method for such a problem was also investigated in Shi and Han (2007b), where the cross-validation based normalization parameters determination made it very intensive computing cost.

With regard to the practical demand of the energy system in a steel plant, there are very few research outcomes related to the prediction problem with high level noises and outliers (Liu et al.,

* Corresponding author. Fax: +86 411 84707581.
E-mail address: zhaoj@dlut.edu.cn (J. Zhao).

http://dx.doi.org/10.1016/j.conengprac.2015.10.003
0967-0661/

2010; Liu, Liu, Wang, et al., 2012). The existing studies were basically relevant to the data processing technology before modeling, which usually generated a negative impact on the system modeling because of the effectiveness of data processing, thus is not appropriate for the real-time online application (Sáez, Luengo, & Herrera, 2013). On the other hand, it should be also noticeable that the uncertainty involved in the industrial data might bring about a poor performance if only considering the point-oriented methods, which neglected the quantitative reliability analysis for the forecasting error and failed to be really applied to the industrial practice (Khosravi, Nahavandi, & Creighton, 2013). Fortunately, the prediction interval mode can effectively resolve the drawbacks of the point-oriented forecasting, which can take into account not only the uncertainty of data resource, but the difference between the established model and the real values as well, and finally provide the confidence intervals and the prediction ones corresponding to the predicted values (Heskes, 1997; Hwang & Ding, 1997). While, in the existing prediction intervals constructions, a number of methods concentrated only on the noises analysis (Sheng, Zhao, Wang, et al., 2013), which paid few attentions on the anomalies data widely existed in original industrial dataset.

Considering the drawbacks of ESN based modeling and the industrial application demand, a quantile based ESN ensemble modeling method is proposed in this study for the PIs construction, which aims at the non-Gaussian distributed model residuals of industrial energy system. This paper employs the quantile regression to compute the output weight of the ESN models and optimize the modeling error for the purpose of the robustness and the stability when facing with the noises and outliers. And, a bootstrap based method is then proposed to train the proposed ensemble model. To verify the effectiveness of the proposed method, the practical data coming from a steel plant are employed. A series of experiments relevant to the three practical scenarios indicate that the proposed method can greatly enhance the stability of the prediction model and has good performance for the PIs construction of the BFG generation. And, an application software developed by the proposed method has been applied to the energy management system in a steel plant for over six months, and the operating results indicate that the PIs results can help the energy scheduling workers to learn the flow tendency of the BFG system and provide sound scheduling guidance.

The rest of this paper is organized as follows. The general issues related to the BFG system and the demand for the gas flow prediction are briefly introduced in Section 2. Section 3 proposes the quantile regression-based echo state network ensemble (QR-ESNE) model for industrial data with high noises and outliers, and gives the PIs constructed by the proposed method. The validation of the proposed method is experimentally addressed in Section 4. Finally, this paper is summarized in Section 5.

2. Problem description

Byproduct gas is one of the most important energy resources in steel industry, and its operation scheduling can directly influence the production cost and the energy consumption of steel enterprise. BFG, regarding as the key byproduct for energy scheduling, has the largest generation and consumption amount among the gases. Taking a certain steel plant as an example, there are four blast furnaces viewed as the BFG generation units. The gas transportation system consists of pipeline network, gas holders, pressure stations, power plants, low pressure boilers, etc., and the gas consumption units involve coke oven, rolling process, chemical production units, etc. The brief structure of the BFG system can be illustrated in Fig. 1. Due to the processes of hot-blast stoves switching, blast furnaces damping down or reducing wind, there are large gas flow fluctuations transported into the entire BFG system, which will lead to a negative impact on the gas balance of generation and consumption amounts, and influence the production of the other gas consumption units in the BFG system.

Nowadays, the BFG generation is approximately estimated on a basis of personal experiences of the energy scheduling workers. However, since the experiences coming from various workers lack of consistency, the inaccurate energy flow estimation occurs. Fig. 2 shows a series of typical BFG generation within continuous 2 h. It can be seen that, there are a lot of uncertainties and disturbances during the industrial process, which made the collected data contain high level noises and a number of outliers.

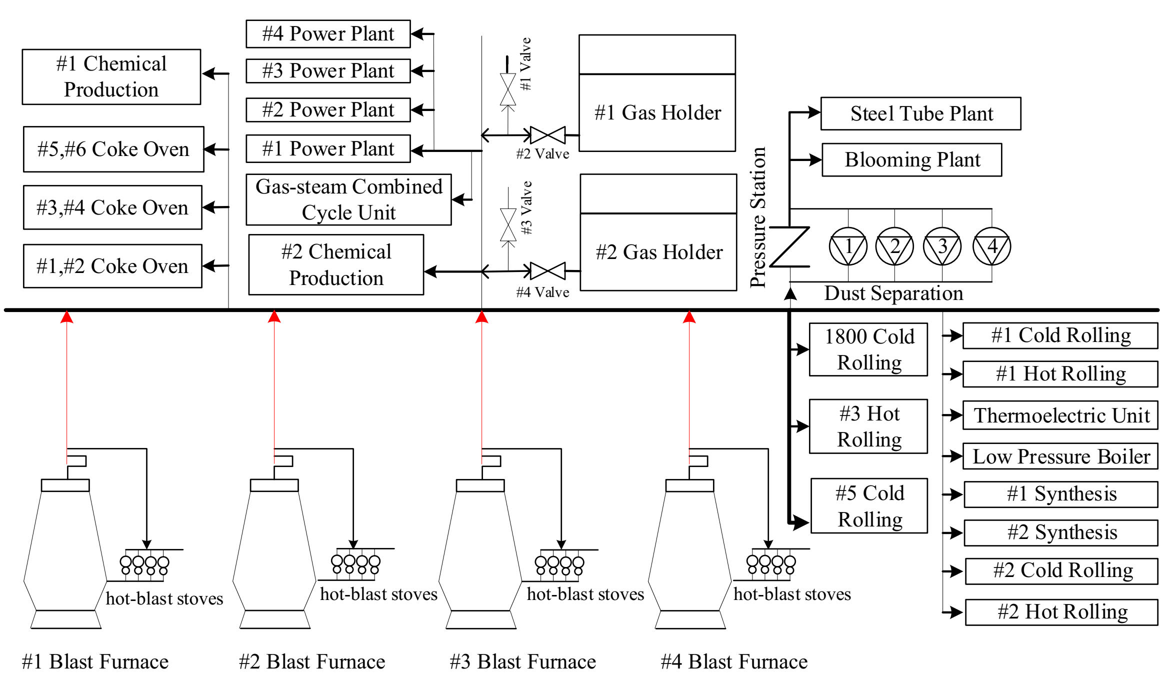

Fig. 1. The structure of BFG system.