Received September 9, 2017, accepted October 5, 2017, date of publication October 17, 2017, date of current version November 28, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2763630

Knowledge-Based Control and Optimization of Blast Furnace Gas System in Steel Industry

HAIXIA WANG, CHUNYANG SHENG, AND XIAO LU
College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao 266590, China

Corresponding author: Chunyang Sheng (chunyang860124@163.com)

This work was supported in part by the National Nature Science Foundation of China under Grant 61773245, Grant 61603068, Grant 61503224, and Grant 61273197, in part by the China Postdoctoral Science Foundation under Grant 2015M582115 and Grant 2016T90640, in part by Taishan Scholarship Construction Engineering, and in part by the Shandong University of Science and Technology Public Visiting Scholar and Scientific Research Foundation of Shandong University of Science and Technology for Recruited Talents.

ABSTRACT Aiming at the control and optimization problem of blast furnace gas (BFG) systems in the steel industry, a knowledge-based optimal control algorithm combining fuzzy rules extraction with neural networks (NNs) ensemble-based prediction is proposed. On one hand, a fuzzy model is designed to extract the expert control knowledge from the historical data of the industrial process after community detection, and then, a great deal of scheduling knowledge is employed to compose a fuzzy rule base, which can be used for fuzzy inference of control scheme with a new input. On the other hand, data-driven NNs ensemble is built to model the BFG system for prediction. Meanwhile, the prediction results can provide the inputs when using fuzzy rule base for control and optimization. Finally, a BFG system of one steel enterprise is studied in this paper for experiments, which verifies the effectiveness and practicability of the proposed method.

INDEX TERMS Blast furnace gas, knowledge base, fuzzy rules, control, optimization.

I. INTRODUCTION

From Aug. 2005, energy saving and emission reduction has been a key topic of the development of the industry, since beautiful scenery is the gold and silver mines proposed by the president Xi. In China, steel enterprises are not only a significant part of the national industry, but also with great deal of energy consumption and sewage [1]. Thus, one important objective of the industrial intelligent manufacturing is to improve the ability of energy saving and emission reduction. Blast furnace gas, the key component of the byproduct gas system of steel industry, is one kind of important secondary energy. To improve the capability of optimal control of blast furnace gas will make sense to the energy saving and emission reduction.

Recent years, there are many researches about the control, scheduling and optimization of the blast furnace gas and even the byproduct gas system. Firstly, some researchers pointed out that it is very necessary to build the mathematical programming model for the byproduct gas system. A mixed integer linear program-based model was firstly proposed in [2] to realize the long-term optimization of the by-product gas system, in which the objective functions include minimizing the fluctuation of the gas holder levels, minimizing the fuel cost, penalty costs and the purchased power cost, maximizing the generation of the power system, and the constraints include restriction of process mechanism, energy balance and security constraints. However, it seems that the mathematical programming model reported in [2] is reasonable and effective for the optimal control of the byproduct gas system. However, there are some deficiencies and drawbacks which lead to the model proposed in [2] cannot be applied to the practical field, so some improved researches are conducted in [3]–[5]. In [3], a dynamic mixed integer linear programming model for multi-period optimization of byproduct gas is proposed to optimize byproduct gas distribution. In [4], a green mixed integer linear programming model is also proposed for byproduct gas distribution, in which the environmental cost caused by pollutants discharge is factored in total cost. In [5], aiming at the time-delay problems of mathematical programming, a forecast model of gas supply, gas demand and surplus gas in a steel plant was proposed for the dynamic programming of byproduct gas system to develop an optimal utilization strategy of byproduct gas. Although mathematical programming-based techniques are proposed relatively earlier, it has not been widely utilized due to the following reasons. First, the over ideal or over relaxation of the parameters of byproduct process leads to a decreasing accuracy of the model. Second, the complexity of the byproduct gas system leads to the difficulties of solution procedure. Generally, it is difficult to choose a suitable initial solution for the optimization model. Thus, the feasible solution of the model might not exist. Third, it is lack of the forecast or prediction for the future, the time-delay of the programming model cannot be avoided, which will also lead to the low accuracy of the model.

As for one byproduct gas system, gas holder is generally viewed as the key component, which can effectively regulate the balance of the supply and demand of system [6], [7]. When the supply is superior to the demand, the surplus gas can be reserved by gas holder. And when the supply is inferior to the demand, the gas reserved can be supplied to the pipeline network. However, the capability of the gas holders is limited, so some imbalance conditions will occur in the gas system, which will cause some unnecessary waste and sewage. Therefore, some researchers insisted that the optimal control problem of the byproduct gas system can be transformed to an equilibrium control problem of the gas holder level. Recently, the most common-used method for the equilibrium control of gas holder level is data-driven modeling technique, which can be divided into two steps. In the first step, a machine learning method is employed to identify the nonlinear relationship between the gas holder level and the flow of the generation and consumption units of the gas system. Meanwhile, the gas holder level and the flow of the generation and consumption units are predicted. In the second step, the nonlinear relationship and the prediction results are used to determine the rough control scheme. The above stated research is firstly reported in [8], where a two-stage online prediction method is proposed for a blast furnace gas system. In the first stage, the flow of the generation and consumption units is predicted based on echo state networks. And in the second stage, the relationship between the gas holder level and the flow of the generation and consumption units are identified, furthermore, the gas holder level is also predicted based on the identified relationship. Based on the above modeling and prediction, the gas holder level can be effectively controlled. Subsequently, similar researches reported in [9] and [10] are done for coke oven gas system and converter gas system, the mechanisms of which are also like the blast furnace gas system. The above-mentioned optimal control technique is relatively novel and extensively applied to many steel enterprises. In addition, some researchers considered that the probability graphical model can be used to build the causality relationship between the gas holder level and adjustable users, and the corresponding research was reported in [11]. The Bayesian network is adopted to model the control problem of the blast furnace gas holder level, i.e., a probability relationship between the gas holder level and the adjustable users is established. And then, the different adjustable users will be chosen with different indicated probabilities in an operation scheme, and the users with the biggest probability will be chosen firstly.

So far data-driven optimal control methods are still in dominant position for the byproduct gas system. However, it is not difficult after analyzing to find that all the above-mentioned methods are not suitable for the special operation condition and abnormal service condition. In the real field of steel industry, the optimal control schemes of special or abnormal conditions mostly depend on the expert experience and knowledge. Therefore, the combination of the expert knowledge and scientific techniques will be more effective way for the optimal control. Generally, the expert knowledge can be got through the communication with the expert, yet this way is very difficult and with low efficiency. Moreover, the expert knowledge should be transformed into the logical language that computer can identify, which requires more computational consumption. After investigation and study, the expert knowledge is certainly applied into the industrial process in the form of the control strategy, so there is plenty of latent expert knowledge in the monitor data of industrial process. Another way of knowledge extraction is data-driven machine learning techniques, in which the latent knowledge in data could be mined for application. However, the optimal control scheme depends not just on the expert knowledge that might be with time-delay. In our opinion, the expert knowledge is combined with quantitative modeling and prediction will be more effective to improve the capability of optimal control of byproduct gas system.

Aiming at the optimal control problem of the blast furnace gas system in steel industry, a knowledge-based optimal control algorithm combining fuzzy rules extraction with neural networks (NNs) ensemble-based prediction is proposed in this study. First, the community detection of complex network-based samples selection method is proposed to partition the different operational conditions of blast furnace gas system. Meanwhile, the typical and valuable data samples are selected from the original large-scale dataset. Second, a fuzzy model is designed to extract the expert scheduling knowledge from the historical data of the industrial process after community detection. And then, a great deal of scheduling knowledge is employed to compose a fuzzy rules base, which can be used for fuzzy inference of control scheme with a new input. Third, data-driven NNs ensemble is built to model the blast furnace gas system for prediction. Based on the prediction results, the control scheme can be reasoned by using the fuzzy rule base constructed above. Finally, a byproduct gas system of one steel industry is studied for experiments to verify the effectiveness and practicability of the proposed method, which shows the proposed technique is very meaningful to the energy reservation and emission reduction of industrial enterprises.

This rest is organized as follows. The problem about the control and optimization of the BFG system is described in Section II. The samples selection and the extraction of the expert knowledge-based control rules are discussed in Section III. In Section IV, the NNs ensemble-based prediction model for the BFG system is proposed, and meanwhile the parameters of the prediction model are estimated by the Bayesian regularization technique. The effectiveness of the proposed method is experimentally verified

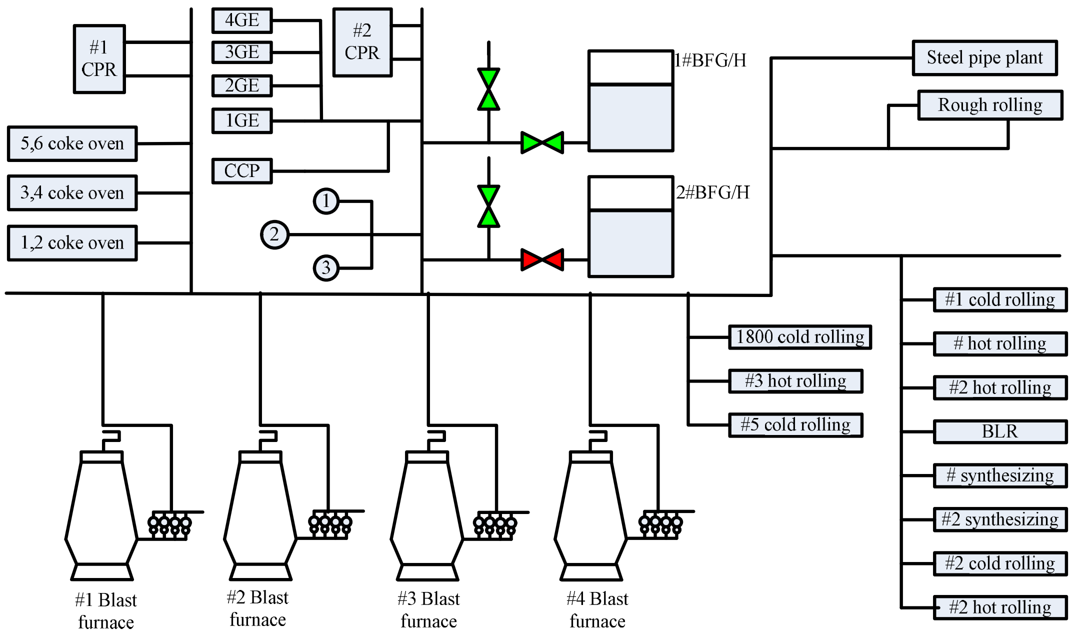

FIGURE 1. The structural diagram of the blast furnace gas system.

in Section V. Finally, some conclusion remarks are drawn in Section VI.

II. PROBLEMS DESCRIPTION

A typical BFG system, which contains the gas generation units, the consumption units, transmission pipeline network and storage units, is very complicated. Generally speaking, the storage units are viewed as the core of the system, which can balance the difference between the supply and demand of the pipeline network to guarantee the security operation. However, since the capability of the storage units is limited, the surplus gas will be burnt in air when the gas supply is superior to the gas demand, which leads to energy waste and air pollution. On the contrary, when the gas supply is inferior to the gas demand, some outsourcing energy leading to additional economic cost, such as natural gas or fuel-oil, will be consumed. A real blast furnace gas system of Bao-steel industry, the structural diagram of which is shown in Fig. 1, is employed here to illustrate the specific mechanism. Four blast furnaces viewed as the generation units can supply into the transportation network on average 1.8 million BFG per hour, whose calorific power is about 3200KJ/Nm^3 [8]. The transportation system usually includes pipelines, mixing stations and pressure stations. The consumption users primarily consist of coking oven, hot rolling plant, cold rolling plant, chemical products recovery, low pressure boiler, and power plant. Since the hot blast stoves of blast furnace expend quite a quantity of BFG, and be continuously switched, the generation amount flowed into transportation will frequently fluctuate. In Bao-steel, taking the overlapping of wave crests and troughs into consideration, the variation of generation amount may reach 500 thousand m^3 BFG per hour, which will exhibit a drastic impact on the whole gas system [8]. Although the gas holder can be treated as a buffer storage unit, its total capacity of only 300000 m^3 is hardly enough to completely respond to the variation present in the BFG system [8]. Besides, there are often abnormal conditions occurring in the production process such as blast reduction, user shutdown or equipment fault. Such circumstance also leads to system imbalance.

Based on the above statement, the operational level of gas holder is one of the most important indicator for studying the optimal control of the blast furnace gas system. In addition, it is necessary to note that the blast furnace gas system is not an isolated one but a key component of the total energy system. It has a strong connection with the consumption of outsourcing energy, coke oven gas, converter gas, the generation of steam heat and the electrical energy. Since the byproduct gas system is not only with complicated operational mechanism, but also with coupling phenomenon, it is hard to build a mechanism-based optimal control model. In recent years, data-driven techniques are more and more applied to the industrial filed and achieve some considerable performance. However, most of the existing data-driven methods belong to the category of quantitative calculation, which cannot cope with the sudden accident and some special conditions, since they lack high level intelligence. In addition, there are several controlled devices of the blast furnace gas system. It is difficult to adopt the quantitative calculation-based techniques to determine the control scheme. Thus, so far data-driven researches cannot be effectively applied to the practical industrial field.

In the industrial field, when the domain experts intend to make a control scheme, operational states of some important users are considered firstly. According to the monitor states, whether a control scheme is required or not is determined. And if one scheme is necessary, the domain experts will make it according to their accumulated knowledge. Thus, if the expert knowledge is available, one feasible and effective control scheme can be obtained. However, expert knowledge-based control algorithm is short of the judgment of the future. It is necessary to combine the expert knowledge with the prediction technique to obtain one more effective control scheme. The major advantage of the knowledge-based control scheme can be applied in the industrial field.

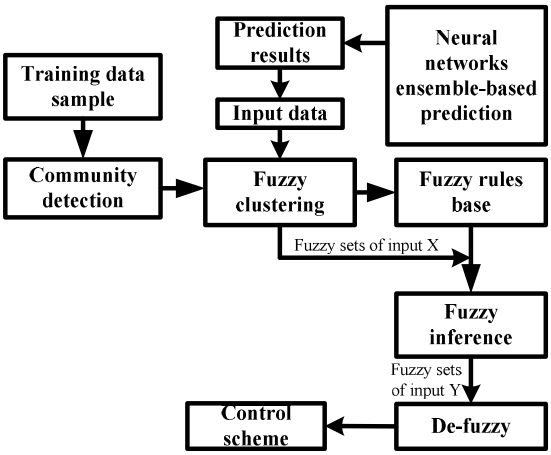

FIGURE 2. Chart of knowledge-based optimal control method.

III. EXPERT KNOWLEDGE-BASED CONTROL ALGORITHM

The proposed expert knowledge-based optimal control method involves three parts, as shown in Fig 2. First, the community detection of complex networks is employed to choose the samples with typical and valuable characteristics from the original training dataset. Second, the training data sample after community detection are discretized by means of the fuzzy C-means clustering, and the fuzzy modeling techniques are adopted to extract the fuzzy rules hidden in data. Then, a fuzzy rule base is built, which can be used for fuzzy inference of the control scheme.

Finally, to avoid the time delay of expert knowledge-based decision, a neural networks ensemble is built for modeling the blast furnace gas system. Based on the nonlinear model, the operational conditions of the BFG system can be predicted or forecasted. From the prediction results, whether one control scheme is required can be determined. And if one control scheme is required, the prediction results can be chosen as the novel input data for the fuzzy rules based after fuzzy clustering. After the fuzzy inference and the de-fuzzy technique, the control scheme can be obtained. This method can effectively utilize the expert knowledge and the large amount of the historical process data.

A. COMMUNITY DETECTION-BASED SAMPLES SELECTION

The community detection method takes the module maximum as the optimization objective to obtain the optimal community division and to further select the valuable training samples, which can avoid the influence coming from the randomness. We consider one original training dataset coming from the industrial field.

$$S = \{[\mathbf{x}_i(t), y_i(t-1)], y_i(t)|i = 0, 1, \cdots, N\} \tag{1}$$

where $\mathbf{x}_i(t)$ and $y_i(t-1)$ are the input samples, $y_i(t)$ is the output sample. As for the BFG system, $\mathbf{x}_i(t)$ denotes the generation or consumption flow of the units in the system, and $y_i(t)$ denotes the gas holder level.

In this study, we choose the Euclidean distance to calculate the connected relationship of a complex network. The formula of the Euclidean distance is written as

$$m_{ij} = \sqrt{\sum_{i\neq j} \left(s_{ik} - s_{jk}\right)^2} \tag{2}$$

Using the above formula, we can compute the Euclidean distance m_{ij} between two samples $[\mathbf{x}_i(t), y_i(t-1)], y_i(t)$ and $\left[\mathbf{x}_j(t), y_j(t-1)\right], y_j(t)$. Thus, the distance matrix $\mathbf{M}$ containing the Euclidean distance between any two samples is written as

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1N} \\ m_{21} & m_{22} & \cdots & m_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ m_{N1} & m_{N2} & \cdots & m_{NN} \end{bmatrix} \tag{3}$$

where $m_{ij} = m_{ji}$ and $m_{ii} = 0$.

To obtain an adjacency matrix $\mathbf{M}'$, a threshold parameter R is set here. All the elements in the matrix $\mathbf{M}$ will be compared with R. If the value of some element is inferior to R, the two related samples are viewed as adjacent samples. Otherwise it will be viewed as non-contiguous.

$$m'_{ij} = \begin{cases} 1, & if\ node\ i\ and\ j\ are\ connected \\ 0, & else \end{cases} \tag{4}$$

Then, we can get the adjacency matrix $\mathbf{M}'$ as follows.

$$\mathbf{M}' = \begin{bmatrix} m'_{11} & m'_{12} & \cdots & m'_{1N} \\ m'_{21} & m'_{22} & \cdots & m'_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ m'_{N1} & m'_{N2} & \cdots & m'_{NN} \end{bmatrix} \tag{5}$$

To explain how to use the community detection of complex networks to select valuable samples, the community detection of complex networks is firstly introduced. A typical network is commonly composed by nodes or vertices and edges or links. The nodes denote the different elements of one system and the edges can describe the relationship among different nodes. If there is one relationship between two nodes or the logical relationship between two nodes is very close, these two nodes are linked with an edge. The edge can be directed, also can be undirected. The degree of a vertex is the number of the edges connected to that vertex. For undirected networks, it can be computed as

$$k_i = \sum_j a_{ij} \tag{6}$$

where the value of a_{ij} denotes whether the two nodes i and j are connected. If i and j are connected, $a_{ij} = 1$, otherwise $a_{ii} = 0$.

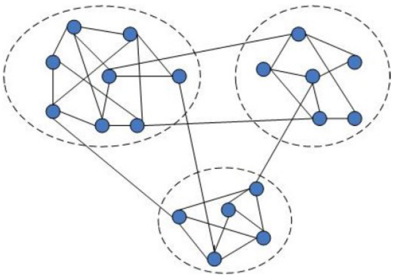

FIGURE 3. Community structure of complex networks.

As shown in Fig. 3, we can see a small network with community structure. In this case there are three communities, denoted by the dashed circles, which has dense internal links but between which there is only a lower density of external links. To determine the optimal community division and the number of the community, the evaluation index named modularity is proposed in [12]:

$$Q = \sum_i \left(e_{ii} - a_i^2\right) \tag{7}$$

where e_{ii} denotes the ratio between the edges in the ith community and the edges of the whole network. $a_i = \sum_j e_{ij}$ denotes the ratio between the edges of the nodes in community i connected to other community and the edges of the whole network. The stronger the structure of the community is, the larger the value of Q will be. Here, we use hierarchical clustering to integrate different community. Originally, each node can be viewed as one community, and we can integrate each two communities in one step. The integration of the community should follow one rule that is we should choose the integration resulting in the maximal increase of Q. The variety of the value of Q after the community integration can be described as

$$\Delta Q = e_{ij} + e_{ji} - 2a_i a_j = 2(e_{ij} - a_i a_j) \tag{8}$$

After n steps, when $\Delta Q < 0$, the value of Q can achieve the maximum. Now, the structure of the network is optimal after community division. Since the original communities are described by the nodes, if we assume that node i and j are connected by one edge, then $e_{ij} = 1/2m$, otherwise $e_{ii} = 0$. Meanwhile, $a_i = k_i/2m$. The original value of the matrix can be set as

$$\Delta Q = \begin{cases} \dfrac{1}{2m} - \dfrac{k_i k_j}{(2m)^2}, & \textit{if } i \textit{ is connected with } j \\ 0, & \textit{otherwise} \end{cases} \tag{9}$$

where m is the number of edges of the whole network.

Based on the above principle about the community detection, the optimal community can be detected by computing the value of ΔQ. When using the community detection algorithm for the matrix $\mathbf{M}'$, the samples with the similar characteristics will be divided into the same community.

As for the data-driven modeling, the quality of the training samples has a great influence on the generalization ability and the accuracy. When the similar training samples account for a large proportion in the sample population to other samples, the learning process of the prediction model will face the over-fitting due to redundancy problem. After community detection, the samples located on the edge of the community and connected with the other community can be viewed as the redundancy samples. While, the samples in one community that do not connect with the samples in other community or has very sparse connection can be viewed as the samples with typical characteristics. To avoid the redundancy problem of the training samples and to effectively quantify the specification of the samples, an evaluation index named "joint binding degree" is defined as follows.

$$c_i = k_{i-in}/k_i \tag{10}$$

where k_{i-in} is the number of the edges linked with node i in the community and k_i is the degree of node i. Generally, a large "joint binding degree" represents that the node belongs to the community with a higher probability. $c_i = 1$ means the node i belongs to this community completely while $c_i < 1$ means the node i is located on the edge of the community and probably has a similarity relation with the nodes in other community.

According to (10), we can compute the joint binding degree of each node. Further, we sort the nodes in the community in descending order in accordance with the joint binding degrees. Finally, we can choose the effective samples from different communities according to the value of the joint binding degree of different samples. The training dataset composed by the effective samples contains most of operational conditions in the industrial process, based on which the generalization ability of the prediction model can be improved obviously.

The detailed steps of implementation are listed as follows when using the community detection of complex networks for the sample selection.

Step 1: Construct the training dataset by sampling from the industrial real-time database. Each training sample can be viewed as one node of the complex network.

Step 2: Compute the Euclidean distance m_{ij} of any two samples i and j in the training dataset according to (3). And then the distance matrix $\mathbf{M}$ is obtained.

Step 3: Choose a suitable threshold parameter R. Compare each element m_{ij} of $\mathbf{M}$ with the parameter R to obtain the adjacent matrix $\mathbf{M}'$.

Step 4: Compute the value of ΔQ and use the community detection algorithm for the complex network represented by $\mathbf{M}'$.

Step 5: Based on the results of community detection, the joint binding degree of the nodes can be computed according to (10). Sort the nodes of one community in descending order.

Step 6: Choose the effective samples according to the joint binding degrees to construct the training dataset.

B. CONSTRUCTION OF FUZZY RULES BASE

According to the above-mentioned problems, the gas flow of the generated units and the consumers, the reservation of the storage units and the amount of the adjustable units at k are considered as the inputs of the fuzzy model. And the variable amount of the adjustable units from k to $k+h$ is set as the corresponding outputs. It should be noted that the amount of the adjustable users cannot be varied without manual intervention. If the number of the input units is n and the number of the adjustable units is l, the inputs are written as $\mathbf{X}(k) = \{x_1(k), x_2(k), \cdots, x_n(k)\}$ and the outputs are written as $\mathbf{Y}(k) = \{y_1(k), y_2(k), \cdots, y_l(k)\}$. To extract the control knowledge hidden in data, the original input and output data is transformed into fuzzy sets first by the means of fuzzy C-means clustering [13]. Take one input component x_i as an example, assumed that C_i different clusters with cluster centers v_{ij} are obtained, and then the cluster information and the degree $\mu_i(k)$ of membership of $x_i(k)$. And so on, all the fuzzy sets and their membership information of the inputs and outputs variables can be obtained, and then the fuzzy rules are extracted as the following form.

$$\begin{aligned} R_i: \ & \textit{If } x_1(k) \textit{ is } A_{1,1} \textit{ and } x_2(k) \textit{ is } A_{2,1} \\ & \quad \textit{and} \cdots \textit{and } x_m(k) \textit{ is } A_{m,1} \\ & \text{Then } y_1(k) \textit{ is } B_{i,1} \textit{ and } y_2(k) \textit{ is } B_{2,1} \\ & \quad \textit{and} \cdots \textit{and } y_l(k) \textit{ is } \mathrm{B}_{l,1} \end{aligned} \tag{11}$$

where R_i is the jth rule. $x_i(k)$ is the ith component of inputs at k. $A_{i,j}$ denotes the jth cluster of the ith component $x_i(k)$ of inputs. $y_i(k)$ is the ith component of outputs at k. $B_{i,j}$ denotes the jth cluster of the ith component $y_i(k)$ of outputs. Based on the above statement, a set of fuzzy rules are built to construct an expert knowledge base.

When $A_{i,j}$ and $B_{i,j}$ are viewed as the domain of the inputs $\mathbf{X}$ and outputs $\mathbf{Y}$, the relationship of each fuzzy rule can be described as

$$R_j = (A_{1j} \times A_{2j} \times \cdots \times A_{mj}) \times (B_{1j} \times B_{2j} \times \cdots \times B_{lj}) \tag{12}$$

And the membership function of R_j can be written as

$$\mu_{R_j}(\mathbf{X}, \mathbf{Y}) = \mu_{A_{1,j}}(x_1) \wedge \cdots \wedge \mu_{A_{m,j}}(x_m) \wedge \mu_{B_{1,j}}(y_1) \wedge \cdots \wedge \mu_{B_{l,j}}(y_l) \tag{13}$$

The fuzzy relationship of all rules is obtained by combination of all the relationships, that is

$$\mathbf{R} = \bigcup_{j=1}^{N} R_j \tag{14}$$

And the membership function of $\mathbf{R}$ can be written as

$$\mu_{\mathbf{R}}(\mathbf{X}, \mathbf{Y}) = \bigvee_{j=1}^{N} \mu_{R_j}(\mathbf{X}, \mathbf{Y}) \tag{15}$$

C. EXPERT KNOWLEDGE-BASED FUZZY INFERENCE

Fuzzy inference is to deduce a conclusion based on some fuzzy pre-conditions. If the fuzzy relationship between the inputs and the outputs is available, a new output can be reasoned when given a set of new values of the inputs $\mathbf{X}^*(k) = \{x_1^*(k), x_2^*(k), \cdots, x_n^*(k)\}$. First, the clustering technique is also used here to obtain the fuzzy membership matrix of each input data. And then the fuzzy sets and degree of the membership of the different output are reasoned based on the fuzzy relationship. Finally, the fuzzy sets of the output with the largest degree of the membership are obtained. Based on the de-fuzzy technique, the control scheme can be obtained. To sum up, the expert knowledge-based inference in this study provides the feasible control scheme, and sometimes the optimal one, which is meaningful to the industrial utilization since it is almost impossible to find a globally optimal control scheme. As for the fuzzy inference, the effective inputs are vital elements. However, historical or real-time data cannot be chosen for inputs, since the means of control might lag, which is fatal for industrial application. Thus, the effective and accurate prediction is required for the BFG system to obtain a set of advanced inputs.

IV. PREDICTION MODEL OF BFG SYSTEM

A. NEURAL NETWORKS ENSEMBLE-BASED PREDICTION MODEL

According to the operational mechanism of the BFG system, we design one specific NNs ensemble to predict the variety of the BFG system whose structure is shown in Fig. 4. The networks ensemble shown in Fig. 4 is composed by two ESNs and one perceptron model, where two ESNs are used for prediction the total generations and consumptions of the BFG system, respectively. The perceptron model is designed to predict the gas holder levels. y_1 can be viewed as the total generation amount of BFG and y_{l-1} can be viewed as the total consumption amount of BFG except consumption of the adjustable users. From y_2 to y_{l-2} is the consumption amount of the adjustable users, such as the boiler, power plant and synthesizing users. z is the output of the ensemble and denotes the gas holder level, and y_l is the gas holder level in the previous moment. From the structure of the proposed prediction model, the gas holder level at time k is determined by many factors, including the total generation amount of BFG at time k, the total consumption amount of BFG at time k, the consumption amount of adjustable users at time k and the gas holder level at time $k-1$. The total generation amount of BFG denotes the total generation of all the blast furnace. The total consumption amount denotes the consumption of all consumption users, including hot rolling, colder rolling, steel pipe plant, et al. Since the consumption of the adjustable users cannot be changed without human interruption, so we do not make predictions for $y_2, y_3, \cdots, y_{l-2}$. The total generation amount and the total consumption amount are series varying with time, the two ESNs are employed to predict y_1 and y_{l-1}.

Based on the structure of NNs ensemble, the formulas of the proposed method are composed by the formulas of these two ESNs and the formula of the perceptron. First, the formula of the ESN model is written as [14], [15]

$$\begin{aligned} \mathbf{x}(k) &= f\left(\mathbf{W}^{in}\mathbf{u}(k) + \mathbf{W}\mathbf{x}(k-1)\right) \\ y_j(k) &= f^{out}\left(\mathbf{W}^{out}[\mathbf{u}(k), \mathbf{x}(k)]\right) \end{aligned} \tag{16}$$

where $\mathbf{u}(k) \in \mathbb{R}^{m\times 1}$ is the inputs of the network with the form $\mathbf{u}(k) = [u_1(k), u_2(k), \cdots, u_m(k)]$, $m = m_1$ or m_2. $\mathbf{x}(k) \in \mathbb{R}^{N\times 1}$ is the internal states of the dynamical reservoir with the form $\mathbf{x}(k) = [x_1(k), x_2(k), \cdots, x_N(k)]$, $N = N_1 or N_2$. $y_j(k)$ is the output, $j = 1 or l-1$. Here $\mathbf{W}^{in} \in \mathbb{R}^{N\times m}$ is the connection matrix describing the relationships between the elements located in the input and the DR. $\mathbf{W} \in \mathbb{R}^{N\times N}$ is the weight matrix of the neurons in DR. Note that to provide sufficient memorization capabilities, $\mathbf{W}$ should be a sparse matrix whose connectivity level is 1%∼5% and the spectral radius is less than 1. $\mathbf{W}^{out} \in \mathbb{R}^{(N+m)\times 1}$ is the output weights. f is the activation of internal neurons, usually a hyperbolic tangent function and f^{out} is the output activation, usually an identity. The formula of the perceptron is written as

$$z = f(\mathbf{w}, \mathbf{y}) = f(\sum_{j=1}^{l} w_j y_j) \tag{17}$$

where w_j is the weight parameters and z is the output of the ensemble.

As for the NNs ensemble, the parameters estimation is generally a difficult task due to the complexity of the structure of a NNs ensemble and large quantity of unknown parameters. To reduce the difficulties of parameters estimation, the ESNs are chosen for the individuals of the ensemble since ESN has a lower computational complexity. The characteristics of the ESN is that the input weights $\mathbf{W}^{in}$ and the internal weights $\mathbf{W}$ are given before training and fixed in the training process, thus only the output weights $\mathbf{W}^{out}$ need to be estimated [15]. And as for the ensemble, only the weights $\mathbf{w} = \{w_j\}_{j=1}^{l}$ and the output weights of two ESNs $\mathbf{W}_1^{out}$ and $\mathbf{W}_{l-1}^{out}$ are unknown. For more intuitive description, here the notation $\boldsymbol{\theta}$ is employed here to replace $\mathbf{w}, \mathbf{W}_1^{out}, \mathbf{W}_{l-1}^{out}$. Thus, we can construct the Bayesian regularization method to estimate the unknown parameters.

B. BAYESIAN REGULARIZATION OF THE POSTERIOR DISTRIBUTION

Consider the observed outputs with additive noise

$$t_i = z_i + \varepsilon_i \tag{18}$$

where ε_i is a Gaussian noise with zero mean and the variance σ_ε^2. Based on the regression described in (17), the output $z(\mathbf{u}^*, \mathbf{y}^*, \boldsymbol{\theta})$ of the ensemble depends on the indirect input $\mathbf{u}^*$, direct inputs $\mathbf{y}^*$ and a set of model weights $\boldsymbol{\theta}$. Then, the conditional distribution can be written as the integral over these parameters

$$p(t^*|\mathbf{u}^*, \mathbf{y}^*, \boldsymbol{\theta}) = \int p(t^*|\mathbf{u}^*, \mathbf{y}^*, \boldsymbol{\theta}) \cdot p(\boldsymbol{\theta}|D)d\boldsymbol{\theta} \tag{19}$$

where $p(t^*|\mathbf{u}^*, \mathbf{y}^*, \boldsymbol{\theta})$ is a likelihood function, which denotes the difference between the real observed targets t^* and the output z^* of the ensemble when given the parameters $\mathbf{w}$ and $\mathbf{W}^{out}$.

$$\begin{aligned} &p(t^*|\mathbf{u}^*, \mathbf{y}^*, \boldsymbol{\theta}) \\ &\quad= \left(\frac{\beta}{2\pi}\right)^{1/2} \exp\left(\frac{\beta}{2}\left\{f(\mathbf{u}^*, \mathbf{y}^*, \boldsymbol{\theta}) - t^*\right\}^2\right) \end{aligned} \tag{20}$$

where β is a hyper-parameter with the value $1/\beta = \sigma_t^2$. $\mathbf{y}^* = [y_1^*, y_2^*, \cdots, y_l^*]^T$, in which $y_2^*, \cdots, y_{l-2}^*$ is the known consumption of the adjustable users and y_l^* is the gas holder levels at the precious moment. y_1^* and y_{l-1}^* are the total generation and consumption of the BFG system, respectively.

$$y_1^* = \mathbf{W}_1^{out}\left[\mathbf{u}_1(k), \mathbf{x}_1(k)\right] \tag{21}$$

$$y_{l-1}^* = \mathbf{W}_{l-1}^{out}\left[\mathbf{u}_{l-1}(k), \mathbf{x}_{l-1}(k)\right] \tag{22}$$

On the right side of (19), another posterior distribution $p(\boldsymbol{\theta}|D)$ is still unknown. Using the Bayes' rule, $p(\boldsymbol{\theta}|D)$ can be written as

$$p(\boldsymbol{\theta}|D) = \frac{p(D|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(D)} \tag{23}$$

If we assume the priori of $p(\boldsymbol{\theta})$ is a Gaussian distribution, the distribution $p(\boldsymbol{\theta}|D)$ is written as [16]

$$p(\boldsymbol{\theta}|D) = \frac{1}{Z_S}\exp\left(-\frac{\beta}{2}E_D - \frac{\boldsymbol{\alpha}}{2}E_W\right) = \frac{1}{Z_S}\exp\left(-S(\boldsymbol{\theta})\right) \tag{24}$$

where Z_S is a normalizing constant.

The term E_D is the contribution from likelihood $p(D|\boldsymbol{\theta})$, which assumes that the data is independent, can be written as the product

$$\begin{aligned} p(D|\boldsymbol{\theta}) &= \prod_{i=1}^{N} p(t_i^n|\mathbf{u}_i^n, \mathbf{y}_i^n, \boldsymbol{\theta}) \\ &= \frac{1}{Z_D(\beta)}\exp\left(-\frac{\beta}{2}\sum_{i=1}^{N}\left\{f(\mathbf{u}_i^n, \mathbf{y}_i^n, \boldsymbol{\theta}) - t_i^n\right\}^2\right) \\ &= \frac{1}{Z_D(\beta)}\exp\left(-\frac{\beta}{2}E_D\right) \end{aligned} \tag{25}$$

where now Z_D is the normalizing constant given by the integral over $\mathbf{w}$ and $\mathbf{W}^{out}$ which gives $Z_D(\beta) = (2\pi/\beta)^{N/2}$. $\mathbf{y}_i^n = [y_{i,1}^n, y_{i,2}^n, \cdots, y_{i,l}^n]^T$, in which $y_{i,2}^n, \cdots, y_{i,l-2}^n$ is the known consumption of the adjustable users and $y_{i,l}^n$ is the gas holder levels at the precious moment. $y_{i,1}^n$ and $y_{i,l-1}^n$ can be described as

$$y_{i,1}^n = \mathbf{W}_1^{out}[\mathbf{u}_{i,1}^n(k), \mathbf{x}_{i,1}^n(k)] \tag{26}$$

$$y_{i,l-1}^* = \mathbf{W}_{l-1}^{out}\left[\mathbf{u}_{i,l-1}^n(k), \mathbf{x}_{i,l-1}^n(k)\right] \tag{27}$$

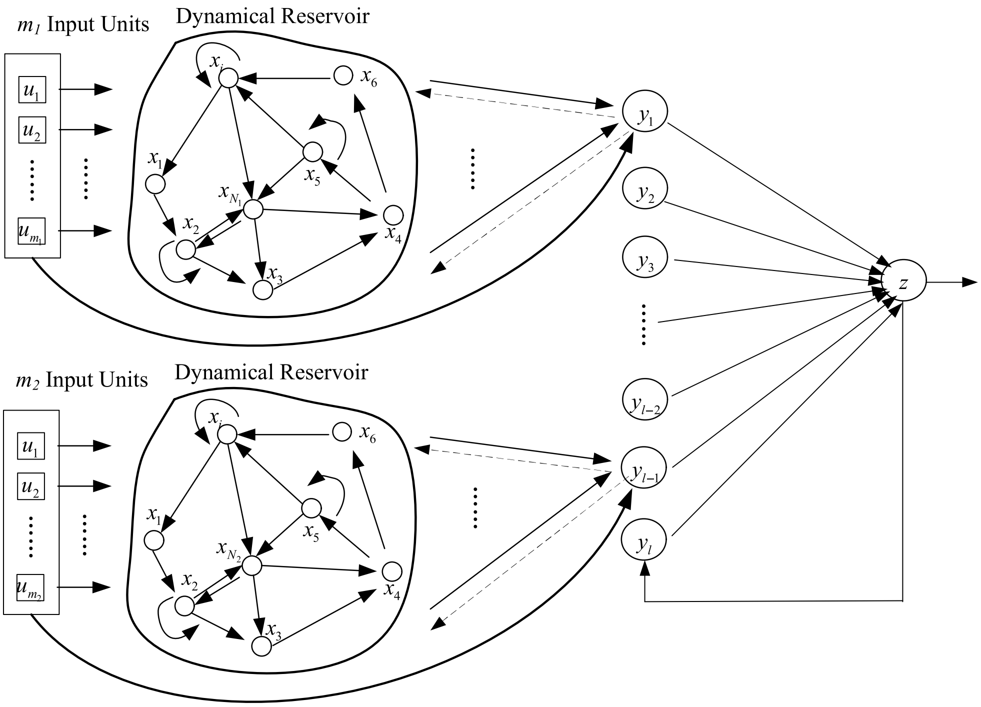

FIGURE 4. Structure of NNs ensemble for BFG system.

The second term E_W is the contribution from the prior over the weights [16]

$$\begin{aligned} p(\boldsymbol{\theta}|\boldsymbol{\alpha}) &= p(\mathbf{W}_1^{out}, \mathbf{W}_{l-1}^{out}, \mathbf{w}|\boldsymbol{\alpha}) \\ &= \frac{1}{Z_W(\boldsymbol{\alpha})} \exp\left(\sum_{i=1}^{3}\left[-\frac{\alpha_i}{2}\|\boldsymbol{\theta}_i\|^2\right]\right) \\ &= \frac{1}{Z_W(\boldsymbol{\alpha})} \exp\left(-\frac{\boldsymbol{\alpha}}{2}E_W\right) \end{aligned} \tag{28}$$

where $\boldsymbol{\theta}_1 = \mathbf{W}_1^{out}$, $\boldsymbol{\theta}_2 = \mathbf{W}_{l-1}^{out}$, $\boldsymbol{\theta}_3 = \mathbf{w}$, again $Z_W(\boldsymbol{\alpha})$ is a normalizing constant

$$Z_W(\boldsymbol{\alpha}) = \int \exp(-(\boldsymbol{\alpha}/2)E_W)d\boldsymbol{\theta}$$

To evaluate the most probable weights $\boldsymbol{\theta}_{MP}$, we can maximize the posterior distribution $p(\boldsymbol{\theta}|D)$ that is equivalent to minimizing the function $S(\boldsymbol{\theta})$. First, the function $S(\boldsymbol{\theta})$ can be linearized by Taylor series expansion and neglect the third-order terms, it leads to the approximation

$$S(\boldsymbol{\theta}) \approx S_{MP} + (\Delta\boldsymbol{\theta})^T\mathbf{A}(\Delta\boldsymbol{\theta}) \tag{29}$$

where $\mathbf{A}$ is the Hessian matrix $\mathbf{A} = \nabla_{\boldsymbol{\theta}}\nabla_{\boldsymbol{\theta}}S_{MP}$, $\Delta\boldsymbol{\theta} = \boldsymbol{\theta} - \boldsymbol{\theta}_{MP}$ and $S(\boldsymbol{\theta}_{MP})$ has been written as $S(\boldsymbol{\theta})$

$$S(\boldsymbol{\theta}) = \sum_{i=1}^{3}\left[\frac{\alpha_i}{2}\|\boldsymbol{\theta}_i\|^2\right] + \frac{\beta}{2}\sum_{i=1}^{N}\left\{f(\mathbf{u}_i^n, \mathbf{y}_i^n, \boldsymbol{\theta}) - t_i^n\right\}^2 \tag{30}$$

Substituting (24) (20) into (19) leads to the relationship

$$p(t^*|\mathbf{u}^*, \mathbf{y}^*, D) \propto \int \left\{\exp\left(-\frac{\beta}{2}\left\{f(\mathbf{u}^*, \mathbf{y}^*, \boldsymbol{\theta}) - t^*\right\}^2\right) \cdot \exp\left(-\frac{1}{2}(\Delta\boldsymbol{\theta})^T\mathbf{A}(\Delta\boldsymbol{\theta})\right)\right\} d\boldsymbol{\theta} \tag{31}$$

The function $f(\mathbf{u}^*, \mathbf{y}^*, \boldsymbol{\theta})$ may now be linearly approximated by Taylor expanding about $\boldsymbol{\theta}_{MP}$, i.e.,

$$f(\mathbf{u}^*, \mathbf{y}^*, \boldsymbol{\theta}) \approx f(\mathbf{u}^*, \mathbf{y}^*, \boldsymbol{\theta}_{MP}) + \mathbf{g}^T\Delta\boldsymbol{\theta} \tag{32}$$

where $\mathbf{g} = \nabla_{\boldsymbol{\theta}}f(\mathbf{u}^*, \mathbf{y}^*, \boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{MP}}$.

Substituting into (30) and evaluating the integral over $\boldsymbol{\theta}$ gives

$$p(t^*|\mathbf{u}^*, \mathbf{y}^*, D) = \frac{1}{(2\pi\sigma_{t^*}^2)^{1/2}} \exp\left(\frac{\{t^* - f(\mathbf{u}^*, \mathbf{y}^*, \boldsymbol{\theta}_{MP})\}^2}{2\sigma_{t^*}^2}\right) \tag{33}$$

where $\sigma_{t^*}^2 = 1/\beta + \mathbf{g}^T\mathbf{A}\mathbf{g}$, β is a hyper-parameter related to the distribution of the output.

C. OPTIMIZATION OF HYPER-PARAMETERS AND WEIGHTS

The optimal parameters including $\boldsymbol{\alpha}, \beta, \boldsymbol{\theta}$ correspond to the maximum of the posterior distribution of parameters $p(\boldsymbol{\theta}, \boldsymbol{\alpha}, \beta|D)$. According to Bayes' rules, the maximum of the posterior distribution of parameters can also be interpreted as an error function to minimize by taking the logarithm of the likelihood function $p(D|\boldsymbol{\alpha}, \beta, \boldsymbol{\theta})$, thereby giving

$$\begin{aligned} \log p(D|\boldsymbol{\alpha}, \beta, \boldsymbol{\theta}) = &-\alpha_1 E_{W_1}^{MP} - \alpha_2 E_{W_2}^{MP} - \alpha_3 E_{W_3}^{MP} \\ &- \beta E_D^{MP} - \frac{1}{2}\ln(\det\mathbf{A}) \\ &+ \frac{W_1}{2}\ln\alpha_1 + \frac{W_2}{2}\ln\alpha_2 + \frac{W_3}{2}\ln\alpha_3 \\ &+ \frac{N}{2}\ln\beta - \frac{N}{2}\ln(2\pi) \end{aligned} \tag{34}$$

Here, $\mathbf{A}$ is the Hessian matrix $\mathbf{A} = \nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}} S_{MP}$ and can be written as

$$\mathbf{A} = \mathbf{H} + \boldsymbol{\alpha}\mathbf{I} \tag{35}$$

where $\mathbf{H} = \beta \nabla\nabla E_D$. If $\{\lambda_i\}_{i=1}^{W}$ is the eigenvalue of the matrix $\mathbf{H}$, and then the eigenvalue of the matrix $\mathbf{A}$ equals to

$$eig(\mathbf{A}) = \begin{cases} \lambda_i + \alpha_1, & 1 \le i \le W_1 \\ \lambda_i + \alpha_2, & W_1 \le i \le W_2 \\ \lambda_i + \alpha_3, & W_2 \le i \le W \end{cases} \tag{36}$$

To solve the optimal hyper-parameter $\boldsymbol{\alpha}$, we should calculate the partial differential equations of (34) with respect to $\boldsymbol{\alpha}$. Firstly, differential coefficient $d \ln(\det \mathbf{A})/d\alpha_j$ should be computed.

$$\begin{aligned} \frac{d}{d\alpha_j} \ln(\det \mathbf{A}) &= \frac{d}{d\alpha_j} \ln\left(\prod_i (\lambda_i + \alpha_j)\right) \\ &= \frac{d}{d\alpha_j} \sum_i \ln((\lambda_i + \alpha_j) = \sum_i \frac{1}{\lambda_i + \alpha_j} \end{aligned} \tag{37}$$

If we neglect the influence of $d\lambda_i/d\alpha_j$, (34) reaches the maximal value, when the following equation is found.

$$2\alpha_j E_{W_j}^{MP} = W_j - \sum_{i=1}^{W_j} \frac{\alpha_j}{\lambda_i + \alpha_j} \tag{38}$$

If we define

$$\gamma_1 = \sum_{i=1}^{W_1} \frac{\lambda_i}{\lambda_i + \alpha_1} \tag{39}$$

$$\gamma_2 = \sum_{i=W_1}^{W_2} \frac{\lambda_i}{\lambda_i + \alpha_2} \tag{40}$$

$$\gamma_3 = \sum_{i=W_2}^{W} \frac{\lambda_i}{\lambda_i + \alpha_3} \tag{41}$$

Substitute (39) - (41) into (38), (38) can be rewritten as

$$2\boldsymbol{\alpha} E_W^{MP} = \gamma_1 + \gamma_2 + \gamma_3 \tag{42}$$

Now, we consider the maximization of (34) with respect to β. Firstly, differential coefficient $d \ln(\det \mathbf{A})/d\beta$ should be computed.

$$\begin{aligned} \frac{d}{d\beta} \ln(\det \mathbf{A}) = \frac{d}{d\beta} \sum_{i=1}^{W_1} \ln(\lambda_i + \alpha_1) + \frac{d}{d\beta} \sum_{i=W_1}^{W_2} \ln(\lambda_i + \alpha_2) \\ + \frac{d}{d\beta} \sum_{i=W_2}^{W} \ln(\lambda_i + \alpha_3) \end{aligned} \tag{43}$$

Since λ_i is the eigenvalues of matrix $\mathbf{H} = \beta\nabla\nabla E_D$, λ_i is proportionate to β. And then

$$\begin{aligned} 2\beta E_D^{MP} &= N - \sum_{i=1}^{W_1} \frac{\lambda_i}{\lambda_i + \alpha_1} - \sum_{i=W_1}^{W_2} \frac{\lambda_i}{\lambda_i + \alpha_2} - \sum_{i=W_2}^{W_3} \frac{\lambda_i}{\lambda_i + \alpha_3} \\ &= N - (\gamma_1 + \gamma_2 + \gamma_3) \end{aligned} \tag{44}$$

Based on the above derivation, the update equations of $\boldsymbol{\alpha}$ and β are obtained.

$$\alpha_i^{new} = \gamma_i / 2E_{W_i}, \quad \beta^{new} = (N - \gamma)/2E_D \tag{45}$$

where $\gamma = \gamma_1 + \gamma_2 + \gamma_3$.

The optimal parameters $\boldsymbol{\theta}_{MP}$ and the hyper-parameters $\boldsymbol{\alpha}$ and β can be optimized by an iterative process based on the scale conjugate gradient. The detailed implementation process that can be found in [17] will not go into details.

V. EXPERIMENTS AND RESULTS

In this study, one steel enterprise in China is selected as the research object, and the historical data of the blast furnace gas system is adopted for experiments. Firstly, some prediction experiments are conducted for the total generation amount of BFG, the total consumption amount of BFG and the gas holder level of BFG, which can be used for monitoring the variation of the system in the future. Secondly, according to the prediction results, knowledge-based optimal control technique is employed for adjusting some unbalance conditions of the BFG system by using the prediction results as the input of the knowledge-based control model.

A. SOME PREDICTION EXPERIMENTS

First, we conduct some prediction experiments to demonstrate the reliability of the proposed prediction model for the generation amount, consumption amount, gas holder level of the BFG system. The generation amount and the consumption amount (excluding the adjustable users) belong to the time series prediction. The prediction is completed by the two ESN individuals. The embedding dimension of the two ESN individuals of the prediction model is set as 60 and the number of the neurons in the dynamic reservoir of the ESN is set as 100. The number of the adjustable users of the BFG system equals to 8, including the #1,2,3,4 boilers of the power plant (corresponding to #1,2,3,4 GE in Fig. 1), the CCPP power station, low pressure boiler (LPB) and #1,2 synthetizing units. The parameters of the prediction model are optimized by the Bayesian regularization method.

The prediction results of the total generation amount, and the consumption amount of the BFG system are shown in Figs. 5 and 6, respectively. From the Figs. 5 and 6, the proposed method can predict the trend of the total generation amount and consumption amount of the BFG system. Although, there are some prediction errors, the prediction results can be used for the knowledge-based control model.

As for the prediction of the gas holder level, it belongs to the factor-based prediction. The gas holder level is determined by its own state in the precious moment, the total generation amount, the total consumption amount and the consumption of the adjustable users. In this study, there are

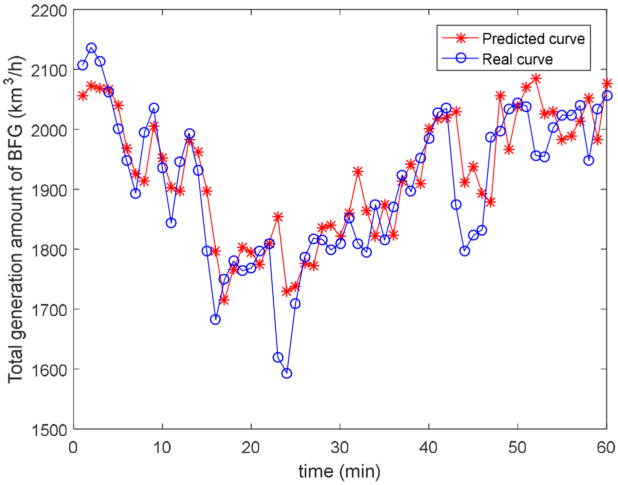

FIGURE 5. The prediction results of total generation amount of BFG.

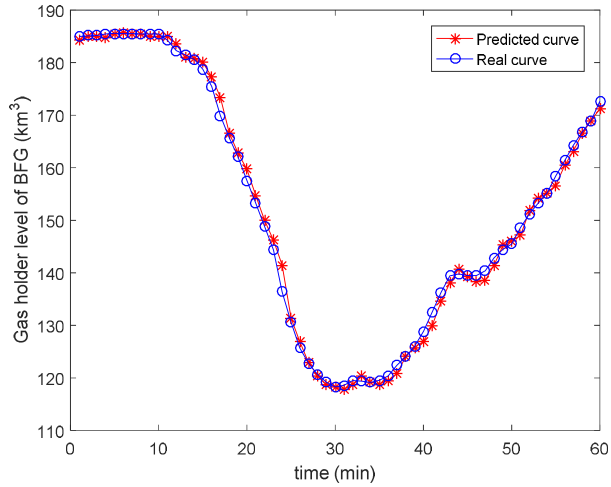

FIGURE 7. The prediction results of the gas holder level of BFG.

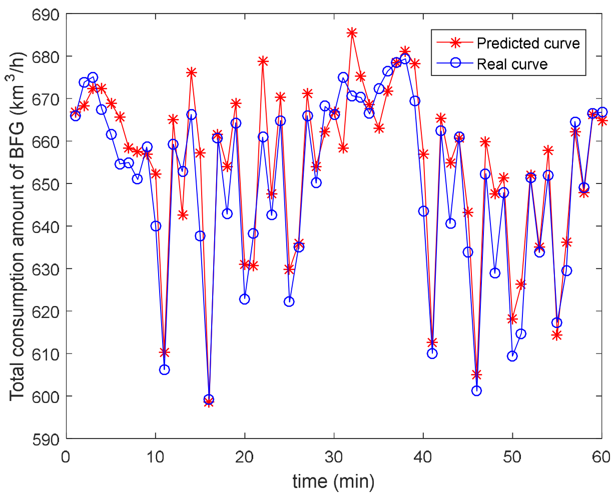

FIGURE 6. The prediction results of total consumption amount (excluding the adjustable users) of BFG.

eleven influential factors of the gas holder level. The prediction results of the gas holder level of the BFG system are shown in Fig. 7, from which we can see that the gas holder level is certainly determined by these influential factors. The generation amount and the consumption amount (excluding the adjustable users) cannot be controlled, since they are related to the production process. All the adjustable users can be viewed as the controllable object, i.e., the consumption of the adjustable users can be adjusted when the BFG system is unbalance.

B. KNOWLEDGE-BASED CONTROL EXPERIMENTS

Knowledge-based control technique is based on the fuzzy rules extracted from the large scale of operational data of the domain expert. According to the prediction results, we can monitor the variation of the system. If the performance of the system cannot satisfy the demand of the production, a control scheme is required for the system. In this case, knowledge-based control scheme can be obtained by fuzzy inference. The prerequisite of the fuzzy inference is the states of the users of the system. As for the BFG system, the outputs of the fuzzy inference are the states of the adjustable users, based on which the adjusted amount of the adjustable users can be calculated. The inputs of the fuzzy inference include the gas holder level, the total generation amount, the total consumption amount, the original states of the adjustable users. Notably, the states of the adjustable users cannot be changed without human interruption. The control objective is the economic cost of the BFG system, including the consumption of the outsourcing energy, the generation of the heat energy and power energy, the waste of the byproduct energy and the safety.

Firstly, we choose one adjustment point in the historical data to make experiments. The goal of optimization is to reduce the economic cost and control the gas holder level of the BFG system below 225 km^3, simultaneously. After prediction, the gas holder level will exceed 225 km^3. Moreover, the efficiency of the energy utilization is forecasted based on the prediction results with low performance. Thus, the extracted knowledge base is employed here to obtain one effective control scheme. Based on the fuzzy rules-based inference, the control scheme is reasoned as shown in Table 1. From Table 1, the adjusted amount based on the proposed method is obviously less than the original expert-based adjusted amount in the industrial field. If the economic cost can be reduced after the proposed method-based adjustment, the proposed method is effective and at least superior to the most-common-used method in the industrial field.

The goal of the control scheme is to save the economic cost, which reflects in the following aspects shown in Table 2, such as the reduced outsourcing energy, the reduced gas consumed by boiler, the increased steam generated by boiler, the increased power. From Table 2, the proposed method can achieve one much better performance. It is obvious that the proposed method can save 31.25kg standard coal equivalent

TABLE 1. The control scheme of the proposed method.

Adjustable units	The adjusted amount based on the proposed method (km^3/h)	The human-based adjusted amount (km^3/h)
#1 boiler	-2×20	-3×20
#2 boiler	0	0
#3 boiler	0	0
#4 boiler	0	
CCPP PS	-25.13	-20
LPB	-19.15	-15
#1 synthesizing	0	0
#2 synthesizing	0	0
Total	-84.28	-95

TABLE 2. The optimal results produced by the control scheme.

Optimized indexes	Proposed method	The human-based mode
Reduced outsourcing energy	0	0
Reduced gas consumed by boilers	$26.87km^3$	$23.15km^3$
increased steam generated by boilers	$5037.71km^3$	$4476.62\ km^3$
Increased power generated by turbines	1742. 37kwh	1554.26kwh
Increased energy efficiency	0.17	0
Saved standard coal equivalent	**31.25kg**	0

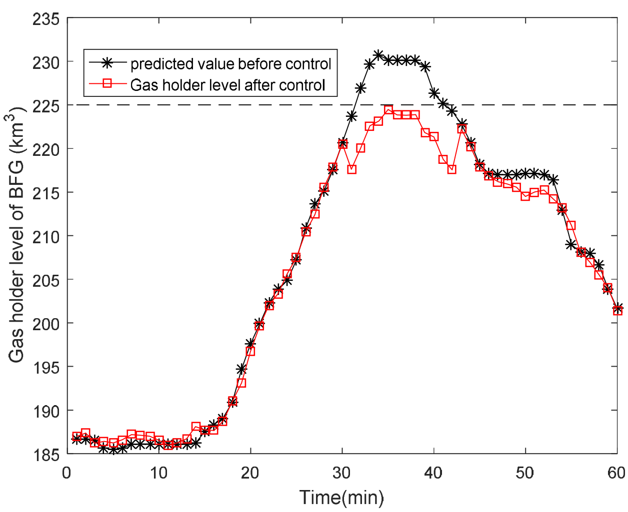

FIGURE 8. The effect of the gas holder level after adjustment.

compared to the original human-based mode. This is because the human-based control mode is carried out based on the experiences of one domain expert while the proposed method integrates the knowledge and experiences coming from many experts and provides one comprehensive judgment. Besides the results shown in Table 2, the operational effect of the gas holder level after the implementation of the control scheme is shown in Fig. 8. From Fig. 8, the gas holder level is predicted to exceed 225 km^3 and reach 230 km^3. After the implementation of the control scheme based on the proposed method, the gas holder level can be reduced in the desired region.

VI. CONCLUSIONS

Aiming at the optimal control problem of byproduct gas system in steel industry, a combination of knowledge-driven control technique and NNs ensemble-based prediction is proposed in this study. First, the community detection of complex network-based samples selection method is proposed to partition the different operational conditions of blast furnace gas system. Meanwhile, the typical and valuable data samples are selected from the original large-scale dataset. Second, a fuzzy model is designed to extract the expert scheduling knowledge from the historical data of the industrial process after community detection. And then, a great deal of scheduling knowledge is employed to compose a fuzzy rule base, which can be used for fuzzy inference of control scheme with a new input. Third, data-driven NNs ensemble is built to model the blast furnace gas system for prediction. Based on the prediction results, the control scheme can be reasoned by using the fuzzy rule base constructed above. The most important advantage of the proposed method is that the feasible control scheme can be obtained under any operational circumstance due to the introduction of expert knowledge. Theoretically, a more reasoning and optimal operation scheme can be achieved with the consideration of both the expert knowledge and the prediction results. Finally, a byproduct gas system of one steel industry is studied for experiments to verify the effectiveness and practicability of the proposed method, which shows the proposed technique is very meaningful to the energy reservation and emission reduction of industrial enterprises, and it can be used to deal with more complicated problems.

REFERENCES

[1] W. Feng and L. Cong, "Energy management system of entire iron and steel plants," *Control Eng. China*, vol. 12, no. 6, pp. 597–600, 2005.
[2] J. H. Kim, H. S. Yi, and C. Han, "A novel MILP model for plant-wide multi-period optimization of byproduct gas supply system in the iron- and steel-making process," *Chem. Eng. Res. Des.*, vol. 81, no. 8, pp. 1015–1025, 2003.
[3] H. Kong, E. Qi, and H. Li, "An MILP model for optimization of byproduct gases in the integrated iron and steel plant," *Appl. Energy*, vol. 87, no. 7, pp. 2156–2163, 2010.
[4] H. N. Kong, "A green mixed integer linear programming model for optimization of byproduct gases in iron and steel industry," *J. Iron Steel Res. Int.*, vol. 22, no. 8, pp. 681–685, 2015.
[5] W. Q. Sun, J. J. Cai, and J. Song, "Plant-wide supply-demand forecast and optimization of byproduct gas system in steel plant," *J. Iron Steel Res. Int.*, vol. 20, no. 9, pp. 1–7, 2013.
[6] X. Zhang, J. Zhao, and W. Wang, "Multi-output least squares support-vector-machine for level prediction in Linz Donaniz gas holder [J]," *Control Theory Appl.*, vol. 27, no. 11, pp. 1463–1470, 2010.
[7] X. Zhang, J. Zhao, and W. Wang, "An optimal method for prediction and adjustment on byproduct gas holder in steel industry," *Expert Syst. Appl.*, vol. 38, no. 4, pp. 4588–4599, 2011.
[8] J. Zhao, W. Wang, Y. Liu, and W. Pedrycz, "A two-stage online prediction method for a blast furnace gas system and its application," *IEEE Trans. Control Syst. Technol.*, vol. 19, no. 3, pp. 507–520, May 2011.

[9] J. Zhao, Q. Liu, W. Wang, and W. Pedrycz, "Hybrid neural prediction and optimized adjustment for coke oven gas system in steel industry," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 3, pp. 439–451, Mar. 2012.
[10] J. Zhao, W. Wang, and W. Pedrycz, "Online parameter optimization-based prediction for converter gas system by parallel strategies," *IEEE Trans. Control Syst. Technal.*, vol. 20, no. 3, pp. 835–845, May 2012.
[11] J. Zhao, W. Wang, and K. Sun, "A Bayesian networks structure learning and reasoning-based byproduct gas scheduling in steel industry," *IEEE Trans. Autom. Sci. Eng.*, vol. 11, no. 4, pp. 1149–1154, Oct. 2014.
[12] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 69, no. 2, p. 026113, 2004.
[13] W. Pedrycz and F. Gomide, "Rule-base fuzzy models," in *Fuzzy Systems Engineering: Toward Human-Centric Computing*, vol. 4, 1st ed. Hoboken, NJ, USA: Wiley, 2007, ch. 11, pp. 280–298.
[14] H. Jaeger and H. Haas, "Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication," *Science*, vol. 304, no. 5667, pp. 78–80, Apr. 2004.
[15] H. Jaeger, "Tutorial on training recurrent neural networks, covering BPTT, RTRL, EKF and echo state network approach," German Nat. Res. Center Inf. Technol., GMD, Sankt Augustin, Germany, Tech. Rep. 159, 2002.
[16] C. M. Bishop. "Sparse kernel machines," in *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006, ch. 7, pp. 345–357.
[17] P. Lauret and E. Fock, "Bayesian neural network approach to short time load forecasting," *Energy Convers. Manage.*, vol. 49, no. 5, pp. 1156–1166, 2008.
[18] Y. Li, S. Tong, L. Liu, and G. Feng, "Adaptive output-feedback control design with prescribed performance for switched nonlinear systems," *Automatica*, vol. 80, pp. 225–231, 2017.

HAIXIA WANG received the bachelor's degree from Dalian Jiaotong University in 2002, and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, in 2013. She is currently a Lecturer with the Key Laboratory for Robot and Intelligent Technology of Shandong Province, Shandong University of Science and Technology. Her research interests include robot technology, robot visual servo, and pattern recognition.

CHUNYANG SHENG was born in Rongcheng, China, in 1986. He received the Ph.D. degree in control theory and control engineering from the Dalian University of Technology in 2014. He has done post-doctoral research with the School of Computer Science and Technology, Dalian University of Technology, from 2015 to 2017. He is currently a Lecturer with the Shandong University of Science and Technology. His research interests include machine learning, data-based prediction, control, and optimization.

XIAO LU received the bachelor's degree from Dalian Jiaotong University in 2002, and the Ph.D. degree in control theory and control engineering from the Dalian University of Technology in 2008. He is currently an Associate Professor with the Key Laboratory for Robot and Intelligent Technology of Shandong Province, Shandong University of Science and Technology. His research interests include linear estimation, fault detection, wireless sensor networks, and robot control.