

Research Article

Application of Model-Based Deep Learning Algorithm in Fault Diagnosis of Coal Mills

Yifan Jian,¹ Xianguo Qing,¹ Yang Zhao,¹ Liang He,¹ and Xiao Qi^{1,2} 

¹Science and Technology on Reactor System Design Technology Laboratory, Nuclear Power Institute of China, Chengdu, China

²Energy and Electricity Research Center, Jinan University, Zhuhai, Guangdong, China

Correspondence should be addressed to Xiao Qi; qixiao@jnu.edu.cn

Received 5 May 2020; Revised 3 July 2020; Accepted 20 July 2020; Published 14 August 2020

Guest Editor: Mohammed Abouheaf

Copyright © 2020 Yifan Jian et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The coal mill is one of the important auxiliary engines in the coal-fired power station. Its operation status is directly related to the safe and steady operation of the units. In this paper, a model-based deep learning algorithm for fault diagnosis is proposed to effectively detect the operation state of coal mills. Based on the system mechanism model of coal mills, massive fault data are obtained by analyzing and simulating the different types of faults. Then, stacked autoencoders (SAEs) are established by combining the said data with the deep learning algorithm. The SAE model is trained by the fault data, which provide it with the learning and identification capability of the characteristics of faults. According to the simulation results, the accuracy of fault diagnosis of coal mills based on SAE is high at 98.97%. Finally, the proposed SAEs can well detect the fault in coal mills and generate the warnings in advance.

1. Introduction

The coal mill is one of the important auxiliary equipment of coal-fired units, and its operating status is directly related to the safe and stable operation of the units. When a fault occurs in the coal mill, the fuel supply of the boiler cannot be guaranteed which creates the mismatch between boiler energy output and the turbine power output. Under this situation, a quick load rejection operation will occur, which directly leads to fire extinguishing in the furnace. The fault in the coal mill will cause large economic loss to power generation enterprises and decrease the safety and stability of the power system. Therefore, it is of great necessity to guarantee the normal operation through effective fault warning and diagnosing of coal mill.

Agrawal et al. [1] divided the fault diagnosis methods into three categories: model-, signal-, and historical operation data-based fault diagnosis methods. Model-based fault diagnosis methods need to establish the mathematical model of the coal mill. Odgaard and Mataji [2] used a simplified energy balance equation to monitor and diagnose abnormal energy flow in the coal mill. Andersen et al. [3] designed a Kalman filter to estimate the moisture in the coal that enters

and exists a coal mill to determine whether the energy in the coal mill is in normal condition. Based on the multisegment model of coal mills established by Wei et al. [4], Guo et al. [5] realized the monitoring of the state of coal mills by identifying the abnormal variation in the model parameters. Model-based fault diagnosis methods analyze the mathematical model of the actual object for fault diagnosis, and thus, the physical meaning is clear. However, establishing the exact model in practical application is difficult [6–10]. Thus, the operability of these methods is poor.

Signal-based fault diagnosis systems are widely used to evaluate the health of mechanical equipment. Many signals of high frequency change during the operation of the coal mill, such as current of coal mill, outlet primary air flow of coal mill, and differential pressure of primary air. Su et al. [11] designed a system that records the vibration signals of the coal mill and shifts them to energy amplitudes by use of wavelet analysis. Whether the coal mill is in coal interruption or coal choking, other fault operations can be determined by analyzing the relationship between the vibration signals and the amount of coal in the mill. Kisić et al. [12] proposed a method to detect the wear degree of grinding roller and analyzed the multivariate control chart on

the frequency spectrum to find the appropriate time to replace the worn parts. Collura et al. [13] utilized model identification and signal processing techniques to develop a coal mill performance monitoring tool based on real-time detection of the fineness of pulverized coal. Compared with the model-based fault diagnosis method, the signal-based fault diagnosis method does not need to establish complex object model. Only through the analysis of collected data can a fault in the system be found. However, these methods often need to install a large number of sensors to collect signal, thereby resulting in high implementation and maintenance costs.

Fault diagnosis based on historical operation data is mainly done by analyzing the differences between the normal operation data and fault operation data to determine the health status of coal mill. Han and Jiang [14] proposed a fault diagnosis method based on fuzzy decision clustering and used a single-layer neural network to realize three kinds of fault identification of coal mill. Qin et al. [15] utilized the abnormal operation data of the coal mill to establish an expert system to determine the operating status of the coal mill by comparing the trend of the model output with the expert system. A data-based fault analysis method is a data-driven approach, and even researchers who are unfamiliar with the system can use relevant algorithms for analysis. However, fault types and fault data in the mass historical data of the thermal power units are incomplete and a data-driven method requires analyzing a large amount of fault data. Thus, selecting the fault data from the vast amount of historical data one by one is difficult [16–18].

The model-based fault diagnosis method needs to establish an accurate model of the coal mill in order to obtain good fault diagnosis results. However, the coal mill is a complex object with multiparameter coupling. It is difficult to establish an accurate mathematical model. The premise of applying signal-based fault diagnosis methods is to be able to measure the monitored parameters. Therefore, it is necessary to install a large number of new sensors on the shell of the coal mill. However, when the coal mill was initially constructed, it usually did not consider reserving the mechanical interface for new sensors. So, it is not easy to install new sensors on the shell of the coal mill. The fault diagnosis method based on historical operation data firstly needs to obtain a large amount of fault operation data of the coal mill. However, the fault data of the coal mill is usually mixed with the normal operation data, which is difficult to classify and identify. Based on the above analysis, the existing methods are difficult to achieve good application results for the fault diagnosis of coal mills. Although the above three types of traditional methods have shortcomings, combining their advantages can find a simpler and more effective method to solve the fault diagnosis of coal mills. The basic idea is to obtain fault simulation data based on a simplified model and use big data analysis for fault identification. In recent years, the rapid development of deep learning algorithms has provided the possibility of big data analysis. Guo et al. [19] constructed an adaptive convolution neural network, which greatly improves accuracy of fault diagnosis of motor bearing. Duan et al. [20] used a deep learning algorithm to study the missing traffic data to implement the interpolation

of missing data. In this study, a model-based data-driven fault diagnosis method is proposed to obtain a fault diagnosis method with simple operation, low cost, and high accuracy. First, on the basis of the simplified coal mill model, massive fault data of the coal mill are obtained by analyzing the fault principle and simulating the fault operation status of the coal mill. This method solves the difficulty in obtaining a large amount of fault data manually from the massive data. Then, stacked autoencoders (SAEs) with multilayer neural networks are established on the basis of the theory of deep learning algorithm. The numerous fault data obtained by the steps above are used in the training of the networks to fully motivate the nonlinear characteristics of deep neural networks. Thus, the built network can accurately learn the essential characteristics of all kinds of faults and then achieve the early warning and diagnosis of the fault in coal mills. The above method can greatly improve the fault diagnosis accuracy of the coal mill, and at the same time, it can also provide fault warning to the operator, which is of great significance for ensuring the safe operation of the power plant and ensuring the safety of the equipment.

The rest of the paper is organized as follows. Section 2 introduces the working principle of the coal mill and its nonlinear dynamic model. Section 3 analyzes the mechanism of two typical coal mill faults and obtains a large number of fault data by simulation experiments. Section 4 introduces the working principle of the SAEs and makes certain improvements to the model. Section 5 is the simulation analysis that aims to verify the effectiveness of the proposed method in the fault diagnosis of the coal mill. Section 6 elaborates the conclusions of the study.

2. Brief Introduction of the Coal Mill System

2.1. Working Principle. MPS-type medium-speed coal mill [21] is a roller-type coal mill designed and manufactured by Babcock, Germany. Such mills are characterized by smooth output, low energy consumption, and long maintenance period. In this study, MPS180-HP-II medium-speed coal mill is used in the analysis. The maximum output is 44.496 t/h, and the fineness of coal powder R90 is 22% (Figure 1). R90 indicates the probability that coal powders cannot pass through a sieve with a pore size of 90 μm .

The raw coal falls into the coal mill through the coal dropping pipe and is milled into coal powder under the squeezing effect of two milling parts (grinding disks and rollers) [4]. The primary air enters the coal mill through the annulus around the grinding disk to dry the coal powders and bring them into the coarse coal separator for separation. The qualified fine coal powders are blown into the boiler for combustion while large ones return into the coal for subsequent milling.

2.2. Mathematical Model of the Coal Mill. The operation of the coal mill involves the mass balance of coal and the energy balance of the entire coal mill. Establishing an effective dynamic mathematical model of coal mills is an important prerequisite for the state monitoring of coal mills. Zeng et al. [22–24] established an MPS medium-speed mill model (equation (1)),

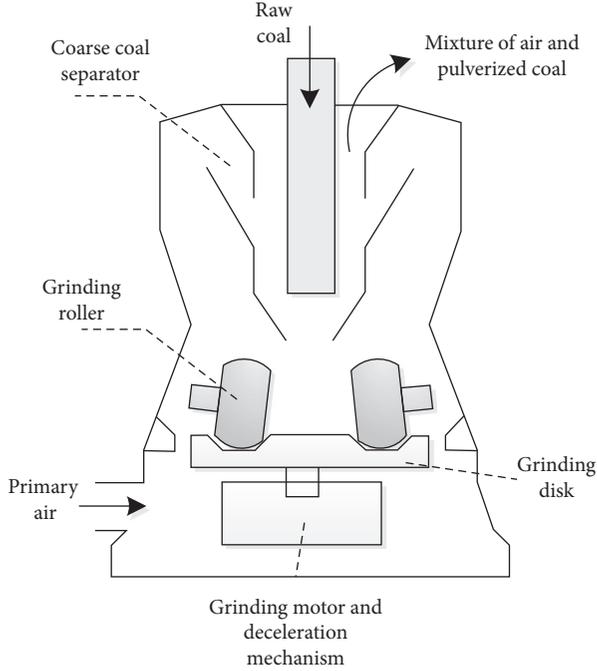


FIGURE 1: Schematic structure of the MPS medium-speed mill.

which includes three inputs and three outputs based on the mass and energy balance of the primary air and coal moisture in the mill. The proposed method is based on this model, and the symbolic description of the model is shown in Table 1:

$$\begin{aligned}
 \dot{W}_{\text{air}} &= \frac{1}{T_1} (-W_{\text{air}} + W_L^{\text{max}} u_L + W_H^{\text{max}} u_H), \\
 \dot{T}_{\text{in}} &= \frac{1}{T_2} \left[-T_{\text{in}} + \frac{C_L W_L^{\text{max}} u_L T_L + C_H W_H^{\text{max}} u_H T_H}{C_{\text{in}} (W_L^{\text{max}} u_L + W_H^{\text{max}} u_H)} \right], \\
 \dot{M}_c &= W_c - K_{10} M_c, \\
 \dot{M}_{pf} &= K_{10} M_c - W_{pf}, \\
 W_{pf} &= K_{11} \Delta P_{pa} M_{pf}, \\
 \Delta P_{pa} &= \frac{22.4}{28.8} \cdot \frac{273 + T_{\text{in}}}{273} \cdot \left(\frac{W_{\text{air}}}{10} \right)^2, \\
 I &= K_6 M_{pf} + K_7 M_c + K_8, \\
 \dot{T}_{\text{out}} &= [K_1 T_{\text{in}} + K_2] W_{\text{air}} + K_3 W_c \\
 &\quad - [K_4 T_{\text{out}} + K_5] [W_{\text{air}} + W_c] + K_9 I \\
 &\quad + K_{12} T_{\text{out}} - K_{14} W_{\text{free}}^{\text{water}}, \\
 \dot{M}_{pc} &= \frac{1}{M_c + M_{pf}} (M_{ar} W_c - W_{\text{free}}^{\text{water}} - M_{pc} W_{pf}), \\
 W_{\text{free}}^{\text{water}} &= K_{13} (W_c M_{ar}) T_{\text{out}} \left(1 - e^{(W_{\text{air}}/K_{15})} \right).
 \end{aligned} \tag{1}$$

TABLE 1: Nomenclature.

Symbol	Meaning
W_L^{max}	Maximum flow of cold air (kg/s)
W_H^{max}	Maximum flow of hot air (kg/s)
u_L	Valve position of cold air
u_H	Valve position of hot air
T_L	Temperature of cold air ($^{\circ}\text{C}$)
T_H	Temperature of hot air ($^{\circ}\text{C}$)
ΔP_{pa}	Differential pressure of primary air (mbar)
C_{in}	Specific heat capacity of mixed primary air (kJ/(kg $\cdot^{\circ}\text{C}$))
C_L	Specific heat capacity of cold air (kJ/(kg $\cdot^{\circ}\text{C}$))
C_H	Specific heat capacity of hot air (kJ/(kg $\cdot^{\circ}\text{C}$))
T_{in}	Inlet primary air temperature of coal mill ($^{\circ}\text{C}$)
W_{air}	Inlet primary air flow of coal mill (kg/s)
W_{pf}	Outlet pulverized coal flow of coal mill (kg/s)
M_{pf}	The mass of pulverized coal in coal mill (kg)
W_c	Inlet coal flow of coal mill (kg/s)
M_c	The mass of raw coal in coal mill (kg)
T_{out}	Outlet temperature of coal mill ($^{\circ}\text{C}$)
$W_{\text{free}}^{\text{water}}$	Evaporation of coal moisture (kg/s)
M_{pc}	Pulverized coal (%)
M_{ar}	Coal moisture (%)
I	Current of coal mill (A)
T_1	Delay time of the valve position to primary air flow (s)
T_2	Delay time of valve position to primary air temperature (s)
K_i	Identified model parameters ($i = 1, 2, \dots, 15$)

In Equation (1), U_L , U_H , and W_c are the control quantities of the model; W_{air} , T_{out} , and W_{pf} are the output quantities of the model; and K_i and T_i ($i = 1, 2, \dots, 15, j = 1, 2$) are the model parameters to be identified and of which the values are shown in Table 2.

3. Model-Based Coal Mill Fault Simulation

The fault types and fault data in the vast amount of historical data of the thermal power units are incomplete, and selecting the fault data one by one from the massive historical data is difficult. Therefore, effectively obtaining a large number of fault data is the key to solve the fault diagnosis of coal mill. The simulation results in [22] showed that the mathematical model of MPS-type medium speed coal mill presents high precision. In the current study, the coal mill model is used in the analysis and two typical coal mill faults (coal interruption and coal choking) are simulated by analyzing the fault mechanism of coal mill. The simulation experiments obtain a large number of fault data, which can effectively solve the difficulty in obtaining fault data manually from the massive data.

First, a control scheme is designed for the coal mill model. The purpose is to ensure that the simulation experiments are conducted under the closed loop regulation, such that the fault data obtained by the simulation experiments can be significantly close to the real operation status of coal mill. The control scheme is shown in Figure 2. The entire control scheme consists of three controlled, three control, and four state variables as presented in Table 3. The control circuits are composed of three single-loop

TABLE 2: Identified model parameters.

Parameter	Value
K_1	0.00069
K_2	0.19549
K_3	0.00999
K_4	0.00109
K_5	0.09338
K_6	0.17999
K_7	0.88836
K_8	34.2065
K_9	0.01656
K_{10}	0.41378
K_{11}	0.07005
K_{12}	-0.05987
K_{13}	0.01152
K_{14}	0.26254
K_{15}	14.9867
T_1	10.3004
T_2	3.6765

proportion-integral-differential (PID) controllers. Specifically, PID1, where the setting parameters, respectively, are $K_p=1$, $K_i=0.05$, $K_d=0$, controls the outlet temperature of coal mill by adjusting the valve position of cold air. PID2, where the setting parameters, respectively, are $K_p=2$, $K_i=0.5$, $K_d=0$, controls the inlet primary air flow of coal mill by adjusting the valve position of hot air, and PID3, where the setting parameters, respectively, are $K_p=0.1$, $K_i=0.1$, $K_d=0$, controls the outlet pulverized coal flow of coal mill by adjusting the inlet coal flow of coal mill. After designing the control scheme, the fault operation status of the coal mill can be simulated by adjusting the corresponding controllers.

3.1. Fault Simulation of Coal Interruption. When an obstruction exists in the coal dropping pipe or a fault occurs in the coal feeder, the amount of coal into the coal mill will reduce directly and coal interruption will occur when the case is serious, thereby endangering the stability of the boiler combustion. The process of simulating coal interruption is as follows. When the coal mill is in stable operation, a negative slope signal is superimposed on PID3, such that the mass of coal entering the coal mill is gradually reduced to 0. The data generated in this process can be considered the coal interruption samples. A large number of coal interruption samples can be obtained by adjusting the set value to run the coal mill in other operation status and repeating the steps above to record fault data.

To verify the effectiveness of the simulation experiments of coal interruption, the variables that change significantly during the period of coal interruption are selected and their varying curves are drawn. Figure 3 shows the result of an arbitrary selection of experimental data. Figure 3(a) shows that coal interruption decreases the mass of coal entering the coal mill, which then decreases the outlet pulverized coal flow of the coal mill. Meanwhile, the heat consumption of the inlet primary air flow of coal mill through the coal mill reduces, thereby resulting in an upward trend in the outlet temperature of coal mill; the valve position of cold air is then

rapidly opened, thereby making the outlet temperature of coal mill fall (Figure 3(b)). The reduction in the mass of coal stored in the coal mill results in the reduction of the current of coal mill and the differential pressure of primary air. The trend is consistent with that of the curves described in Figures 3(c) and 3(d). The ramp signal is removed after 75 s, and the variables are returned to the original set value under the control of the controller.

The research object in [19] is a MPS-type medium speed coal mill in a power plant in Hainan, China. The current study obtains sets of fault data of coal interruption by looking for the historical operation data of the coal mill and draws the varying curves of key variables as shown in Figure 4. According to the accident analysis, the coal interruption fault occurs because of the malfunctioning of the coal feeder; as a result, the actual supply of coal gradually reduces to 0 (Figure 4(a)). Figure 4 shows that, when coal interruption fault occurs, the outlet temperature of the coal mill rises (Figure 4(b)); however, the current of coal mill (Figure 4(c)) and the differential pressure of primary air (Figure 4(d)) decrease. This changing trend is similar to that of the key variables in the simulation of coal interruption. Therefore, the simulation of coal interruption in this study is reasonable. Accordingly, the data in the rectangular frame in Figure 3 can be recorded as fault samples.

3.2. Fault Simulation of Coal Choking. Coal choking may be caused by too little inlet primary air flow of coal mill, excessive coal feed, or too much moisture in raw coal. The process of simulating coal choking is as follows. A positive step signal is superimposed on PID3 to make the mass of coal in coal mill quickly reach the upper limit, and the data are recorded as the coal choking samples. Similar to previous simulation experiments, a large number of coal choking samples can be obtained by adjusting the set value to run the coal mill in other operation status and repeating the steps above to record fault data. The upper limit of the mass of coal stored in the coal mill is set as 60 kg.

The varying curves of the variables are drawn, and Figure 5 shows the result of an arbitrary selection of experimental data. As shown in the figure, the sudden increase in the set value of the outlet pulverized coal flow of coal mill causes the mass of raw coal in the coal mill to rise continuously (Figure 5(a)) and increases the resistance along the way, which results in the increase in differential pressure of primary air (Figure 5(c)). At the same time, the work load of coal mill increases accordingly, such that the current of the coal mill increases as well (Figure 5(b)). When the thickness of the raw coal reaches a certain degree, the grinding efficiency drops significantly, which then reduces the current of the coal mill. The outlet pulverized coal flow of coal mill is reduced to 0 until the mass of raw coal in the coal mill reaches the upper limit, at which time the primary air pipe is blocked and the pulverized coal cannot be blown out (Figure 5(d)). The analysis above shows that the simulation experiment results are consistent with the fault characteristics of coal choking; therefore, the data in the rectangular frame can be used as fault samples.

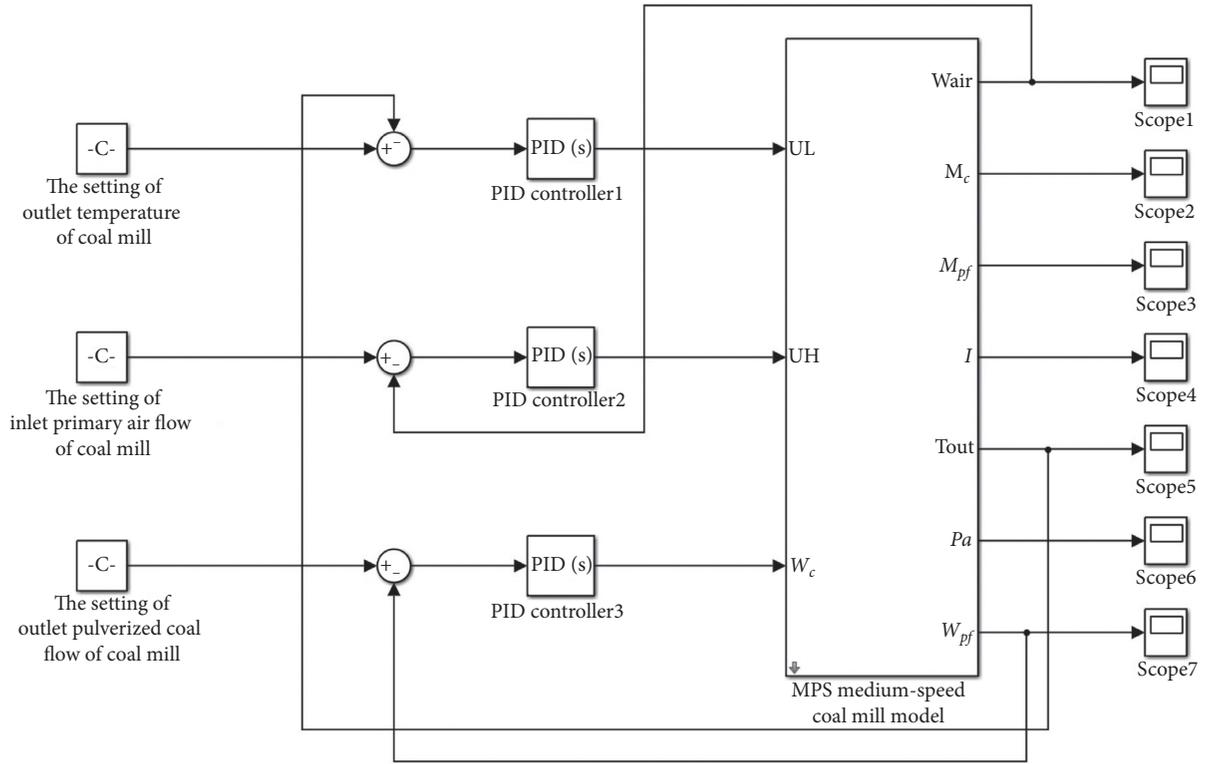


FIGURE 2: Control scheme for fault simulation of the coal mill.

TABLE 3: Variables declaration of control scheme.

Variable	Symbol
Controlled variable	T_{out}
	W_{air}
	W_{pf}
Control variable	u_L
	u_H
	W_c
State variable	M_c
	M_{pf}
	I
	ΔP_{pa}

4. Stacked Autoencoders

The fault diagnosis method based on historical operation data is a data-driven approach, which aims to obtain the nonlinear mapping relationship between the data and fault features. When sufficient data are available for learning, the deep neural networks can theoretically approximate any nonlinear function. This section describes a deep neural network called SAE, which is stacked by autoencoders (AEs), for fault diagnosis of coal mill and proposes two ways to improve the network performance.

4.1. *Fundamentals of Autoencoder.* An AE neural network can be considered a three-layer neural network. This network applies unsupervised learning algorithm to train and adjust the network weight and ultimately sets the network output to be equal to the network input. A typical example is

shown in Figure 6, where $\{x_1, x_2, \dots, x_n; x_i \in R^n\}$ can be treated as a set of unlabeled raw data and $\{x'_1, x'_2, \dots, x'_n; x'_i \in R^n\}$ represents the network output. The circles with b are called bias units and correspond to the intercept term.

The transfer process of raw data from the input layer to the hidden layer is called encoding, and the transfer process from the hidden layer to the output layer is called decoding, which can be described by

$$d = S(W_1x + b_1), \tag{2}$$

$$y = S(W_2x + b_2), \tag{3}$$

where $S(\cdot)$ represents sigmoid function, W_1 represents the weight matrix between the input and hidden layers, W_2 represents the weight matrix between the hidden and output layers, and b_1 and b_2 represent the bias.

According to the concepts mentioned in this section, AE tries to learn a function $h_{w,b}(x) \approx x$. In other words, AE is trained to learn an approximate function such that the network output is similar to the network input. In fact, by putting constraints into AE, such as limiting the number of nodes in hidden layer, AE can obtain the low-dimensional feature of the data by compressing the high-dimensional input data. In Section 3, a large number of fault data have been obtained by fault simulation experiments. The remaining parts focus on the establishment of a suitable AE network to find the relationship between data and fault characteristics by effective learning of the complex fault data.

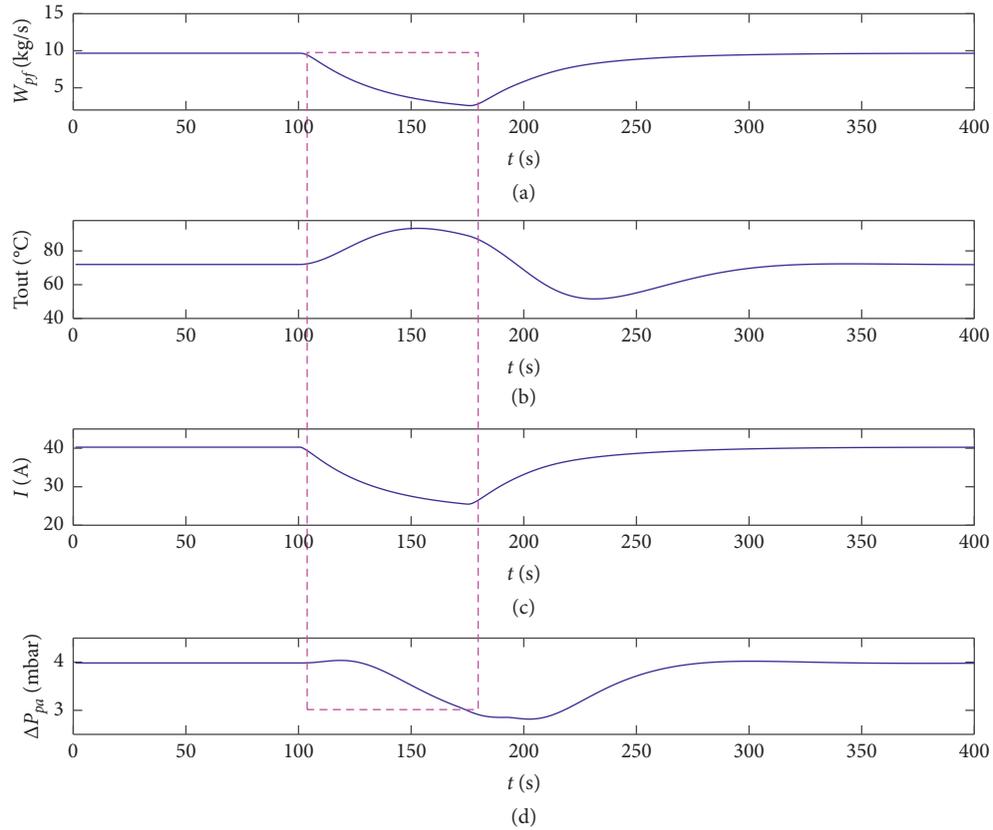


FIGURE 3: Fault simulation of coal interruption in coal mill. (a) Outlet pulverized coal flow of coal mill; (b) outlet temperature of coal mill; (c) current of coal mill; (d) differential pressure of primary air.

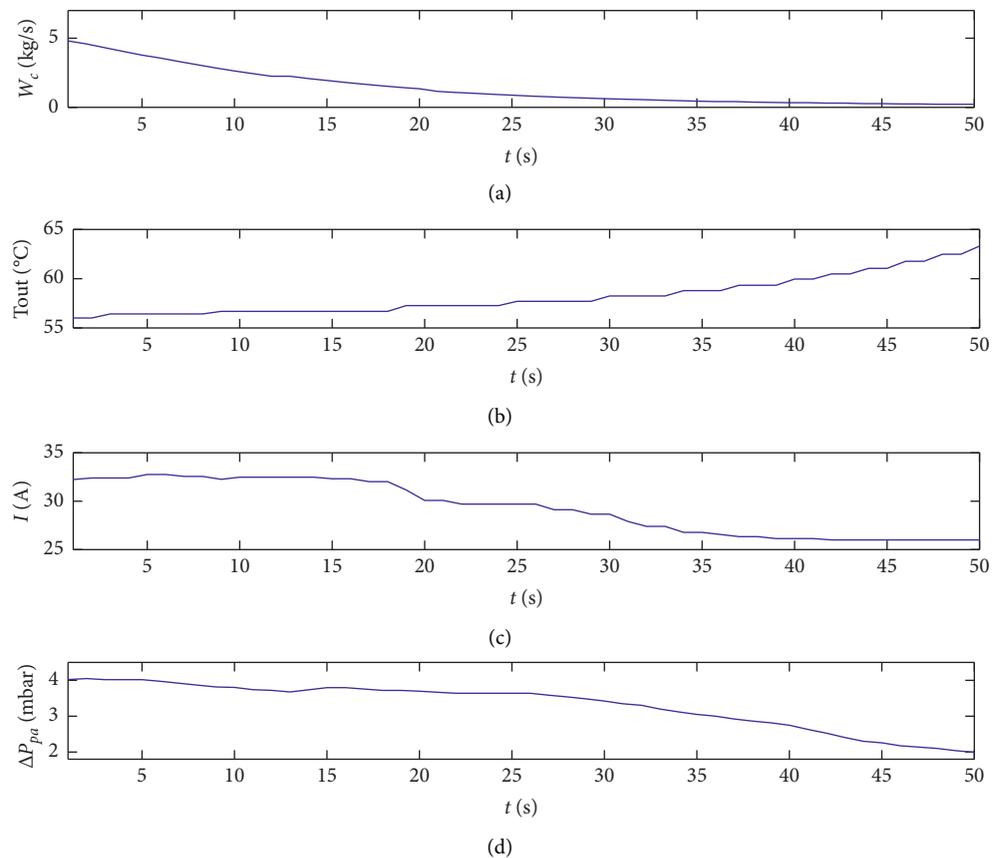


FIGURE 4: Actual fault data of coal interruption in coal mill. (a) Outlet pulverized coal flow of coal mill; (b) outlet temperature of coal mill; (c) current of coal mill; (d) differential pressure of primary air.

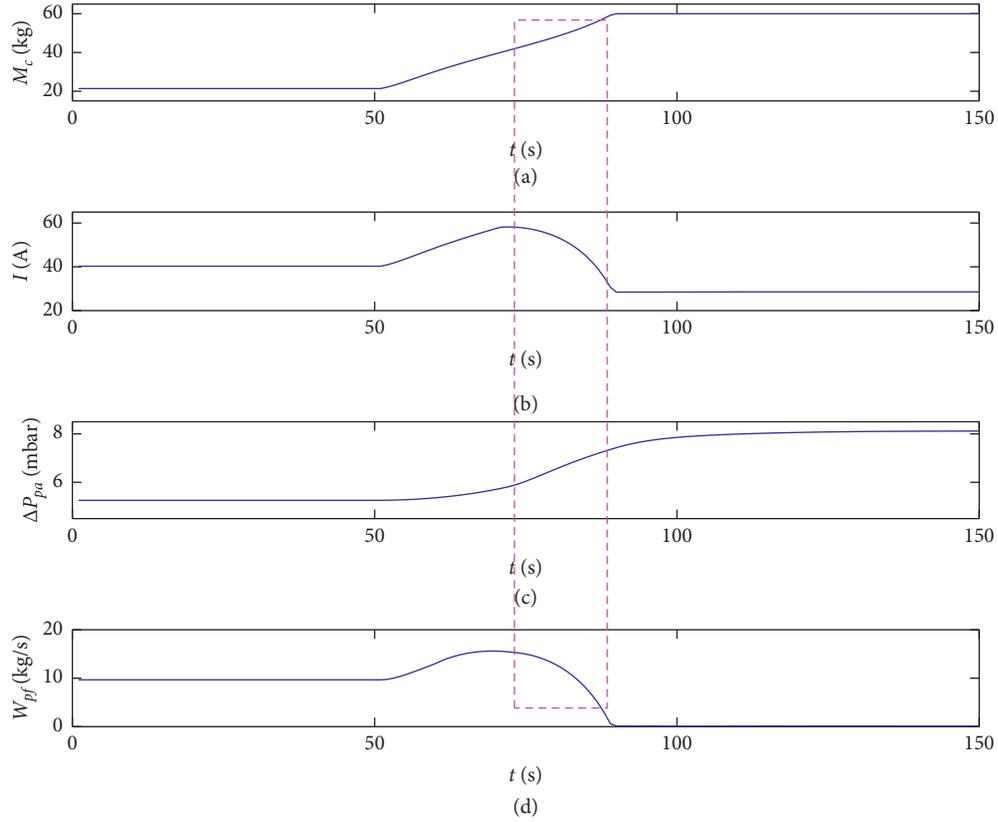


FIGURE 5: Fault simulation of coal choking in coal mill. (a) The mass of raw coal in coal mill; (b) current of coal mill; (c) differential pressure of primary air; (d) outlet pulverized coal flow of coal mill.

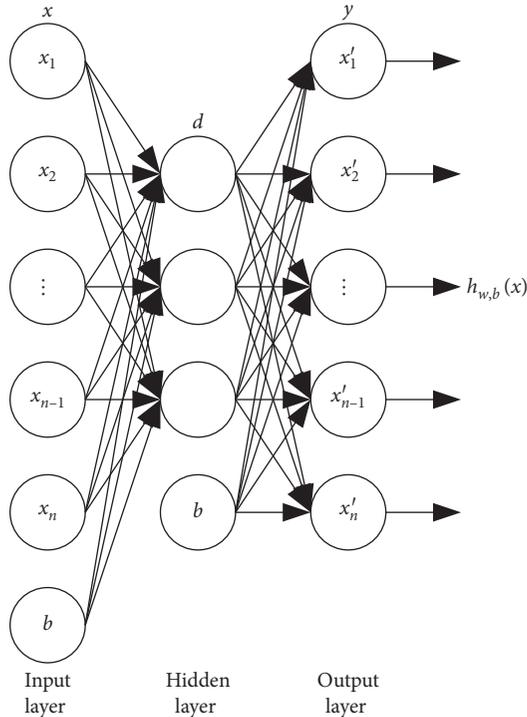


FIGURE 6: Network structure of AE.

The backpropagation algorithm is used for AE training. A training set $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$ of m training samples is assumed. The network can be trained using batch gradient descent. For a single training example (x, y) , the cost function can be defined as

$$J(W, b; x, y) = \frac{1}{2} \|h_{w,b}(x) - y\|^2. \quad (4)$$

For a training set of m samples, the overall cost function is

$$J(W, b) = \left[\frac{1}{m} \sum_{i=1}^m \left(\frac{1}{2} h_{w,b}(x) - y^2 \right) \right] + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2, \quad (5)$$

where λ represents weight decay coefficient that controls the relative importance of the two terms in equation (5). $W_{ji}^{(l)}$ represents the synaptic weight between the i -th neuron in layer l and j -th neuron in layer $l+1$. n_l represents the number of layers in AE. In other words, n_l can represent the output layer of the network, and s_l represents the number of the total neurons in layer l . The first term in the definition of $J(W, b)$ is an average sum-of-squares error term. The second term is a weight decay term that can decrease the magnitude of the weights and prevent overfitting.

The weight W and bias b are updated with gradient descent as follows:

$$W'_{ij} = W_{ij} - \alpha \frac{\partial}{\partial W_{ij}} J(W, b), \quad (6)$$

$$b'_i = b_i - \alpha \frac{\partial}{\partial b_i} J(W, b),$$

where α represents the learning rate. The partial derivatives in the equations above are derived as follows:

$$\frac{\partial}{\partial W_{ij}^{(l)}} J(W, b) = \left[\frac{1}{m} \sum_{i=1}^m \frac{\partial}{\partial W_{ij}^{(l)}} J(W, b, x^{(i)}, y^{(i)}) \right] + \lambda W_{ij}^{(l)},$$

$$\frac{\partial}{\partial b_i^{(l)}} J(W, b) = \frac{1}{m} \sum_{i=1}^m \frac{\partial}{\partial b_i^{(l)}} J(W, b, x^{(i)}, y^{(i)}), \quad (7)$$

where

$$\frac{\partial}{\partial W_{ij}^{(l)}} J(W, b; x, y) = a_j^{(l)} \delta_i^{(l+1)}, \quad (8)$$

$$\frac{\partial}{\partial b_i^{(l)}} J(W, b; x, y) = \delta_i^{(l+1)},$$

where $a_j^{(l)}$ represents the activation of unit j in layer l and $\delta_i^{(l+1)}$ represents the error term of layer $l+1$, given by

$$\delta_i^l = \left(\sum_{j=1}^{s_{l+1}} W_{ji}^{(l)} \delta_j^{(l+1)} \right) f'(z_i^{(l)}), \quad (9)$$

where $z_i^{(l)}$ represents the input weighted sum of unit i in layer l and $f'(\cdot)$ represents partial deflection of sigmoid function. The error term of the output layer n_l is given by

$$\delta_i^{(n_l)} = - \left(y_i - a_i^{(n_l)} \right) f'(z_i^{(n_l)}), \quad (10)$$

$$f'(z_i^{(n_l)}) = a_i^{(l)} (1 - a_i^{(l)}),$$

where $a_i^{(l)}$ represents the activation of unit i of layer l , $a_i^{(n_l)}$ represents the activation of unit i in the output layer, and $z_i^{(n_l)}$ represents the input weighted sum of unit i in the output layer.

Repeating the above equations can make the output of AE equal to the input of AE by minimizing the overall cost function (equation (5)).

4.2. Improvement of AE. As mentioned in Section 4.1, limiting the number of nodes in hidden layer is conducive to helping AE learn the relationship between the input data and fault features, because reducing the number of neurons can simplify the structure of the hidden layer and reduce the dimension of the input data.

Restricting the number of neurons can reduce the dimension of the data, but the network can learn few features in the hidden layer. On the basis of guaranteeing the diversity of the features in the hidden layer, a method called sparse constraint is introduced in this study to improve AE. The main idea is not to reduce the number of neurons but to

consider restrictions to limit the activities of the neurons and thus reduce the dimension of the input data. Accordingly, the original overall cost function (equation (5)) should be modified to introduce an additional penalty factor, given by

$$J_{\text{sparse}}(W, b) = J(W, b) + \beta \sum_{j=1}^s \text{KL}(\rho \parallel \hat{\rho}_j), \quad (11)$$

where

$$\text{KL}(\rho \parallel \hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j}, \quad (12)$$

where $\sum_{j=1}^s \text{KL}(\rho \parallel \hat{\rho}_j)$ represents the sparsity penalty term, β controls the weight of the sparsity penalty term, $\hat{\rho}_j$ represents the average activation of unit j in hidden layer, ρ represents a sparsity parameter, and s represents the number of units in one hidden layer.

The penalty term has the following property: if $\hat{\rho}_j = \rho$, then $\text{KL}(\rho \parallel \hat{\rho}_j) = 0$; the value increases monotonically with the difference between $\hat{\rho}_j$ and ρ . Therefore, the activations of hidden units are sufficiently small when ρ is set close to zero.

Random noise is introduced into the input data to make the network learn rich information and thus prevent the AE from learning only the equivalent representation of the original data. The main idea is to set a small number of nodes in the input layer to zero at a small probability. However, the probability of introducing random noise should be appropriate; otherwise, the noise may cause irreversible damage to the input data.

SAEs are deep neural networks consisting of multiple layers of the improved AEs in which the output of each layer is wired to the input of the next layer. The SAE model is connected with a Softmax classifier to complete the construction of the deep neural network (Figure 7). The SAE model can identify the fault in the coal mill by learning the labeled data obtained from the fault simulation experiments of the coal mill.

5. Fault Diagnosis Based on SAE

5.1. Data Preprocessing and Health State Definition. In accordance with the fault simulation method described in Section 3, the simulation experiments are conducted repeatedly, and then 5000 sets of experimental data are obtained including three kinds of data samples; namely, coal mill operates in coal broken condition (coal interruption), full-of-coal condition (coal choking), and normal condition (normal operation). To facilitate the training of the SAE model, the three different operation conditions of the coal mill are labeled. The definition is shown in Table 4. The 5000 sets of data are randomly divided into training data and test data (Table 5). At the same time, two experiments are conducted to validate the effects of the proposed fault diagnosis method. The test samples of two experiments are the same.

5.2. Establishment of the SAE Model. The optimum SAE model has important effects on the accuracy rate of fault identification. In [25, 26], the unsupervised learning effect of SAE is reported to be affected by parameters of model, such

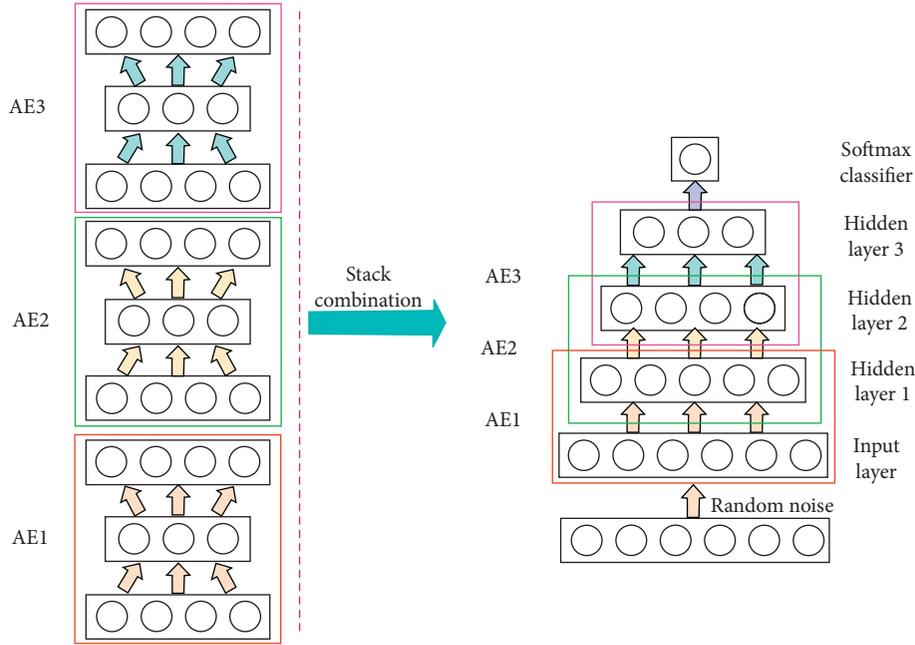


FIGURE 7: Deep neural network with SAE.

TABLE 4: Definition of different conditions of the coal mill.

Label	Condition
1	Coal interruption
2	Coal choking
3	Normal operation

TABLE 5: Training and testing data for fault diagnosis of the coal mill.

Experiment	Sample	Label		
		1	2	3
Experiment 1	Training samples	1500	1500	500
	Testing samples	61	58	75
Experiment 2	Training samples	1500	1500	1500
	Testing samples	61	58	75

as the number of nodes in the input and hidden layers, sparse parameter, and the number of times of network training. The experimental data of Experiment 1 are used as training samples, and relevant experiments are conducted to determine the optimum parameters of the SAE model. The evaluation index is the reconstruction error of the first layer of the SAE model, which is calculated by equation (4), and the experimental results are shown in Figure 8. According to the analysis in Section 3, the significantly changed variables during the fault period of the coal mill include differential pressure of primary air, outlet temperature of coal mill, and current of coal mill. Therefore, the three variables are used as the input nodes of the SAE model.

The SAE model can learn much information when the number of nodes in the input layer is large. However, the number of nodes in the input layer cannot be increased indefinitely because of the computational complexity. Figure 8(a) shows that, when the number of nodes in the

input layer increases from 40 to 100, the reconstruction error of the network decreases continuously. If the number of nodes in the input layer increases further, then the reconstruction error will remain unchanged.

The number of nodes in the hidden layer determines the degree to which the model compresses the input data. The degree of compression is high when the number of nodes in the hidden layer is small. An experiment is employed using the first layer of SAE. In the experiment, the input size is set to 120 on the basis of the experiment above to determine the appropriate hidden layer parameters by analyzing the influence on the reconstruction results. As shown in Figure 8(b), when the number of hidden layer nodes is less than the input layer nodes, the reconstruction error fluctuates in a small range. This result indicates that the original data can obtain better compression when the number of hidden layer nodes is small, and this situation is conducive for the model to learn data characteristics. However, when the number of nodes in the hidden layer exceeds the number of nodes in the input layer, the reconstruction error increases rapidly, and the training effect is poor. The reason is that the sparse parameter ρ is set to 0 at this time, and the activities of neurons in the hidden layer cannot be limited, which then leads to the poor compression effect of the SAE model on original data. Combining the constraints of complexity of network structure and computational efficiency, the number of hidden layers is set to three, and the number of nodes in each layer is 100, 50, and 25.

Sparse constraint is introduced to improve the capability of the SAE model to compress input data. Figure 8(c) shows that, when the value of ρ is between 0.05 and 0.15, the reconstruction error of the network continues to decrease, showing that the inhibitory effect on neurons is appropriate. With the increase in the value of ρ , the inhibitory effect on neurons is excessive, and the reconstruction error increases rapidly.

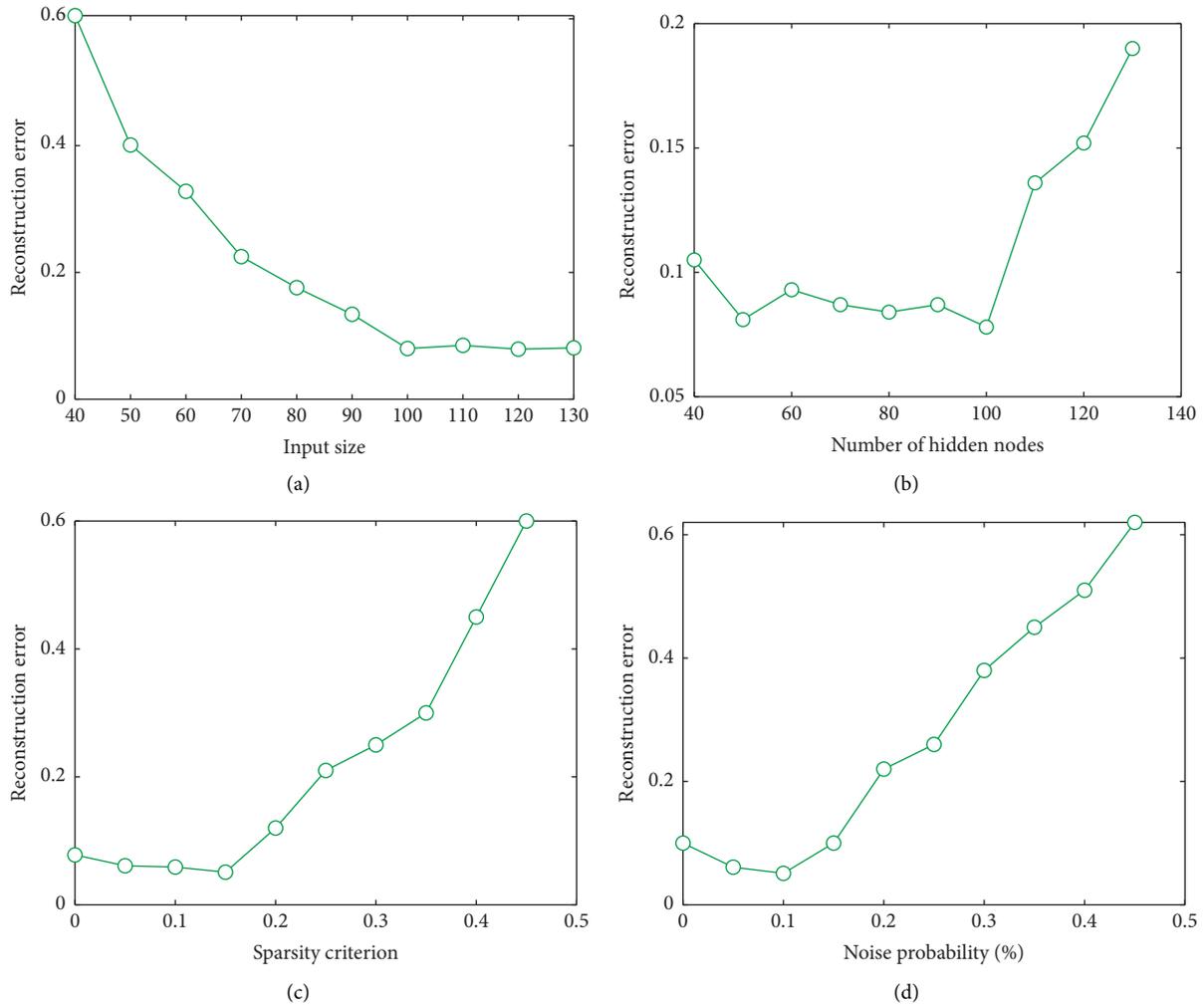


FIGURE 8: Reconstruction error curves for different SAE model parameters. (a) Input size; (b) number of hidden nodes; (c) sparse parameter; (d) noise probability.

TABLE 6: SAE network parameters.

Structure parameters	Input neurons 120	Hidden layer 1 100	Hidden layer 2 50	Hidden layer 3 20	Output layer 4	Transfer function Sigmoid
Learning parameters	Number of training 500	Batch size 100	ρ 0.15	Noise probability 0.1	β 0.05	Learning rate 0.2

Random noise is introduced to the input data to prevent the SAE model from learning only the equivalent representation of the original data. As shown in Figure 8(d), when the probability of introducing noise is in the range of 0 to 0.1, reconstruction error decreases with the increase in noise. However, with the increase in noise, the reconstruction error increases rapidly because excessive noise causes non-destructive damage to the raw data.

In combination with the analysis above, the key parameters of the SAE model are shown in Table 6. The allocation of the nodes in the input layer is shown in Table 7. In the table, 1–40 nodes are sampled values of the differential pressure of primary air in four seconds (the sampling time is set to 0.1 s), 41–80 nodes are sampled values of outlet temperature of coal mill, and 81–120 nodes are sampled values of current of coal mill.

TABLE 7: Assignment situation of input layer nodes.

No.	Variable
1–40	Differential pressure of primary air
41–80	Outlet temperature of coal mill
81–120	Current of coal mill

5.3. *Validation of the Proposed Method.* After determining the parameters of SAE, the network is trained by the training data obtained from simulation experiments, and then the test data are sent into the network to test the result of fault identification of the SAE model, as shown in Figure 9. The simulation results of Experiment 1 show that all samples of false diagnosis in the 194 groups of test samples are from the normal operation data. Two sets of normal operation

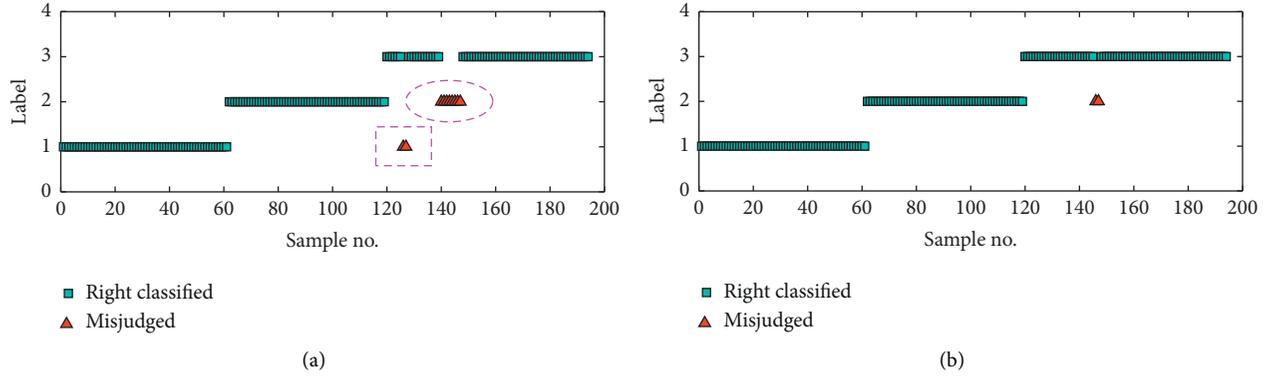


FIGURE 9: SAE-based classification result. (a) Experiment 1; (b) Experiment 2.

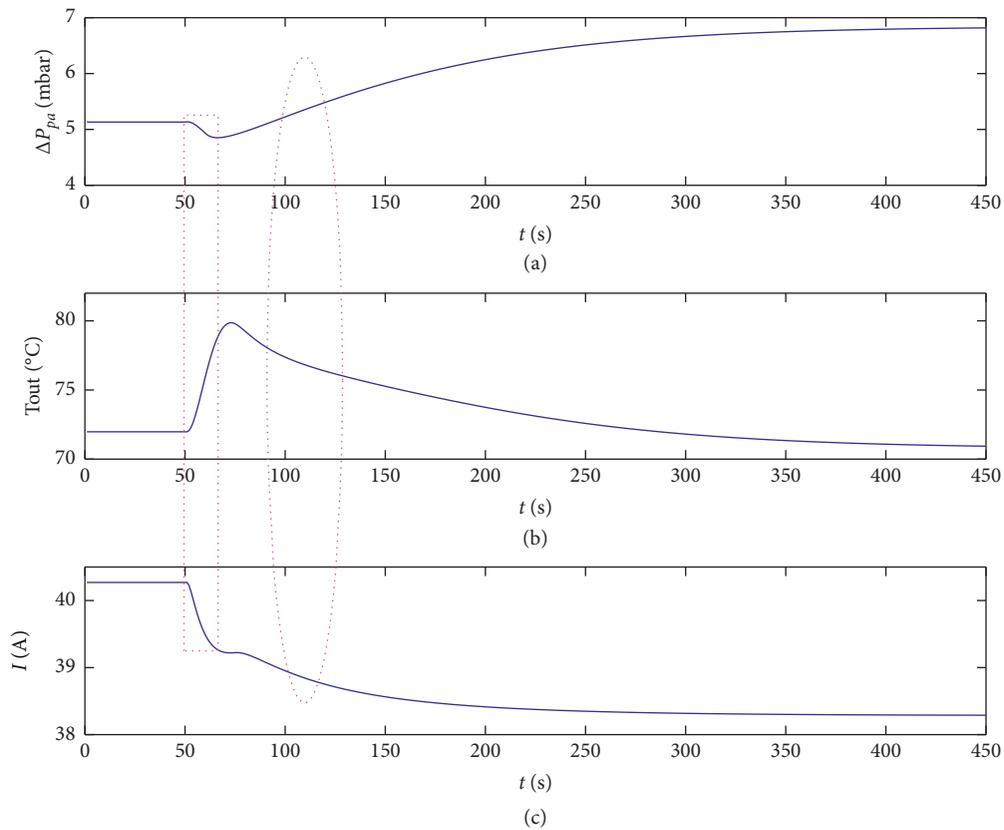


FIGURE 10: Data curve of test samples. (a) Differential pressure of primary air; (b) outlet temperature of coal mill; (c) current of coal mill.

samples are mistakenly diagnosed as coal interruption, and seven sets of normal samples are mistakenly diagnosed as coal choking. The accuracy rate of fault identification is 95.4%. To analyze the experimental results, the data curves corresponding to the misdiagnosed test samples are plotted and shown in Figure 10. The two sets of misdiagnosed samples contained in the rectangular box in Figure 9(a) correspond to data contained in the rectangular box in Figure 10. The change trend of data during this period indicates that differential pressure of primary air decreases, outlet temperature of coal mill rises, and current of coal mill decreases. These trends are consistent with the characteristics of coal mill when the coal interruption fault occurs, and

TABLE 8: Comparison of the effectiveness of network improvement.

Methods	Fault recognition accuracy (%)
SAE without noise and sparse constraint	84.02
SAE with sparse constraint	85.05
SAE with random noise	89.18
Improved SAE	95.4

these characteristics have been described in Section 3.1. Therefore, when normal operation samples are insufficient, SAE cannot fully study the differences between the two types

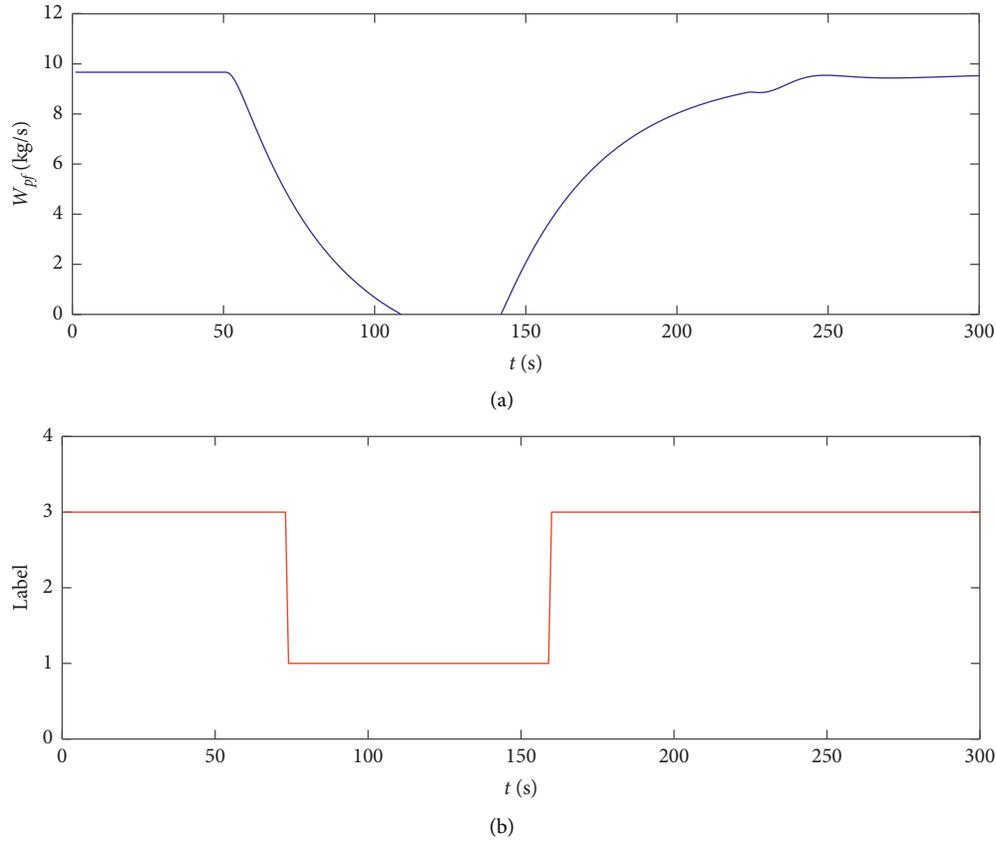


FIGURE 11: SAE-based fault diagnosis of coal interruption. (a) Outlet pulverized coal flow of coal; (b) early warning signal.

of data and wrongly diagnoses the normal operation data as coal interruption fault.

Similarly, from the data contained in the ellipse box in Figure 10, the change trend of the data in this period is found to be consistent with the characteristics of the coal choking, differential pressure of primary air rises, outlet temperature of coal mill decreases, and current of coal mill decreases. Thus, SAE can mistakenly diagnose the normal operation data as coal choking fault. To improve the accuracy of fault diagnosis of SAE, the training samples of normal operation are increased to 1500 groups, and an experiment (Experiment 2) is conducted again. Figure 9(b) shows that, although two sets of misdiagnosed samples are still present, the accuracy of fault diagnosis of SAE has been improved to 98.7%. Therefore, if the training samples continue to increase, then the accuracy of fault diagnosis of SAE will theoretically be close to 100%.

To illustrate the effectiveness of the method that improves the performance of SAE, the accuracy rate of fault diagnosis of SAE before and after the algorithm improvement is compared. The comparison results are shown in Table 8. Training and test samples are from the data in Experiment 1. Table 8 shows that, when no improvement is implemented in SAE, the fault recognition rate is 84.02%. When sparse constraint is introduced in SAE, the fault recognition rate is increased to 85.05%. After introducing random noise into input data, the fault identification rate increases to 89.18%. When two improved methods are introduced into SAE, the network fault recognition rate

further increases to 95.4%. The above analysis shows that the two improved methods proposed in this study can improve the fault recognition capability of SAE.

The coal mill is characterized by a large delay system. Detecting changes in outlet pulverized coal flow of coal mill to find the operation fault in coal mill often cannot establish early warning. Through real-time monitoring of differential pressure of primary air, outlet temperature of coal mill, and current of coal mill, and the three kinds of fast changing signals, the trained SAE can find the operation fault in the coal mill in advance. Coal interruption is taken as an example in this study. As shown in Figure 11, coal interruption fault occurs by artificial simulation. As a result, the outlet pulverized coal flow of coal mill reduces to 0 in 110 s, while the output of the SAE jumps from the normal operation state to coal interruption fault in 75 s. The network has advanced 35 s to predict the fault in the coal mill. With the adjustment of PID, the outlet pulverized coal flow of the coal mill rises gradually and goes back to the safety limit in 160 s. At this point, the output of the SAE returns to normal. Thus, the proposed method based on the deep learning algorithm can play an important role in the fault diagnosis of the coal mill.

6. Conclusions

In this study, a deep learning algorithm based on a data-driven model is proposed for fault diagnosis of coal mills. On the basis of the mechanism model of coal mills, the fault

operation of coal mills is simulated and numerous fault data are obtained. Thus, the difficulty in obtaining the fault data using traditional methods is addressed. The performance of SAE is improved by introducing sparse constraints and random noise in the input layer. At the same time, the accuracy of fault diagnosis of coal mills is effectively improved, thereby enabling the possible prediction of fault in coal mills. The method proposed in the paper greatly improves the accuracy of the fault diagnosis of coal mills, which is of great significance for ensuring the safe operation of power plants. In addition, the proposed method is easy to generalize. Complex mechanical equipment in other industrial fields can use this method for fault diagnosis. The method can reduce the use of sensors for fault diagnosis of large equipment and the investment of human resources, which is essential to improve the economy and safety of the industry.

It should be noted that the paper does not consider online training for SAE. The main reason is that the online training of deep neural networks will take a lot of time, which puts stricter requirements on the performance of computers and optimization algorithms. So, we adopt a simplified method which chooses to directly use the offline training model to achieve fault diagnosis. It greatly saves the cost of calculation. However, using the online training model can continuously optimize the accuracy of the model [27–31] as the data accumulate. Therefore, how to realize the online training of the model will be the focus of our follow-up research.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] V. Agrawal, B. K. Panigrahi, and P. M. V. Subbarao, "Review of control and fault diagnosis methods applied to coal mills," *Journal of Process Control*, vol. 32, pp. 138–153, 2015.
- [2] P. F. Odgaard and B. Mataji, "Observer-based fault detection and moisture estimating in coal mills," *Control Engineering Practice*, vol. 16, no. 8, pp. 909–921, 2008.
- [3] P. Andersen, J. D. Bendtsen, T. S. Pedersen, and B. Mataji, "Coal moisture estimation in power plant mills," in *Proceedings of the 2009 17th Mediterranean Conference on Control and Automation*, pp. 1066–1071, Thessalonik, Greece, 2009.
- [4] J.-L. Wei, J. Wang, and Q. H. Wu, "Development of a multisegment coal mill model using an evolutionary computation technique," *IEEE Transactions on Energy Conversion*, vol. 22, no. 3, pp. 718–727, 2007.
- [5] S. Guo, J. Wang, J. Wei, and P. Zachariades, "A new model-based approach for power plant tube-ball mill condition monitoring and fault detection," *Energy Conversion and Management*, vol. 80, pp. 10–19, 2014.
- [6] Y. Wang, H. R. Karimi, H. Shen, Z. Fang, and M. Liu, "Fuzzy-model-based sliding mode control of nonlinear descriptor systems," *IEEE Transactions on Cybernetics*, vol. 49, no. 9, pp. 3409–3419, 2019.
- [7] Y. Wang, X. Yang, and H. Yan, "Reliable fuzzy tracking control of near-space hypersonic vehicle using aperiodic measurement information," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 12, pp. 9439–9447, 2019.
- [8] E. A. Garcia, Z. Han, and P. M. Frank, "FDI based on parameter and output estimation: an integrated approach," in *Proceedings of the European Control Conference*, Karlsruhe, Germany, 1999.
- [9] T. Hofling and R. Isermann, "Adaptive parity equations and advanced parameter estimation for fault detection and diagnosis," in *Proceedings of the IFAC World Congress*, pp. 55–60, San Francisco, CA, USA, July 1996.
- [10] P. Nomikos and J. F. MacGregor, "Monitoring batch processes using multiway principal component analysis," *AICHE Journal*, vol. 40, no. 8, pp. 1361–1375, 1994.
- [11] Z.-G. Su, P.-H. Wang, X.-J. Yu, and Z.-Z. Lv, "Experimental investigation of vibration signal of an industrial tubular ball mill: monitoring and diagnosing," *Minerals Engineering*, vol. 21, no. 10, pp. 699–710, 2008.
- [12] E. Kisić, V. Petrović, S. Vujnović, Ž. Đurović, and M. Ivezić, "Analysis of the condition of coal grinding mills in thermal power plants based on the T^2 multivariate control chart applied on acoustic measurements," *Facta Universitatis-Series: Automatic Control and Robotics*, vol. 11, no. 2, pp. 141–151, 2012.
- [13] S. Collura, D. Possanzini, D. Pestonesi, and M. Gualerci, *Coal Mill Performances Optimization through Non-invasive Online Coal Fineness Monitoring*, Powergen, Vienna, Austria, 2013.
- [14] X. Han and X. Jiang, "Fault diagnosis of pulverizing system based on fuzzy decision-making fusion method," in *Fuzzy Information and Engineering*, vol. 2, pp. 1045–1056, Springer Berlin Heidelberg, Berlin, Germany, 2009.
- [15] W. Qin, W. Yan, and J. Xu, "Application of fault diagnosis expert system in grinding process," in *Proceedings of the Automation and Logistics (ICAL)*, pp. 290–295, Hong Kong, China, 2010.
- [16] Y. Wang, H. R. Karimi, H.-K. Lam, and H. Shen, "An improved result on exponential stabilization of sampled-data fuzzy systems," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 6, pp. 3875–3883, 2018.
- [17] R. Isermann and P. Balle, "Trends in the application of model based fault detection and diagnosis of technical processes," in *Proceedings of the IFAC World Congress*, pp. 55–60, San Francisco, CA, USA, July 1996.
- [18] P. M. Frank, "Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy," *Automatica*, vol. 26, no. 3, pp. 459–474, 1990.
- [19] X. Guo, L. Chen, and C. Shen, "Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis," *Measurement*, vol. 93, pp. 490–502, 2016.
- [20] Y. Duan, Y. Lv, Y. L. Liu et al., "An efficient realization of deep learning for traffic data imputation," *Transportation Research Part C: Emerging Technologies*, vol. 72, pp. 168–181, 2016.
- [21] Y. Gao, D. Zeng, J. Liu, and Y. Jian, "Optimization control of a pulverizing system on the basis of the estimation of the outlet coal powder flow of a coal mill," *Control Engineering Practice*, vol. 63, pp. 69–80, 2017.
- [22] D.-L. Zeng, Y. Hu, S. Gao, and J.-Z. Liu, "Modelling and control of pulverizing system considering coal moisture," *Energy*, vol. 80, pp. 55–63, 2015.
- [23] D. Simon, "Kalman filtering with state constraints: a survey of linear and nonlinear algorithms," *IET Control Theory & Applications*, vol. 4, no. 8, pp. 1303–1318, 2010.

- [24] V. Rashtchi, E. Rahimpour, and S. Fazli, "Genetic algorithm application to detect broken rotor bar in three phase squirrel cage induction motors," *International Review of Electrical Engineering*, vol. 6, no. 5, pp. 2286–2292, 2011.
- [25] H. Pan and X. W. Wang, "Survey on collaborative filtering recommendation algorithm based on extreme learning machine stacked denoising autoencodes," *Application Research of Computers*, vol. 33, no. 8, 2016.
- [26] Y. Jian, X. Qing, H. Liang, Y. Zhao, X. Qi, and M. Du, "Fault diagnosis of motor bearing based on deep learning," *Advances in Mechanical Engineering*, vol. 11, no. 9, pp. 1–9, 2019.
- [27] A. Das and F. L. Lewis, "Distributed adaptive control for synchronization of unknown nonlinear networked systems," *Automatica*, vol. 46, no. 12, pp. 2014–2021, 2010.
- [28] S. El-Ferik, H. A. Hashim, and F. L. Lewis, "Neuroadaptive distributed control with prescribed performance for the synchronization of unknown nonlinear networked systems," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 12, pp. 2135–2144, 2017.
- [29] S. El-Ferik, A. Qureshi, and F. L. Lewis, "Neuro-adaptive cooperative tracking control of unknown higher-order affine nonlinear systems," *Automatica*, vol. 50, no. 3, pp. 798–808, 2014.
- [30] H. A. Hashim, S. El-Ferik, and F. L. Lewis, "Adaptive synchronisation of unknown nonlinear networked systems with prescribed performance," *International Journal of Systems Science*, vol. 48, no. 4, pp. 885–898, 2017.
- [31] F. L. Lewis, H. Zhang, K. Hengster-Movric, and A. Das, *Cooperative Control of Multi-Agent Systems: Optimal and Adaptive Design Approaches*, Springer Science & Business Media, Berlin, Germany, 2013.