Control Engineering Practice 78 (2018) 79–88

Contents lists available at ScienceDirect

Control Engineering Practice

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

Construction of prediction intervals for gas flow systems in steel industry based on granular computing

Zhongyang Han [a], Jun Zhao [a,*], Henry Leung [b], Wei Wang [a]

[a] *School of Control Sciences and Engineering, Dalian University of Technology, Dalian, China*
[b] *Department of Electrical and Computer Engineering, University of Calgary, Calgary, Canada*

A R T I C L E I N F O

Keywords:
Steel industry
Byproduct gas
Long-term prediction
Prediction intervals

A B S T R A C T

Understanding the future flow variation of byproduct gas is very crucial for energy scheduling in steel industry. An accurate prediction of the tendencies is significantly beneficial for raising the economic profits of steel enterprise. Given that most existing techniques focus on short term or numeric prediction that can hardly meet the practical requirements on the predicting horizon, the guidance effect of the results imposing on energy scheduling is limitative. In this study, a granular computing (GrC)-based method for the construction of prediction intervals (PIs) is proposed, which considers semantic features of the gas flows and granulate the data so as to form a number of unequal-length granules on the horizontal axis. Dynamic time warping technique is then deployed to equalize the granules' lengths. As for the longitudinal (amplitudes of gas flows) granular expansion, one can regard the data amount covered by the granulation as an objective to optimize the allocation of information granularity for constructing PIs. To verify the performance of the proposed GrC-based approach, this study exhibits a series of comparative experiments by using the practical industrial data, and the developed prediction system is also applied in the energy center of Baosteel Co. Ltd. The results indicate that the application system presents high accuracy and can provide an effective guidance for balancing and scheduling of the byproduct energy.

1. Introduction

Byproduct gas is the significant secondary fuel generated during the manufacturing process in steel industry, such as blast furnace iron-making, converter steel-making, etc. In order to coping with the frequently occurred imbalance between generation and consumption units, the scheduling work has to be operated according to the flow variation characteristics. It could be lack of accuracy and timeliness for making decision when such imbalance already existed. While if the flow variation tendency can be understood in advance, then a relevant energy scheduling scheme can be established ahead, which allows an efficient, low-cost and sustainable manufacturing process of steel industry.

In the literature, most energy prediction methods in steel industry considered the data-driven approaches. An online parameter optimization-based prediction was modeled by Zhao, Wang, Pedrycz, and Tian (2012) for a converter gas system, and a parallel computing strategy based on GPU structure was implemented. A multi-kernel support vector regressor (SVR) was designed to estimate the gas tank level of a steel plant (Zhao, Liu, Zhang, & Wang, 2012). Similarly, the complexity of a converter gas system was analyzed by Han, Liu, Zhao, and Wang (2012), where the multiple tank levels were simultaneously predicted by a multi-output least square SVR. Although these methods provided satisfactory results for predicting horizon limited within one hour, their accuracies will largely degrade if the horizon has to be extended. Through many discussions with the scheduling operators on-site, a long-term prediction of energy variations is of higher practical value, especially for the energy equipments with long term starting or switching phases.

Granular computing (GrC), originated a decade ago as a unified conceptual and processing framework, is substantially capable of associating with fuzzy sets (Xu & Li, 2016), rough sets (Li & Xu, 2015; Xu, Li, & Zhang, 2017), and interval analysis for data-driven modeling and optimization (Chai, Xu, & Wong, 2016; Leite, Palhares, Campos, & Gomide, 2015; Peters, 2011; Tsang, Wang, Chen, Wu, & Hu, 2013). In a general sense, one can regard information granule, i.e., the basic cell in GrC, as a collection of elements drawn together by their closeness, such as resemblance, proximity, functionality, etc., articulated in terms of some functional relationships (Pedrycz, 2016). Partitioning the information granules based on ascending or descending tendencies of data fluctuation, a fuzzy reasoning method was proposed to engage in a long-term prediction reported by Dong and Pedrycz (2008). But, its solving

* Corresponding author.
E-mail address: zhaoj@dlut.edu.cn (J. Zhao).

https://doi.org/10.1016/j.conengprac.2018.06.012
Received 18 September 2017; Received in revised form 21 June 2018; Accepted 22 June 2018
0967-0661/

instances were only a number of artificial or traditional datasets that failed to handle industrial datasets or application-oriented problems. Recently, a time series based long-term prediction was proposed by Zhao, Han, Pedrycz, and Wang (2016), and a factor-based model was established for energy flows and storage predictions by Han, Zhao, Wang, Liu, and Liu (2014), respectively. In addition, a multi-output model addressed by Han, Zhao, Liu, and Wang (2016) adopted a hybrid collaborative clustering for multiple gas tanks levels. It can be noticeable that these prediction studies of long-term versions focused on the numeric mode, which makes the results described by only a series of exact predicted values that lacks of the quantitative evaluation on the prediction reliability.

For industrial practice, it is of concerns to the on-site operators that the reliability of the prediction results could be quantified. There are various methods for prediction intervals (PIs) construction in literature. For instance, regarding the error as a Gaussian distribution near the mean value of the target output, the mean–variance estimates (MVEs) method constructs PIs by estimating the mean and variance. The Delta method predicts the variance between the observation and the output of the model so as to provide PIs results, see the details in Khosravi, Nahavandi, Creighton, and Atiya (2011a, b). In Guan, Luh, Michel, and Chi (2013), a hybrid Kalman filter was designed for the PIs construction of electricity powerload. Combining a particle swarm optimization (PSO) with a support vector machine (SVM), a PIs model for electricity price was also reported by Shrivastava, Khosravi, and Panigrahi (2015). Furthermore, the probabilistic forecasting was also widely applied for PIs construction, such as electricity price (Bello, Bunn, Reneses, & Muñoz, 2017), wind power generation (Wan, Xu, Pinson, Dong, & Wong, 2014), etc. However, these approaches only demonstrated medium or short-term PIs, which cannot satisfy the practical requirements of the energy decision making in steel industry. Based on a class of artificial neural network model, a PIs model was provided for rainfall runoff estimation (Kasiviswanathan, Cibin, Sudheer, & Chaubey, 2013). Besides the predicting horizon was relatively short, a large amount of model parameters was involved in these approaches, which came up with a low availability in dealing with real-time industrial applications. With respect to the chaotic time series prediction, a non-parameter PIs was developed for climate forecasting reported by Rong and Xiao (2013). However, such a method required not only a large amount of training data, but it also provides a rather short term predicting horizon. In the perspective of long-term PIs construction with data-based methodologies, very few research results are reported in the literature, especially in the field of industrial applications.

A GrC-based long-term PIs construction approach is proposed in this study. Considering the phasic characteristics (i.e., industrial manufacturing semantics), data is granulated based on semantics to create a series of information granules with different lengths. And, a time warping normalization technique equalizes these granules for subsequent fuzzy clustering process of the long-term prediction. The PIs are then constructed on the basis of the extension in the longitude axis via optimizing the allocation of the information granularity so as for quantifying the prediction reliability. To demonstrate the effectiveness of the proposed method, a number of experiments by using the practical energy data of a steel plant are conducted in this paper. And, an application software system based on the proposed method is developed and applied to the Baosteel Co. Ltd., China. The results indicate that the application system is capable of providing long-term prediction with satisfactory accuracy and PIs coverage that meet the practical requirements for byproduct gas scheduling.

The rest of this paper is organized as follows. The industrial problem considered here is presented in Section 2. Section 3 describes the modeling processes, including data granulation, PIs construction and long-term prediction. The comparative analyses and the developed system are addressed in Section 4. Finally, Section 5 summarizes this study.

2. Problem descriptions

A typical structure of the byproduct gas system in a steel plant can be illustrated in Fig. 1. Three categories of gases that are produced during the processes of iron making, coking and steel making are important energy resources, i.e., blast furnace gas (BFG), coke oven gas (COG) and LD converter gas (LDG). Due to the complexity of the steel production process, the gas flows of generation and consumption exhibit intensive nonlinearity characteristics. Furthermore, the energy pipeline networks in plant cover a large area and long distance, thus it is hardly for a mechanism-based model to describe such a system accurately. Based on the accumulative massive data obtained from the supervisory control and data acquisition (SCADA) system, data-driven modeling and prediction become practical approach to solve this complicated industrial problem (Zhao, Liu, Pedrycz, & Li, 2012; Zhao, Liu, Wang, Pedrycz, & Cong, 2012; Zhao, Wang, Liu, & Pedrycz, 2011).

Conventional approaches usually suffer from two drawbacks in predicting the energy flows. First, given that the accuracy is limited by using the iterative point-wise methods, they can only provide relatively accurate results when the predicting horizon is within 1–2 h at most. Thus, the energy scheduling can hardly be satisfactory because some devices in steel industry have to be operated in advance for more than 6 h. In other words, a byproduct energy prediction over a period of 6 h is desired to provide an effective production scheme or maintenance scheme. As such, the manufacturing efficiency and the cost reduction could be largely enhanced. Second, the existing studies gave prediction results mostly in terms of numeric values, which are rather absolute and also not in accordance with the sense of human cognition. In other words, it is practically insufficient for the results to merely provide a series of predicted values, and the prediction reliability has also to be quantitatively illustrated for meeting the on-site scheduling demands.

3. GRC-based prediction intervals construction

It is essential for a GrC framework to granulate the data, which has a direct impact on the subsequent modeling performance. In this study, (1) considering the phasic variation characteristics of the gas flows, a horizontal data granulation based on the industrial semantics is established; and (2) as for the longitudinal granulation, one can regard the intervals coverage as the objective for optimizing the allocation of information granularity so as to quantify the prediction reliability (PIs construction). The detailed procedures of the proposed method can be summarized as Fig. 2. In order to clarify the notations of the utilized symbols, a nomenclature is firstly given here as Table 1.

3.1. Industrial semantics based horizontal data granulation

In literature, ascending and descending tendencies are usually viewed as data granulation principle, which obviously lacks of the consideration on practical data variation meanings. Different from generic time series data, the byproduct gas flows in steel industry always exhibit some practical semantics, i.e., phasic production procedures, such as the LDG generation flows during steel-making process, the COG consumption flows during iron-reducing process, etc.

In this study, the industrial data are first granulated by using the semantics characteristics. Here, one can take the BFG consumption of the hot blast stove as an example, which is illustrated in Fig. 3. In this figure, the flow can be partitioned into a series of similar segments (granules), each of which typically consists of two phases corresponding to the practical operation, i.e., the regular burning (the amplitude values more than 180 km^3/h, lasting approximately 30 min) and the temporary suspending (the amplitude values less than 170 km^3/h, lasting approximately 15 min). In practice, the gas flow of the hot blast stove frequently varies due to its running status switching. It is also obviously that the ascending/descending tendency description cannot explicitly represent such industrial features. Therefore, this study