SPECIAL SECTION ON ADVANCES IN MACHINE LEARNING
AND COGNITIVE COMPUTING FOR INDUSTRY APPLICATIONS

IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received March 17, 2020, accepted March 30, 2020, date of publication April 2, 2020, date of current version April 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2985233

Control Strategy for Denitrification Efficiency of Coal-Fired Power Plant Based on Deep Reinforcement Learning

JIGAO FU [1], HONG XIAO [1], HAO WANG [2],(Member, IEEE), AND JUNHAO ZHOU [3]

[1] Faculty of Computer, Guangdong University of Technology, Guangzhou 510006, China
[2] Department of Computer Science, Norwegian University of Science and Technology, 2802 Gjøvik, Norway
[3] Faculty of Information Technology, Macau University of Science and Technology, Macau, China

Corresponding author: Hong Xiao (wh_red@163.com)

This work was supported in part by the Research on self-organizing elasticity enhancement strategy of intelligent manufacturing IoT network, National Natural Science Foundation of China under Grant 61672170, and in part by the R&D and Application of a New Generation of Intelligent Industrial Robots, Core Technology Research Project, Foshan City, under Grant 1920001001367.

ABSTRACT The optimal control of denitrification system in coal-fired power plants in China has recently received widespread attention. The accurate prediction of denitrification efficiency and formulate control strategy of denitrification efficiency can guide the control and operation of the denitrification system better. Meanwhile, it can achieve the effect of energy conservation and Nitrogen oxides (NO_x) reduction. In this paper, we take a domestic 1000 MW unit as an example, consider each of the major factors that affect the denitrification efficiency of selective catalytic reduction (SCR). We put forward a deep reinforcement learning (DRL) model by combining the Long short-term memory (LSTM) model and the Asynchronous Advantage Actor - Critic algorithm (A3C). We first use the LSTM to build a prediction model for denitrification efficiency. We then use the DRL model to obtain a control strategy for SCR denitrification efficiency in coal-fired power plants. The experimental results demonstrate that the accuracy of denitrification efficiency prediction model we established is better than other machine learning models, reaching 91.7%. Our control strategy model is industrially feasible and universally applicable.

INDEX TERMS Coal-fired power plant, denitrification efficiency, selective catalytic reduction (SCR), long short-term memory (LSTM), asynchronous advantage actor critic (A3C), deep reinforcement learning.

I. INTRODUCTION

As the industry continues to develop, Nitrogen oxides (NO_x) emissions are also increasing continuously. Acid rain has changed from sulfuric acid type to composite type of sulfuric acid and nitric acid [1]. NO_x has gradually become the main source of gaseous pollution. For power station boilers burning pulverized coal, the pollutants' NO_x emissions are mainly Nitric oxide (NO) and Nitrogen dioxide (NO_2), of which NO accounts for more than 90%, so NO and NO_2 are generally referred to as NO_x.

In recent years, the government and research institutes have done a lot of research on the control of NO_x pollution, they have developed many practical and efficient new technologies. According to the different control stages of nitrogen oxides during combustion, the emission reduction technologies are generally divided into before combustion, during combustion and flue gas denitrification after combustion. Among them, the most effective emission reduction technology is flue gas denitrification after combustion. Denitrification measures after combustion include hot carbon reduction, wet complex absorption, selective non-catalytic reduction (SNCR), electron beam irradiation, selective catalytic reduction (SCR), plasma, microbiological methods, etc [2]. Among them, SCR is currently the most widely used flue gas denitrification technology in the world. The SCR means that a reducing agent (generally ammonia) selectively reduces NO_x (mainly NO) in the flue gas to N_2 and H_2O under conditions of catalyst, oxygen and a certain temperature range. It has the advantages of high denitrification efficiency, mature technology, no secondary pollution, reliable operation and easy maintenance [3], [4], it is most suitable for vigorous promotion. That is why the SCR is most widely used in Chinese coal-fired power stations.

Therefore, the accurate prediction of denitrification efficiency and formulate control strategy of denitrification efficiency can guide the control and operation of the denitrification system better. How to steadily optimize the

The associate editor coordinating the review of this manuscript and approving it for publication was Hong-Ning Dai.

denitrification efficiency in the SCR denitrification system, so that coal-fired power plants can achieve the energy conservation and emission reduction has become a hot research topic.

A. MOTIVATION

The basis for optimizing the denitrification control system is to accurately establish the relationship between actual denitrification efficiency and the parameters of industrial process. According to the actual operation, the SCR flue gas temperature is related to the boiler load and fuel combustion. The catalyst can usually only be replaced after the failure, so the Spray ammonia mass flow becomes one of the key factors for daily adjustment [5],[6].

The chemical reaction equation for denitrification is:

$$4NO + 4NH_3 + O_2 \rightarrow 4N_2 + 6H_2O, \qquad (1)$$

$$2NO_2 + 4NH_3 + O_2 \rightarrow 3N_2 + 6H_2O. \qquad (2)$$

Consequently, the denitrification efficiency is affected by multiple factors such as Boiler load, Inlet NO_x mass concentration, Inlet O_2 mass concentration, Inlet flue gas temperature, Spray ammonia mass flow and Ammonia/air mixer ammonia inlet regulator valve position feedback, etc.

However, there are some challenges in predictive analysis and control of denitrification efficiency.

1) Currently, the instruments for flue gas monitoring and analysis can monitor the mass concentration of NO_x in the inlet and outlet flue gas of SCR denitrification equipment directly. Meanwhile, they compute the denitrification efficiency. But this way is just a simple feedback on the results of denitrification reaction process, it cannot reflect the relationship between the monitoring process parameters and the denitrification efficiency. Simultaneously, the monitoring equipment is relatively influenced by external factors, and sometimes there are problems such as faults, etc. which will lead to inaccurate measurement results.
2) The current machine learning methods mostly used for predictive modeling of denitrification efficiency in coal-fired power plants are regression analysis, Support Vector Machine (SVM), Artificial Neural Network (ANN), etc. However, there are fewer reasons for selecting auxiliary variables, less data samples and no consideration of the time series characteristics of process parameters, so that the prediction results still have large errors and lack a certain generalization ability.
3) There is no reliable strategy model for controlling denitrification efficiency in the progress of industrial applications.

Existing methods based on formula or data-driven have modeled the denitrification efficiency and achieved not bad predicted results, but the accuracy of prediction needs to be further improved. However, in order to truly optimize the denitrification efficiency in the SCR denitrification system, it is not only necessary to improve the prediction accuracy of the denitrification efficiency, but also to research how to control the effects of the above factors on the denitrification efficiency and the cost in the denitrification system. Establish an effective control strategy model for the related factors of denitrification efficiency.

B. CONTRIBUTIONS

In order to solve the above challenges, we put forward an intelligent control method for denitrification efficiency. This method combines Long short-term memory neural network (LSTM) with Asynchronous Advantage Actor - Critic (A3C) algorithm to build a deep reinforcement learning (DRL) model, and implements a control strategy for SCR denitrification efficiency of coal-fired power plants. As far as we know, we are the first to adopt deep reinforcement learning method in the control strategy of denitrification efficiency. In this paper, our main contributions are summarized as follows.

1) We originally propose a deep reinforcement learning model that combines LSTM and A3C algorithm, which can provide control strategies for the control of denitrification efficiency or related industrial applications.
2) Taking the 1000 MW boiler of a power plant in Guangdong as the research object, we use the LSTM to establish the SCR denitrification efficiency prediction model. The model is able to embody the effect of the time series characteristics of industrial process parameters on actual denitrification efficiency. Experimental results demonstrate that compared with existing models, the accuracy of prediction is improved.
3) We take the SCR system's denitrification efficiency to be not less than 85% as the goal. Then we use DRL algorithm to perform simulation control experiments on the LSTM model. The experimental results indicate that our proposed approach can maintain the denitrification efficiency above 85%. Meanwhile, the value control strategy of the optimal input variable action combination is obtained.

In this paper, the remainder is arranged as follows. We review related works on predictive analysis of denitrification efficiency and control strategies in Section II. Section III describes the main methods that we used. Section IV presents the details of the model which we put forward. Then Section V discusses the results of experiment. At last, Section VI summarizes our work and discuss the direction of research in future.

II. RELATED WORK

This section looks back the latest advances in denitrification efficiency for coal-fired power plants. We divide the recent research into two categories: 1) Predictive analysis of denitrification efficiency based on machine learning approaches; 2) Control strategies based on deep reinforcement learning methods.

A. PREDICTIVE ANALYSIS OF DENITRIFICATION EFFICIENCY BASED ON MACHINE LEARNING APPROACHES

With the development of artificial intelligence and computer technology in recent years, the ability to deal with nonlinear problems is gradually increasing. It has achieved notable results in the area of modeling and optimization of power plant boilers. As data-driven modeling methods are widely used in industry, some scholars have proposed using soft-measurement technology to predict and analyze denitrification efficiency [7]. They used multiple regression algorithms, neural networks, genetic algorithms or partial least squares to model and predict the denitrification efficiency, and achieved good prediction results. However, due to fewer auxiliary variables and less data samples, it makes the prediction results in the literature still have a lot of errors and lack of the ability to popularize. Currently, most of the methods for predictive modeling of denitrification efficiency in coal-fired power plants are regression analysis, SVM, ANN or other methods. Zhao *et al.* [8] adopted Principal Component Analysis (PCA) to select the dominant factors affecting denitrification efficiency, they established a predictive model on the basis of Least Squares Support Vector Machine (LS-SVM) to obtain good generalization ability and prediction accuracy. However, none of the foregoing methods take into consideration the time series characteristics of industrial process parameters.

Deep Learning (DL) has become an important research hotspot in the area of machine learning [9], which has achieved remarkable success in the field of image analysis [10], machine translation [11], video classification [12], speech recognition [13], etc. The basic idea of DL is combining low-level features through non-linear transformation and multi-layer network structure to form easily distinguishable, abstract high-level representations to discover distributed feature representations of data [14]. Therefore, the DL places extra emphasis on the expression and perception of things. But training a DL model is a very time consuming task because DL models usually involve numerous parameters [15], it means a large amount of data, high speed streaming data and different types of data, which poses a challenge for DL models [16].

The LSTM is a kind of deep learning methods which is an improved time recurrent neural network on the basis of the RNN. It was originally put forward by Hochreiter and Schmidhuber [17] to solve the gradient explosion, gradient disappearance, lack of long-term memory ability, etc. during the use process of RNN, enables RNN to be effectively used for the time series information of long distance [18]. Recently, with the continuous development of DL, the models of LSTM have been successfully applied in a number of different areas such as sentimental analysis [19], traffic speed prediction [20], electrical load forecasting [21], failure time series prediction [22], etc. Compared with our work in [23], we expand on it and combine the method of reinforcement learning (RL) in this paper, which can better formulate industrially applicable control strategies. Because the optimization of industrial processes not only needs to improve the prediction accuracy of certain factors, but also needs to know how to control their controllable variables.

Another research hotspot in the area of machine learning is RL, which has been widely used in industrial manufacturing [24], robot control [25], optimization and scheduling [26], game theory of games [27] and other fields. The main idea of RL is learning the optimal strategy to achieve the goal through maximizing the cumulative reward value got by the agent from the environment [28]. So the RL places extra emphasis on learning problem-solving strategies.

B. CONTROL STRATEGIES BASED ON DEEP REINFORCEMENT LEARNING METHODS

With the high-speed development of human society, in more and more complex real-world tasks. It is necessary to utilize DL for learning automatically the abstract representation of extensive input data, then use RL can self-encourage based on this representation to optimize problem-solving strategies. Therefrom, DeepMind, the artificial intelligence research team of Google, which combines innovatively the perception ability of DL and the decision ability of RL, forming a new research hotspot in the area of artificial intelligence, namely Deep Reinforcement Learning (DRL). It can achieve direct control from initial input to output by the method of end-to-end learning. Since its introduction, DRL methods have made substantive breakthroughs in many tasks that require perception of high-dimensional original input data and control of strategy. The emergence of DRL makes the RL technology truly practical, which can solve complex problems in real scene. Since then, DeepMind team has created many agents of human expert-level in many challenging areas. These agents build and learn their own knowledge from the initial input signal directly, without any domain knowledge and the coding of manual control.

The Policy Gradient is a commonly used method for optimizing strategies, it updates the strategy parameters by continuously computing the gradient of strategy's expected total reward with respect to strategy parameters, and finally converges on the optimal strategy [29]. Therefore, when solving the DRL problem, the deep neural network with parameter θ can be adopted to parameterize the representation strategy, and Policy Gradient is used to optimize the strategy. It is worth noting that when solving DRL problems, the first choice is to adopt an algorithm based on Policy Gradient. The reason is that it can optimize strategy's expected total rewards directly. Meanwhile, search the optimal strategy directly in the policy space by the manner of end-to-end, eliminating the cumbersome intermediate links. Therefore, compared with Deep Q Network (DQN) [30] and its improved model, the DRL method based on Policy Gradient is more applicable and the effect of strategy optimization is better.

The basic idea of the Deep Policy Gradient is to optimize directly the strategy of parameterization representation with deep neural networks through various Policy Gradient

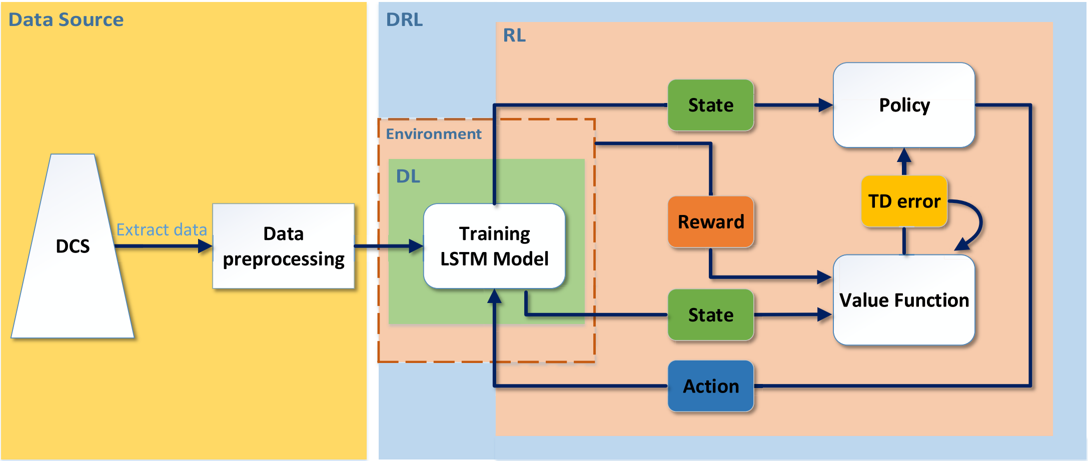


FIGURE 1. Overview of methodology.

approaches. At each iteration of this type of method, it is need to sample the trajectory $\{\tau_i\}\mathrm{N}_i = 1$ with a batch size is N to renew the Policy Gradient. However, it is difficult to get numerous training data online in many complex real-world scenarios. For example, in the manipulation task of a robot in a real scene, collecting and utilizing numerous training data online is very expensive. Meanwhile, the continuous feature of the action makes the method of extracting the batch track online unable to achieve satisfactory coverage. The above problems lead to the emergence of local optimal solutions. To solve this problem, the Actor-Critic (AC) framework in the traditional RL can be extended to the Deep Policy Gradient method. Figure 3 demonstrates the learning framework of a Deep Policy Gradient method based on the AC framework.

Different types of deep neural networks provide a highly efficient representation of the strategy optimization tasks in DRL. So as to alleviate the instability caused by the combination of neural networks and traditional Policy Gradient approaches, various types of Deep Policy Gradient approaches adopt the experience replay [31] to eliminate the correlation between training data, such as Stochastic Value Gradient (SVG) [32], Deep Deterministic Policy Gradient (DDPG) [33], etc. However, the experience replay has two shortcomings: (1) Each real-time interaction between the environment and the agent will consume a large number of computing power and memory; (2) It claims the agent to adopt the off-policy for learning, but the off-policy can only renew the data produced by the old policy. In response to these problems, Mnih *et al.* [34] put forth a lightweight framework of DRL which based on the idea of Asynchronous Reinforcement Learning (ARL). The framework can utilize asynchronous gradient descent method to optimize network controllers' parameters, and it can be combined with a variety of RL algorithms. Among them, the A3C algorithm performs best in the control tasks of various continuous motion spaces.

Taking into account the data of the boiler combustion process and denitrification process also have time series characteristics. Meanwhile, our goal is to achieve a control strategy for SCR denitrification efficiency in coal-fired power plants. There is currently no reliable denitrification efficiency control strategy in the progress of industrial applications. Therefore, in this paper, we put forward a method of combining LSTM with A3C algorithm to build a DRL model. This method can be used to implement control strategies for SCR denitrification efficiency in coal-fired power plants. It has universal applicability and can also be used in other industrial applications.

III. OVERVIEW OF METHODOLOGY

Figure 1 demonstrates an approach for implementing a control strategy for SCR denitrification efficiency. Firstly, the data can be obtained from the Distributed Control System (DCS) of the power plant. In this paper, parameters related to the comparison of SCR denitrification efficiency are selected as data sources. These parameter data are then preprocessed to obtain the parameter data most relevant to the SCR denitrification efficiency. Once the data is ready, all the preprocessed data is input into the LSTM model, and then the trained LSTM model is used as the DRL's environment. After that, the DRL method based on the A3C algorithm uses the State value output by the environment as an input to output a new Action value. Finally, through several iterations, the control strategy of SCR denitrification efficiency was obtained.

A. LONG SHORT-TERM MEMORY NEURAL NETWORK

The LSTM realizes the memory function of time by the switch of the gate to prevent the disappearance of gradient. Each LSTM cell has three gate controllers, which are forget gate, input gate and output gate, forming a new computing unit. The forget gate is responsible for controlling the retention of historical state information of the computing unit. The input gate is responsible for controlling the input of information. The output gate is responsible for controlling the output of information. The activation function Sigmoid makes the output value of the forget gate to be [0,1]. When the output is 1, indicating all the information of previous state is

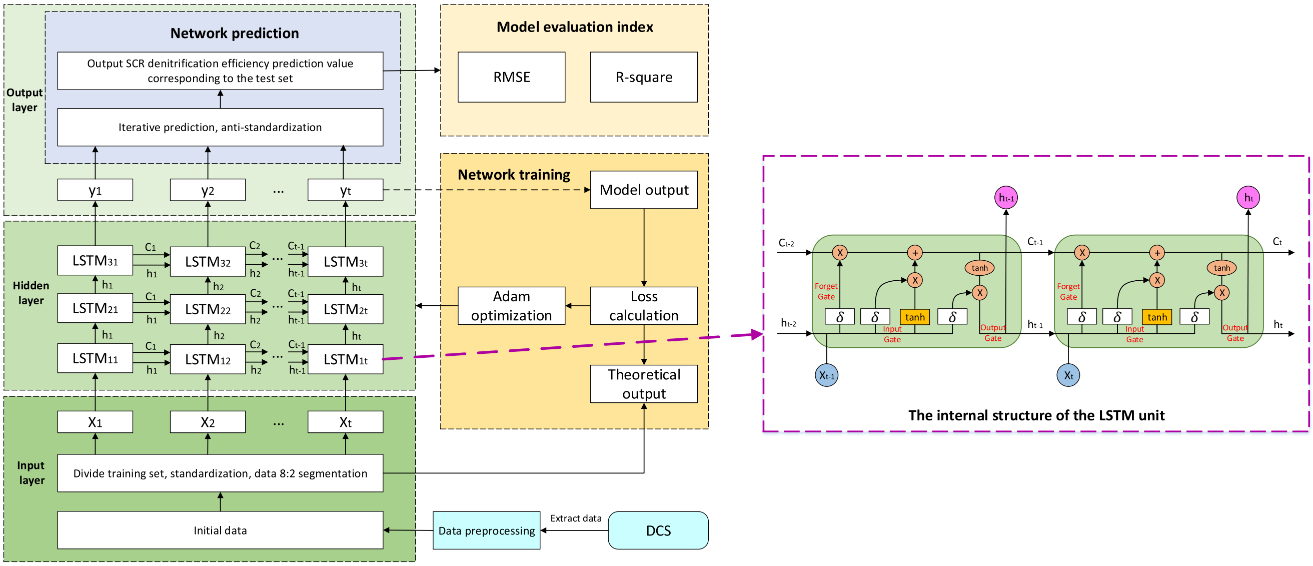


FIGURE 2. Framework of denitrification efficiency prediction model based on LSTM.

retained. When the output is 0, indicating all the information of previous state is discarded by the forget gate. The process of calculation can be represented as:

$$f_t = \delta(W_f x_t + U_f h_{t-1} + V_f c_{t-1} + b_f), \tag{3}$$

$$i_t = \delta(W_i x_t + U_i h_{t-1} + V_i c_{t-1} + b_i), \tag{4}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot tanh(W_c x_t + U_c h_{t-1} + b_c), \tag{5}$$

$$o_t = \delta(W_o x_t + U_o h_{t-1} + V_o C_t + b_o), \tag{6}$$

$$h_t = o_t \cdot tanh(c_t), \tag{7}$$

where f_t, i_t and o_t are the computing rule of the forget gate, the input gate and the output gate at time t respectively, c_t is the computing rule of the LSTM cells at time t, h_t is the output of the computing unit at time t, $\delta(\cdot)$ is the activation function Sigmoid, $tanh(\cdot)$ is the hyperbolic tangent activation function, W, U and V are parameter matrices, b is the bias term.

By the formula (3) - (7) can be seen that input gate, forget gate and output gate each connected to a multiplier to control the state of each cell and the input and output of information. Figure 2 shows the internal structure of the LSTM unit.

The LSTM adopts a Back Propagation Through Time (BPTT) algorithm during training. The optimized gradient algorithm uses the Adaptive Moment Estimation (Adam) algorithm. Adam combines the advantages of the Root Mean Square Prop (RMSProp) algorithm and the Momentum algorithm. It can compute the adaptability of different parameters and occupy less resources of processor. Compared with other algorithms of optimization, Adam demonstrates great advantages in practical applications.

B. DEEP REINFORCEMENT LEARNING BASED ON ASYNCHRONOUS ADVANTAGE ACTOR - CRITIC

DRL is a system of end-to-end perception and control with strong versatility. The learning process can be represented as follows: (1) The agent interacts with the environment to get a high-dimensional observation at each moment, and the DL approaches can perceive the observation to get a specific state feature representations; (2) Evaluate the value function of each action based on the expected reciprocation, and map the current state to the corresponding action through a certain policy; (3) The environment reacts to this action and obtains the next observation. Through continuously circulating the above process, the optimal strategy for achieving the goal can be eventually obtained. The principle framework of DRL is shown in Figure 3.

Specifically, the A3C algorithm performs multiple agents in parallel and asynchronously using the functions of the CPU multi-thread. Therefore, at any time, the parallel agents will go through many different states, eliminating the correlation between the samples of state transitions produced during the training process. Therefore, this low-consumption asynchronous execution method can be a good alternative to the experience playback mechanism.

The A3C algorithm reduces the hardware requirements during training. The depth strategy gradient algorithm relies heavily on a computationally intensive GPU, while the A3C algorithm requires only one standard multicore CPU in the actual operation. By applying multi-threading technology, the A3C algorithm reduces the hardware requirements of the model. In the case of less training time, the average performance of the A3C algorithm on the Atari 2600 game task is significantly improved. Moreover, the A3C algorithm is able to learn an effective strategy for walking a 3D maze based only on the original visual input. In addition, the A3C algorithm also can be widely applied to various continuous action space problems. In summary, the A3C algorithm can be widely applied to various 2D, 3D discrete and continuous motion space tasks. Meanwhile, it achieved the best results in these tasks. This shows that A3C is currently the most versatile and successful DRL algorithm.

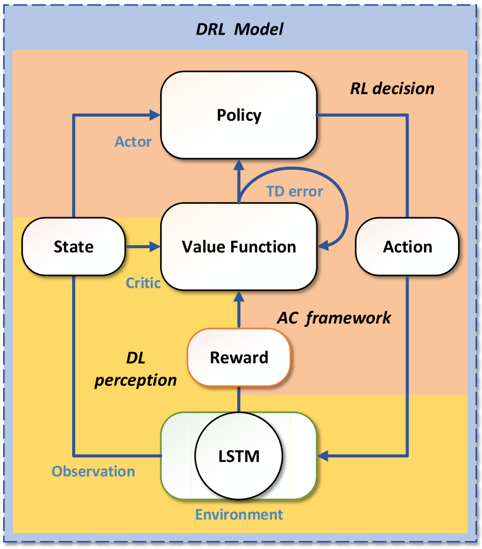


FIGURE 3. The framework of denitrification efficiency model based on DRL.

IV. DENITRIFICATION EFFICIENCY CONTROL STRATEGY MODEL BASED ON DEEP REINFORCEMENT LEARNING

A. THE INFLUENCING FACTORS OF DENITRIFICATION EFFICIENCY

An important index to test the performance of denitrification system is the denitrification efficiency. Denitrification efficiency is the percentage of NO_x removed by the SCR denitrification system to the NO_x entering the SCR denitrification system in the unit time. The SCR denitrification crafts relates a series of chemical and physical reactions. Therefore, the denitrification efficiency is determined by many factors [35].

The power plant DCS records the Boiler load, Inlet O_2 mass concentration, Inlet NO_x mass concentration, SCR ammonia slip, Spray ammonia mass flow, denitrification catalyst inlet and outlet differential pressure, etc. 8 parameters every 1 minute to represent the denitrification status of the SCR denitrification system. We carry out PCA and correlation analysis of each effective parameter on site and the denitrification efficiency in this paper. The analysis results demonstrate that the cumulative contribution rate of the 6 main components of the Boiler load, Inlet flue gas temperature, Inlet O_2 mass concentration, Inlet NO_x mass concentration, Ammonia/air mixer ammonia inlet regulator valve and Spray ammonia mass flow to the original data is greater than 90%, so we select the 6 main components as input to the LSTM model.

Real-time monitoring of the entire SCR denitrification process of a 1000MW unit in China using the monitoring system of the existing fume pollution source emission process (operating condition). The monitoring data per minute for 6 input variables required by the model during the period from 2018-09-01 to 2018-09-07 was selected, a total of 10000 sets of the data were used as the samples of model. Among them, 8000 sets of the data are training data (including verification samples) and 2000 sets of the data are test data. The value range of each parameter in the data is shown in Table 1.

TABLE 1. The value range of parameter in SCR denitrification system of a unit.

Parameter	Ranges
Boiler load /(MW)	394.35∼998.80
Inlet flue gas temperature /°C	302.21∼357.20
Inlet O_2 mass concentration /(mg· m^{-3})	0.99∼6.94
Inlet NO_x mass concentration /(mg· m^{-3})	130.44∼491.22
Ammonia/air mixer ammonia inlet regulator valve	18.66∼90.06
Spray ammonia mass flow /(kg / h)	29.38∼192.49
Measured denitrification efficiency /%	53.26∼98.54

B. SCR DENITRIFICATION EFFICIENCY MODEL BASED ON LSTM

The design scheme of the prediction model for denitrification efficiency based on LSTM is as follows. We adopt 6 main variables as input variables of the LSTM. In the design of the framework of the network, after repeated experiments. It is eventually decided that the LSTM neural network contains 3 LSTM layers, 64 nodes per layer. The optimized method of the model uses the Adam algorithm, the training data is set to 8000 sets, test data is set to 2000 sets, timestep is 10, batch size is 20 and initial learning rate is 0.001.

Figure 2 shows the overall framework of denitrification efficiency prediction model based on LSTM we build, which includes 5 functional modules: Input layer, Hidden layer, Output layer, Network training and Network prediction. Input layer performs preliminary processing on the initial time series to satisfy the input requirements of the network. Hidden layer adopts the LSTM cells, which are demonstrated in Figure 2 to set up a 3-layer recurrent neural network. Output layer is responsible for providing predicted results. Network training adopts the Adam algorithm. Network prediction adopts an iterative approach to predict point by point.

C. ESTABLISHMENT OF DEEP REINFORCEMENT LEARNING MODEL

The design scheme of the SCR denitrification efficiency DRL model based on the AC framework is as follows. Select Boiler load, Spray ammonia mass flow, Inlet O_2 mass concentration, Inlet flue gas temperature, Inlet NO_x mass concentration, Ammonia/air mixer ammonia inlet regulator valve position feedback 6 variables as action value, denitrification efficiency as state value, value range setting as Table 1 is shown.

In the design of the structure of the network, after repeated experimental debugging. It is finally determined that the learning round of the whole model is set to 2000, and 100 times in each round, of which 80 trainings, 20 predictions. The reward function is set to a denitrification efficiency of 85% or more, the reward = 10. When the denitrification efficiency is greater than or equal to 90%, the reward = 20. When the denitrification efficiency is less than 85%, the reward = -30. The Actor learning rate is 0.001

and the Critic learning rate is 0.01. The training process of this model is as follows:

1) Firstly, we set the ranges of Action value and the rules for Reward value;
2) Then, the Actor network (Policy) adjusts the change of the Action value, and output State value (denitrification efficiency value) through Environment (LSTM);
3) The State value gets the Reward value of the corresponding Action through the Critic network (Value Function), and then feeds back to the Actor network;
4) The Actor network adjusts based on the State value and the Reward value fed back by the Critic network to obtain new Action values;
5) Finally, through continuous iterative updates, the optimal Action combination that meets the conditions is obtained.

The overall framework of the SCR denitrification efficiency DRL model based on AC framework is shown in Figure 3.

V. EXPERIMENTAL RESULTS

A. PERFORMANCE METRICS OF THE MODEL

In order to compare the proposed approach with other models, we adopt two performance metrics: root mean square error (RMSE) and R-square (R^2).

The RMSE is the standard deviation of the residuals between observed values and predicted values, it can well reflect the prediction accuracy. The RMSE is calculated by:

$$\delta_{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} \sigma_{error_i}^2}{n}}, \tag{8}$$

where n is the total number of test data, σ_{error} is the predicted error.

The R^2 can represent the quality of a fit through changes in the data.

$$R^2 = 1 - \frac{\sum (Y_{actual} - Y_{predict})^2}{\sum (Y_{actual} - Y_{mean})^2}. \tag{9}$$

The denominator is understood as the discreteness of the initial data. The numerator is the error between the initial data and the predicted data. Dividing the two can remove the influence of the discreteness of the original data. The theoretical value range is $(-\infty, 1)$, the theoretical value range is $[0, 1]$. In actual operation, a curve that fits good is generally chosen to calculate R^2, so $-\infty$ is rarely seen.

If R^2 is closer to 1, it indicates that the equation's variables have a stronger ability to explain y, and the model will fit the data better. If R^2 is closer to 0, it indicates that the model fits worse. Currently, a large amount of experimental results indicate that if R^2 is more than 0.4, fitting effect of the model is good.

B. PERFORMANCE COMPARISON

In this paper, we adopt 10000 sets of data of a 1000 MW coal-fired boiler unit in Guangdong under the operating conditions from 00:01 on September 1, 2018 to 22:40 on September 7, as experimental data. The coal-fired boiler has no obvious external operations that affect combustion during this period. Investigate abnormal operating data and abnormal value of the experimental data, no abnormal operating data and abnormal value were detected. It shows that the unit is basically in stable operating state.

Table 2 presents the process of adjusting the time step of LSTM. The results reflect that when the time step of the LSTM model parameters is 10 and the number of training epochs is 500, the model has the best predictive performance. So as to verify the predicted performance of LSTM model in SCR denitrification efficiency of coal-fired power plants. By comparing the LSTM model with the LSSVM model and the RNN model, we obtain the predicted results of 2000 sets of test data. Figure 4-6 shows us that the 100 sets of experimental results of SCR denitrification efficiency prediction, Table 3 lists the performance comparison of our LSTM model with other models. The key parameters of the LSSVM model are the parameters adjusted to the optimal accuracy. The key parameters of the RNN model and the LSTM model are the same for easy comparison.

TABLE 2. Performance comparison of time step.

Time Step (training epochs = 500)	RMSE	R^2
1	1.876585	0.883327
5	1.984778	0.873541
10	**1.606915**	**0.916849**
15	1.782510	0.887470
20	2.165182	0.847478

TABLE 3. Performance comparison with other models.

Models	Key parameters	RMSE	R^2
LSSVM	RBF kernel; gam = 10; sig2 = 0.1	2.075210	0.850845
RNN	Time Step = 10; epochs = 500	1.969089	0.867800
LSTM	Time Step = 10; epochs = 500	**1.606915**	**0.916849**

It can be seen from Figure 4-6 and Table 3 that the prediction results of the RNN model and the LSSVM model are not much different, the performance of the RNN model is better than the LSSVM model. The RMSE of the LSTM model is 1.606915 and the R^2 of the LSTM model is 0.916849. The performance of the LSTM model are better than the LSSVM

TABLE 4. Distribution of rewards and punishments.

Reward and punishment result	Rounds (2000)	State value /%
+20	28	[90 , +∞)
+10	1905	[85 , 90)
-30	67	(-∞ , 85)

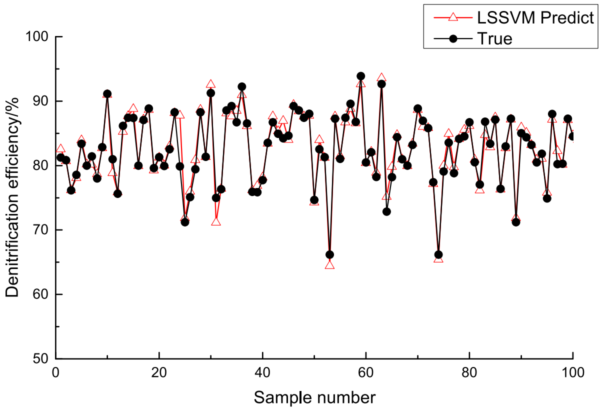


FIGURE 4. Prediction of LSSVM model.

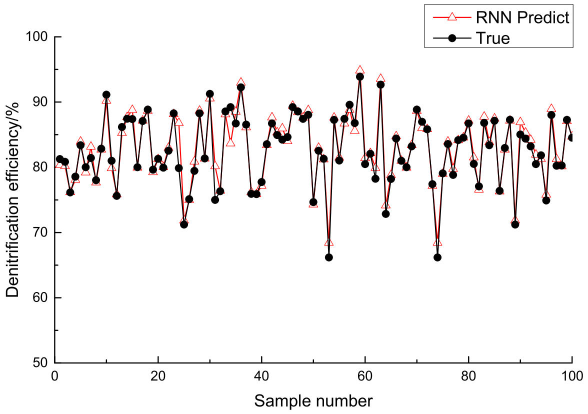


FIGURE 5. Prediction of RNN model.

model and RNN model, which indicates that our LSTM model has higher prediction accuracy than other existing models.

Therefore, the LSTM model is used as the environment in the DRL model. After repeated experimental debugging, it is eventually decided that the reinforcement learning round setting has obvious convergence effect at 2000. There are 2000 learning rounds, 80 trainings and 20 tests per round. Take one prediction result per round as shown in Figure 7-8.

As can be seen from Figure 7, the total reward value of the model tends to be stable after 500 rounds. It can be seen from Figure 7-8 and Table 4, in 2000 rounds, there were 67 rounds of punishment, 28 rounds were rewarded with 20, and the remaining 1905 rounds were rewarded with 10. It can be seen from Table 5 that the most stable predicted state value remains at 85.66. Therefore, the most stable control strategy predicted is that when the Boiler load is 751.96 MW, the Inlet flue gas temperature is adjusted to 335.739 ^{o}C, the Inlet O_2 mass

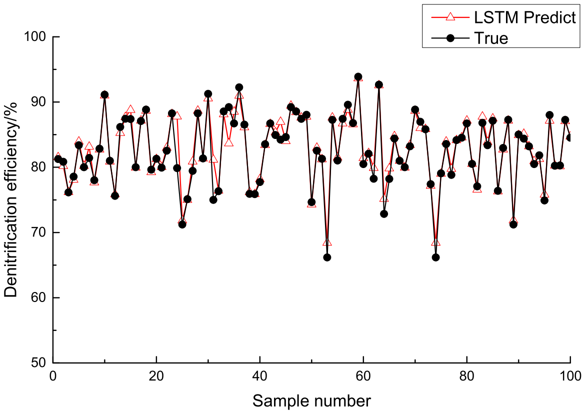


FIGURE 6. Prediction of LSTM model.

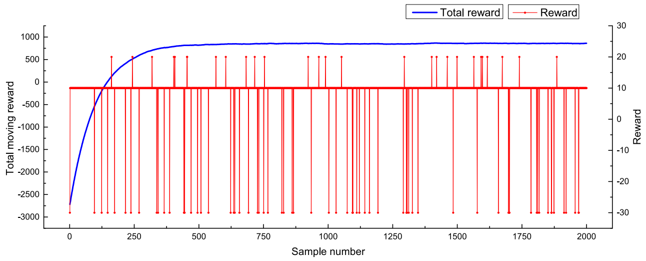


FIGURE 7. Total reward value and reward value results.

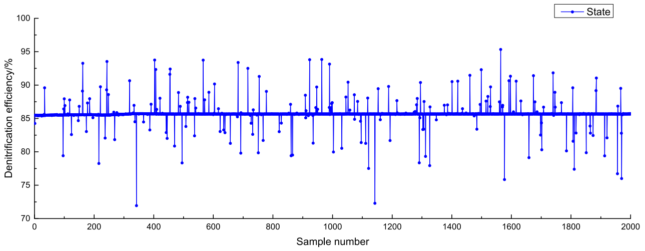


FIGURE 8. Status value prediction results.

TABLE 5. Combined solution with optimal denitrification efficiency under variable Boiler load.

Boiler load /(MW)	**300-400**	**401-500**	**501-600**	**601-700**	**751.96**	**701-800**	**801-900**	**901-1000**
Inlet flue gas temperature / ^{o}C	356.41	337.67	314.83	354.07	**335.74**	335.07	319.82	341.22
Inlet O_2 mass concentration /(mg· m^{-3})	1.18	1.47	1.48	3.73	**4.53**	2.14	6.24	6.49
Inlet NO_x mass concentration /(mg· m^{-3})	179.03	180.15	141.83	132.94	**337.82**	207.17	179.13	138.10
Ammonia/air mixer ammonia inlet regulator valve	27.85	54.81	33.89	62.72	**59.09**	69.73	85.42	27.64
Spray ammonia mass flow /(kg / h)	155.12	107.52	149.53	86.11	**129.08**	105.87	151.01	187.30
Predicted denitrification efficiency /%	93.53	93.76	95.33	93.84	**85.66**	90.51	93.79	92.40

TABLE 6. Combined solution with lowest cost and denitrification efficiency reach the standard under variable Boiler load.

Boiler load /(MW)	**300-400**	**401-500**	**501-600**	**601-700**	**701-800**	**801-900**	**901-1000**
Inlet flue gas temperature / ^{o}C	326.95	347.48	348.17	344.15	322.00	333.87	348.73
Inlet O_2 mass concentration /(mg· m^{-3})	3.57	4.43	4.86	3.52	4.65	3.07	6.10
Inlet NO_x mass concentration /(mg· m^{-3})	377.41	196.92	131.56	166.52	196.93	445.11	489.41
Ammonia/air mixer ammonia inlet regulator valve	41.19	78.06	44.43	32.11	21.70	73.10	88.82
Spray ammonia mass flow /(kg / h)	**50.08**	**29.82**	**47.82**	**29.38**	**38.86**	**31.76**	**34.72**
Predicted denitrification efficiency /%	89.29	88.92	86.95	93.38	87.32	85.82	88.17

concentration is adjusted to 4.53 mg· m^{-3}, the Inlet NO_x mass concentration is adjusted to 337.82 mg· m^{-3}, the Ammonia/air mixer ammonia inlet regulator valve position feedback is adjusted to 59.09 and the Spray ammonia mass flow is adjusted to 129.08 kg/h, the denitrification efficiency will be 85.66%.

Since the Boiler load cannot be adjusted manually. If the 1000 MW unit is in a stable load state. This is equivalent to a Boiler load of about 1000 MW. At this time, the control strategy with a Boiler load of 901-1000 MW should be adopted. If the 1000 MW unit is under variable load condition, refer to Table 5 for the control strategy of the optimal denitrification efficiency. If the cost is considered, the cost of the SCR denitrification process mainly depends on the amount of Spray ammonia mass flow. Therefore, the control strategy with the lowest cost and the denitrification efficiency meeting the standard can refer to Table 6.

VI. CONCLUSION

As the main indicator of SCR denitrification system, denitrification efficiency has great significance for the denitrification system and even the entire power generation unit. Quickly and accurately to predict the denitrification efficiency can contribute the stable operation of unit. The control strategy to obtain the optimal combination of input variable can optimize the energy conservation and emission reduction of the unit. In this paper, we aim at the characteristics of multiple parameters, multiple variables and mutual coupling of coal-fired boilers. First of all, we use correlation analysis and PCA to perform dimensionality reduction on the data of all variables to remove the coupling property between the initial variables. We then adopt the LSTM model which can take advantage of the time series characteristics of industrial process data. Experimental verification of LSTM model by using actual operating data of a coal-fired power plant in Guangdong. Experimental results demonstrate that our approach has better performance than other machine learning approaches. Finally, the A3C algorithm is used to construct the DRL model with the LSTM model to realize the control strategy of SCR denitrification efficiency in coal-fired power plants, which can achieve energy conservation and the emission reduction of NO_x. The experimental results demonstrate that our proposed approach is feasible and universally applicable in industry.

Regarding future directions, we will further investigate the integrated control method of the combustion process, denitrification process and system cost of coal-fired power plants to improve the stability and reliability of control strategies. In particular, we will first build an integrated optimization control model for the entire life cycle of NO_x generation and emissions. At the same time, we will improve the universal applicability of the model and work on researching machine learning models that can be widely used in other industries.

REFERENCES

[1] N. Lu, B. Wei, and Q. Zhu, "Analysis on the construction demand of regional air pollution control management system," *Chin. J. Environ. Manage.*, vol. 7, no. 6, pp. 66–70, 2015.

[2] W. Liu, J.-F. Zhang, and Z. Tong, "Progress in study of NO_x with the selective catalytic reduction," *Ind. Saf. Environ. Protection*, no. 1, pp. 25–28, 2005.

[3] N. Cai, J.-J. Shi, and J.-X. Tang, "Research progress on SCR denitrification technology for industry tail gas," *Guangzhou Chem. Ind.*, vol. 47, no. 1, pp. 14–15 and 23, 2019.

[4] H.-Z. Chen, H.-X. He, and Y.-M. Wan, "REN baozeng. Research progress of coalfired flue gas denitrification technology," *Appl. Chem. Ind.*, vol. 48, no. 5, pp. 1146–1151 and 1155, 2019.

[5] Z.-J. Fang, L.-P. Jin, and M.-L. Yu, "Research on optimization adjustment for ammonia injection and operation of SCR denitrification system in coalfired power plant," *Electr. Power Environ. Protection*, vol. 31, no. 6, pp. 39–42, 2015.

[6] S.-C. Ma, Y. Deng, and W.-L. Wu, "Experimental research on ABS formation characteristics in SCR denitrification process," *J. Chin. Soc. Power Eng.*, vol. 36, no. 02, pp. 143–150, 2016.

[7] T.-M. Qin, J.-Z. Liu, and T.-T. Yang, "SCR denitration system modeling and operation optimization simulation for thermal power plant," *Proc. CSEE*, vol. 36, no. 10, pp. 2699–2703, 2016.

[8] Z.-H. Zhao, C. Han, and W.-J. Zhao, "NO_x content prediction based on principal component analysis and multivariable process monitoring," *Thermal Power Gener.*, vol. 45, no. 7, pp. 98–103, 2016.

[9] K. Yu, L. Jia, Y. Chen, and W. Xu, "Deep learning: Yesterday, today, and tomorrow," *J. Comput. Res. Develop.*, vol. 50, no. 9, pp. 1799–1804, 2013.

[10] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[11] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-cecoder for statistical machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, Doha, Qatar, 2014, pp. 1724–1734.

[12] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA, Jun. 2014, pp. 1725–1732.

[13] Y.-X. Li, J.-Q. Zhang, D. Pan, and H. Dan, "A study of speech recognition based on RNN-RBM language model," *J. Comput. Res. Develop.*, vol. 51, no. 9, pp. 1936–1944, 2014.

[14] Z.-J. Sun, L. Xue, and Y.-M. Xu, "Overview of deep learning," *Appl. Res. Comput.*, vol. 29, no. 8, pp. 2806–2810, 2012.

[15] X. Wang, L. T. Yang, X. Chen, J.-J. Han, and J. Feng, "A tensor computation and optimization model for cyber-physical-social big data," *IEEE Trans. Sustain. Comput.*, vol. 4, no. 4, pp. 326–339, Dec. 2019, doi: 10.1109/TSUSC.2017.2777503.

[16] X. Wang, L. T. Yang, H. Liu, and M. J. Deen, "A big data-as-a-service framework: State-of-the-art and perspectives," *IEEE Trans. Big Data*, vol. 4, no. 3, pp. 325–340, Sep. 2018, doi: 10.1109/TBDATA.2017.2757942.

[17] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[18] A. Graves, *Long Short-Term Memory*. Berlin, Germany: Springer, 2012, pp. 1735–1780.

[19] J. Zhou, Y. Lu, H.-N. Dai, H. Wang, and H. Xiao, "Sentiment analysis of chinese microblog based on stacked bidirectional LSTM," *IEEE Access*, vol. 7, pp. 38856–38866, 2019.

[20] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transp. Res. C, Emerg. Technol.*, vol. 54, pp. 187–197, May 2015.

[21] Z. Chen and L.-X. Sun, "Short-term electrical load forecasting based on deep learning LSTM neural networks," *Electron. Technol.*, vol. 47, no. 1, pp. 39–41, 2018.

[22] X. Wang, J. Wu, and C. Liu, "Exploring LSTM based recurrent neural network for failure time series prediction," *J. Beijing Univ. Aeronaut. Astronaut.*, vol. 44, no. 4, pp. 772–784, 2018.

[23] J. Fu, H. Xiao, T. Wang, R. Zhang, L. Wang, and X. Shi, "Prediction model of desulfurization efficiency of coal-fired power plants based on long short-term memory neural network," in *Proc. Int. Conf. Internet Things (iThings), IEEE Green Comput. Commun. (GreenCom), IEEE Cyber, Phys. Social Comput. (CPSCom), IEEE Smart Data (SmartData)*, Jul. 2019, pp. 40–45.

[24] Y. Gao, R.-Y. Zhou, H. Wang, and Z.-X. Cao, "Study on an average reward reinforcement learning algorithm," *Chin. J. Comput.*, vol. 30, no. 8, pp. 1372–1378, 2007.

[25] S. Levine, C. Finn, T. Darrell, and P. Abbeel, "End-to-end training of deep visuomotor policies," *J. Mach. Learn. Res.*, vol. 17, no. 39, pp. 1–40, 2016.

[26] A. Sallab, M. Abdou, E. Perot, and S. Yogamani, "Deep reinforcement learning framework for autonomous driving," *Electron. Imag.*, vol. 2017, no. 19, pp. 70–76, Jan. 2017.

[27] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, Jan. 2016.

[28] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 1988.

[29] R. S. Sutton, D. A. Mcallester, and S. P. Singh, "Policy gradient methods for reinforcement learning with function approximation," in *Proc. Adv. Neural Inf. Process. Syst.*, Denver, CO, USA, 1999, pp. 1057–1063.

[30] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015.

[31] L. J. Lin, "Self-improving reactive agents based on reinforcement learning, planning and teaching," *Mach. Learn.*, vol. 8, nos. 3–4, pp. 293–321, 1992.

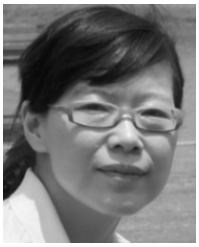
[32] N. Heess, G. Wayne, D. Silver, T. Lillicrap, T. Erez, and Y. Tassa, "Learning continuous control policies by stochastic value gradients," in *Proc. Adv. Neural Inf. Process. Syst.*, Montreal, QC, Canada, 2015, pp. 2944–2952.

[33] T. P. Lillicrap, J. J. Hunt, and A. Pritzel, "Continuous control with deep reinforcement learning," *Comput. Sci.*, vol. 8, no. 6, p. A187, 2015.

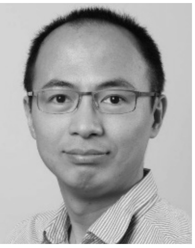
[34] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," in *Proc. Int. Conf. Mach. Learn.*, New York, NY, USA, 2016, pp. 1928–1937.

[35] J.-Q. Gao, S.-Y. Liang, and X.-Y. Wu, "Mechanism modeling and operation simulation for 1000 MW boiler flue gas denitration system," *Hebei Electr. Power*, vol. 38, no. 3, pp. 3–8 and 33, 2019.

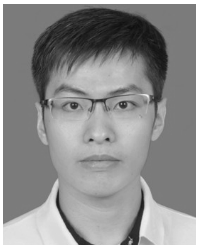
JIGAO FU is currently pursuing the M.Sc. degree in computer technology with the Faculty of Computer, Guangdong University of Technology. His research interests include industrial big data analytics and machine learning.



HONG XIAO received the M.Sc. degree from the Huazhong University of Science and Technology, China, in 2012, and the Ph.D. degree from the South China University of Technology, China, in 2005. She is currently an Associate Professor with the School of Computers, Guangdong University of Technology. Her research interests include big data analytics and the industrial internet of things.



HAO WANG (Member, IEEE) received the B.Eng. and Ph.D. degrees in computer science and engineering from the South China University of Technology, Guangzhou, China, in 2006. He is currently an Associate Professor with the Norwegian University of Science and Technology, Gjøvik, Norway. He has authored or co-authored more than 130 articles in reputable international journals and conferences papers. His current research interests include big data analytics, the industrial internet of things, high performance computing, and safety-critical systems. He is a member of the IEEE IES Technical Committee on Industrial Informatics. He served as a TPC Co-Chair for the IEEE DataCom 2015, IEEE CIT 2017, and ES 2017. He is a Reviewer for many journals, such as the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, and the IEEE TRANSACTIONS ON BIG DATA.



JUNHAO ZHOU is currently pursuing the Ph.D. degree in computer technology and application with the Faculty of Information Technology, Macau University of Science and Technology. His research interests include machine learning and deep learning.

• • •