frontiers | Frontiers in Energy Research

METHODS
published: 12 January 2023
doi: 10.3389/fenrg.2022.953387

Combustion Optimization Under Deep Peak Shaving Based on DYNA-A3C With Dynamic Weight

*Tang Wei-Jie**, *Wang Hai-Tao*, *Liu Ping-Ji* and *Qian Feng-Lei*

China Energy Engineering Group Jiangsu Power Design Institute Co., Ltd., Nanjing, China

The combustion process of boilers under deep peak shaving is a multivariate process which has complex characteristics such as super multivariability, being nonlinear, and large delay. It is difficult to handle complex data and calculate appropriate distributed results. To this end, this study applies the A3C method based on the dynamic weight Dyna structure to the boiler combustion system. This method trains and optimizes the boiler combustion system by establishing a data center and designing appropriate states and reward values, and the simulation results show that this method can be used to optimize the boiler combustion system. It can effectively reduce NO_X emissions and improve the boiler combustion efficiency.

Keywords: data center, deep reinforcement learning, deep peak shaving, the combustion system, A3C

OPEN ACCESS

Edited by:
Xiangdong Liu,
Yangzhou University, China

Reviewed by:
Jianhong Lv,
Southeast University, China
Yi Zhang,
North China Electric Power University, China

****Correspondence:***
Tang Wei-Jie
tangweijie@jspdi.com.cn

Specialty section:
This article was submitted to Process and Energy Systems Engineering, a section of the journal Frontiers in Energy Research

Received: *26 May 2022*
Accepted: *09 June 2022*
Published: *12 January 2023*

Citation:
Wei-Jie T, Hai-Tao W, Ping-Ji L and Feng-Lei Q (2023) Combustion Optimization Under Deep Peak Shaving Based on DYNA-A3C With Dynamic Weight. Front. Energy Res. 10:953387. doi: 10.3389/fenrg.2022.953387

INTRODUCTION

In order to accelerate the completion of China's carbon emission and carbon neutrality goals and obtain digital transformation through data center in the energy industry, the National Energy Administration proposed the implement linkage of three reforms of a coal-fired power unit. The main technical difficulty is how to make the large-capacity coal-fired power unit perform deep peak shaving to the ultra-low load more digitally.

Using the data center to train the combustion system can effectively obtain the maximum amount of information. The boiler combustion process is a complex process with multi-variables, nonlinearity, and large delay. In particular, under deep peak shaving operation, the decrease in the load may lead to instability of boiler combustion, ineffective operation of the denitrification system, over-temperature of the tube wall of the boiler (Shi et al., 2019), etc. How to effectively ensure boiler efficiency and NOx emissions is an important research issue for combustion optimization under deep peak shaving.

The current research on the data application is mainly divided into two aspects. On the one hand, the boiler combustion process is optimized based on optimization algorithms, and the optimization is carried out with the goal of boiler combustion efficiency and environmental protection parameters, such as genetic algorithm (Dal Secco et al., 2015; Pan et al., 2018), particle swarm algorithm (Fang et al., 2012; Sanaye and Hajabdollahi, 2015; Xu et al., 2019), ant colony algorithms (Xu et al., 2008). But the speed of optimization is slow and easy to fall into local optimum. Particularly in deep peak shaving, the boiler combustion situation is more complicated. On the other hand, it is optimized by training neural networks, according to Li and Niu (2016) and Han et al. (2022), in which deep reinforcement learning in the data center has become the focus which has the ability of generalization and decision-making. Bouhamed et al. (2020) and Zou et al. (2020) proposed a deep deterministic policy gradient (DDPG) algorithm based on the actor-critic (AC) framework, which was used to update the policy when solving the DRL problem. Ye et al. (2021) suggested an asynchronous dominant actor-critic

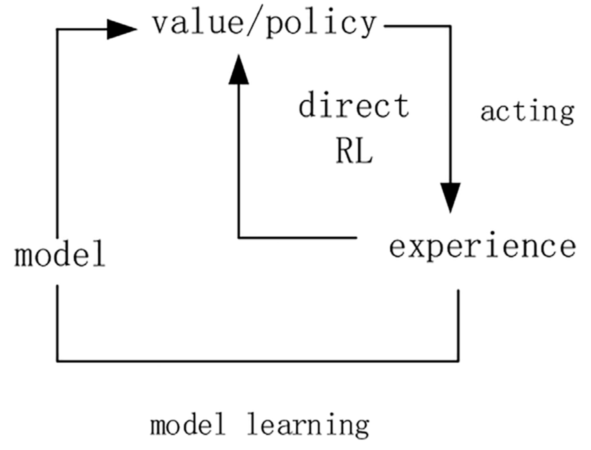


FIGURE 1 | Schematic diagram of the Dyna structure (Tong et al., 2018).

algorithm (AC), which used the multi-threading function of CPU to construct multiple agents in parallel and asynchronously for training at the same time. Therefore, at any time, due to the different states experienced by the parallel agents, the purpose of reducing the correlation between samples in the training process was achieved.

However, sometimes learning through environmental feedback from the data center may cause low learning efficiency, the Dyna structure (as shown in **Figure 1**) can enable agents to act in a virtual environment and generate virtual samples for learning, and combine them with learning experience in the actual environment to improve learning efficiency. So this study improved the asynchronous deep reinforcement learning algorithm (A3C) based on the Dyna structure as an optimization method to find the optimal boiler efficiency and NOx emission, so as to achieve the optimal control target. The simulation results show that the boiler combustion control optimization method based on DW-DYNA-A3C is an effective optimization method.

ESTABLISHMENT OF THE DW-DYNA-A3C METHOD

DW-Dyna-A3C Method

The learning efficiency of deep reinforcement learning is the main factor affecting the application effect. For this reason, this study considers the use of asynchronous methods to improve the learning efficiency. The asynchronous method refers to constructing different environment instances in multiple threads, and using multiple agents in parallel to interact with the environment. The asynchronous method enables the independent exploration in each agent of the thread, and multiple agents will share the acquired experience after joint exploration and parallel computing, and A3C is one such approach.

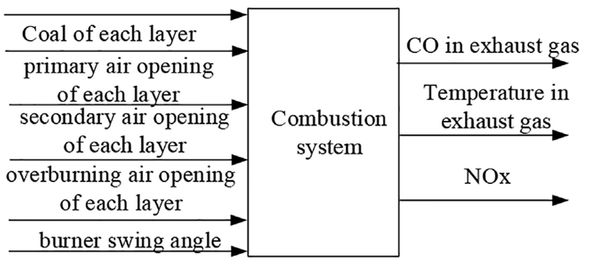

Coal of each layer
primary air opening of each layer
secondary air opening of each layer
overburning air opening of each layer
burner swing angle
Combustion system
CO in exhaust gas
Temperature in exhaust gas
NOx

FIGURE 2 | Input and output variables of the boiler combustion system.

θ_a is the network parameter of critics shared globally, θ_v is the network parameter of the actors, θ_a' and θ_v' are the network parameter of critics and actors of a single thread, and then the gradient accumulation formula of the actors is:

$$d\theta_a \leftarrow d\theta_a + \nabla_{\theta_a'} \log \pi\left(a_i \middle| s_i; \theta_a'\right)\left(R - V\left(s_i; \theta_v'\right)\right), \tag{1}$$

where π is the strategy which refers to the state-to-action mapping, a_i is the action determined by the current strategy (π), and R is the cumulative reward. The critic network cumulative gradient is.

$$d\theta_v \leftarrow d\theta_v + \partial\left(R - V\left(s_i; \theta_v'\right)\right)^2 \Big/ \partial\theta_v'. \tag{2}$$

Although the asynchronous strategy can improve the training speed, the learning process is still very slow if the number of samples obtained is insufficient. Therefore, consider putting the agent in a virtual environment, generating virtual samples for learning, and combining it with the learning experience in the actual environment (Liu et al., 2021). Therefore, it is proposed to add a Dyna structure to each thread in A3C to reduce the interaction with the real environment as well as improve the utilization of the virtual environment.

However, since there is a certain gap between the virtual environment model and real environment, if the learning results in the virtual environment are always dominant, it may cause wrong learning results. Therefore, the dynamic weight method is used to tackle the problem. When meeting the higher cumulative reward of the agent interacting with the real environment or the larger number of global updates, the learning result of the agent in the virtual environment will have less impact.

When updating the network parameters of actors and critics, dynamic weights μ is introduced to the virtual environment model in addition to its own learning rate, which is expressed as

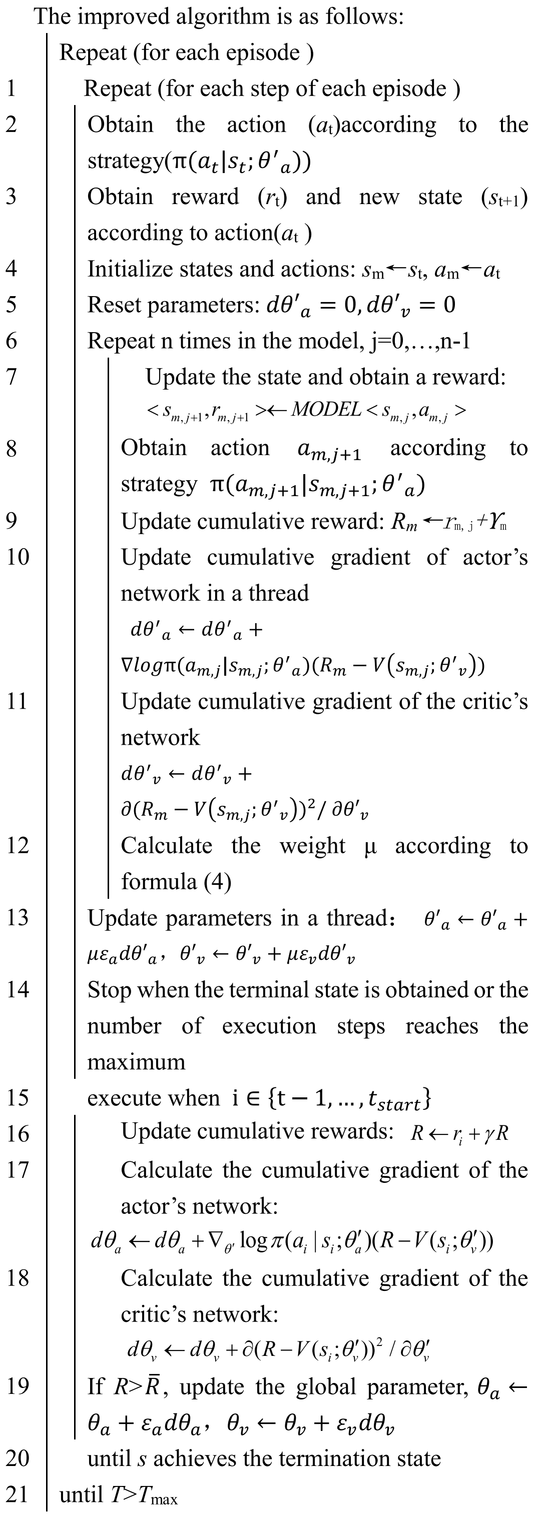
$$\theta_a' \leftarrow \theta_a' + \mu\varepsilon_a d\theta_a', \theta_v' \leftarrow \theta_v' + \mu\varepsilon_v d\theta_v', \tag{3}$$

$$\mu = \left(1 - \frac{T}{T_{\max}}\right) e^{-\frac{R_m}{(j+1) r_{\max}}}. \tag{4}$$

In **formula (4)**, R_m is the cumulative reward in the virtual environment model. When R_m<0, R_m is set to zero to prevent the cumulative reward from being less than zero, which will cause the weight to be overlarge and fail to converge. r_m is the maximum reward given by the virtual environment for each step. j is the number of repeated executions in the virtual environment. μ decreases with the increase of global parameter updates' number and the cumulative reward. T is the global shared count.

METHOD FLOW

The improved algorithm is as follows:



```
The improved algorithm is as follows:
Repeat (for each episode )
  Repeat (for each step of each episode )
    Obtain the action ($a_t$)according to the strategy($\pi(a_t|s_t;\theta'_a)$)
    Obtain reward ($r_t$) and new state ($s_{t+1}$) according to action($a_t$ )
    Initialize states and actions: $s_m \leftarrow s_t$, $a_m \leftarrow a_t$
    Reset parameters: $d\theta'_a = 0, d\theta'_v = 0$
    Repeat n times in the model, j=0,…,n-1
        Update the state and obtain a reward:
        $< s_{m,j+1}, r_{m,j+1} > \leftarrow MODEL < s_{m,j}, a_{m,j} >$
      Obtain action $a_{m,j+1}$ according to strategy $\pi(a_{m,j+1}|s_{m,j+1};\theta'_a)$
      Update cumulative reward: $R_m \leftarrow r_{m,j} + \Upsilon_m$
      Update cumulative gradient of actor's network in a thread
      $d\theta'_a \leftarrow d\theta'_a + \nabla log\pi(a_{m,j}|s_{m,j};\theta'_a)(R_m - V(s_{m,j};\theta'_v))$
      Update cumulative gradient of the critic's network
      $d\theta'_v \leftarrow d\theta'_v + \partial(R_m - V(s_{m,j};\theta'_v))^2 / \partial\theta'_v$
      Calculate the weight μ according to formula (4)
    Update parameters in a thread： $\theta'_a \leftarrow \theta'_a + \mu\varepsilon_a d\theta'_a$， $\theta'_v \leftarrow \theta'_v + \mu\varepsilon_v d\theta'_v$
    Stop when the terminal state is obtained or the number of execution steps reaches the maximum
  execute when $i \in \{t-1, \dots, t_{start}\}$
      Update cumulative rewards: $R \leftarrow r_i + \gamma R$
      Calculate the cumulative gradient of the actor's network:
    $d\theta_a \leftarrow d\theta_a + \nabla_{\theta'} \log \pi(a_i | s_i;\theta'_a)(R - V(s_i;\theta'_v))$
      Calculate the cumulative gradient of the critic's network:
      $d\theta_v \leftarrow d\theta_v + \partial(R - V(s_i;\theta'_v))^2 / \partial\theta'_v$
    If $R > \bar{R}$, update the global parameter, $\theta_a \leftarrow \theta_a + \varepsilon_a d\theta_a$， $\theta_v \leftarrow \theta_v + \varepsilon_v d\theta_v$
  until $s$ achieves the termination state
until $T > T_{max}$
```

ε_a and ε_v are the learning rates of the actor and critic networks, R is the cumulative reward, T is the number of global updates, and the subscript m is the parameters of the algorithm in the virtual model.

The DW-Dyna-A3C algorithm adds an evaluation mechanism for the results of each thread on the basis of the original push mechanism in order to avoid pushing the results of poor operation in a thread to the global and thus affecting the speed and accuracy of convergence. If the cumulative reward of thread running is lower than the average of the cumulative reward of other threads in the last running, this update is only copied from the global parameters to the thread, and the update is not pushed to the global.

DESIGN OF THE COMBUSTION OPTIMIZATION SYSTEM BASED ON DW-DYNA-A3C

Analysis of the Combustion System

In the process of load lifting and lowering, the coordinated control system can calculate the total amount of coal and air required under different loads, after that, it was distributed to the burners and the air of each layer. The distribution method will directly affect the combustion efficiency of the boiler and the emission of NOx in the gas (Wang et al., 2018). At present, the coal of burners in each layer is usually distributed equally, and the air is allocated empirically, which is obviously not the optimal solution.

The boiler efficiency is generally calculated by the reverse balance method (Cheng et al., 2018), in which q_2 is expressed as exhaust heat loss, %. In addition, it is the largest in boiler heat loss and is closely related to temperature of exhaust gas (T_e). q_3 is the loss of the inadequacy burning, %, which is observed on-site by measuring the CO concentration in the exhaust gas while none of the other losses can be measured in real-time (Adams et al., 2021), Therefore, the boiler efficiency is mainly represented by the exhaust gas temperature and the CO concentration in the exhaust gas in this study.

Under deep peak shaving, the wall temperature of the heating surface is the key factor restricting the adjustment. Due to the reduction of the working fluid, the heat transfer of the hydrodynamic cycle is deteriorated, resulting in the over-temperature.

The main objectives of the boiler combustion optimization control system are to (as shown in **Figure 2**): 1) reduce the amount of CO in the exhaust gas (CO); 2)reduce the NOx content at the SCR inlet(NO_x); 3) ensure that temperature of waterwall (Tp) is not overheated; 4) minimize q2.

State Design

The state of the agent can best reflect the optimization goal and the optimization system. Therefore, the state quantity of the combustion optimization system should be composed of the target set value, the actual value, the adjustment value, and the deviation. The set values include the amount of carbon monoxide CO_{sp}, the exhaust gas temperature $T_{e,sp}$, and the set value of NOx concentration $NO_{x,sp}$. The total amount of air D_{AIR}, i-th layer primary air opening $V_{s,j}$, j-th layer secondary air opening $V_{s,j}$, k-layer overburning air opening $V_{c,k}$, and burner swing angle A_f;

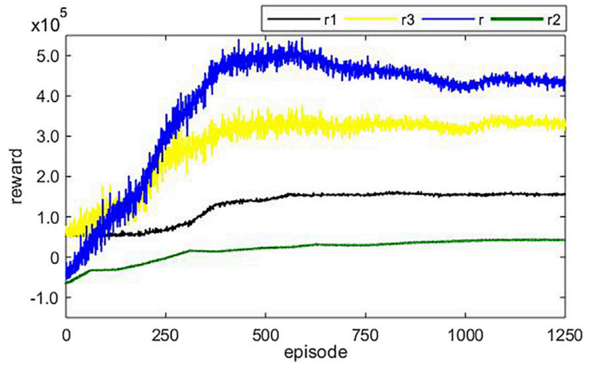


FIGURE 3 | Input and output variables of the boiler combustion system.

e_{Cx}, e_{Te}, e_{NOx} are CO deviation, T_e deviation, and NO$_x$ deviation, respectively. In the optimization process, considering the safety and economy of boiler, the safety margin ΔT_P between the actual maximum wall of T_P and the over-temperature value should be added to the state (Dzikuć et al., 2020). Therefore, the system at time t has a state S_t.

$$S_t = \{T_{e,sp}, NO_{x,sp}, CO_{sp}, D_{AIR}, D_{B,n}, V_{f,i}, V_{s,j} V_{c,k}, A_f, CO, Te, NOx, e_{co}, e_{Te}, e_{NOx}, \Delta T_p\}. \tag{5}$$

Reward Design

Rewards should be able to promote deep reinforcement learning to the optimal strategy. In the reward design, the agent should continue to be rewarded when it learns the optimal strategy, and at the same time, the agent should meet various constraints of the system, such as the rapidity of adjustment, the stability, and the rate of change of the control quantity, so the reward design for the coordinated control system is divided into the following aspects.

Continuing Reward Items

The continuous reward should ensure that the reward increases with the decrease of the deviation in the optimization process, and the reward value reaches the maximum and remains unchanged when the system reaches the optimal value. Since there are three optimization objectives of the combustion optimization system, it is necessary to carry out weighted processing for the deviation of each optimized variable.

$$e_t = [|e_{Te}|, |e_{NOx}|, |e_{co}|] \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix}. \tag{6}$$

In **Formula (6)**, $[\lambda_1, \lambda_2, \lambda_3]^T$ is the deviation weight matrix, whose proportion can be modified according to the regulating target of the combustion system. Take the weighted deviation e_t at time t as an important reference for the continuous reward.

TABLE 1 | Parameters of the actor network and critic network.

Description	Value
Maximum number of global updates	1200
Maximum number of updates for a thread	8000
Actor network learning rate	0.00001
Critic network learning rate	0.00001
Virtual environment model repeats	100
Thread rewards update discount factor	0.99
Global update frequency	5

$$r_1 = \frac{10}{e_t^2 + 1}. \tag{7}$$

Limit Term of the Change Rate of the Control Quantity

In the optimization process, considering that the fast change rate of pipe wall temperature will produce thermal stress and reduce its service life and the adjustment rate of each actuator also has certain limits, it is necessary to limit the change rate of control quantity. The output of the actor network is the increment of each regulation quantity, so it only needs to judge the upper and lower limits of the actor network output, and then reward and punish the reward value.

Suppose that the sampling period of the algorithm is T_s, and the output of the actor network at time t is $\Delta D_{AIR}, \Delta V_{f,i}, \Delta V_{s,j}, \Delta V_{c,k}, \Delta A_f$, so the limiting term of the control variable rate of change is

$$r2 = \begin{cases} 0, & if \begin{cases} \Delta D_{AIR}/T_s \in [d_{AIR\min}, d_{AIR\max}], \\ \Delta V_{f,i}/T_s \in [v_{f\min}, v_{f\max}], \\ \Delta V_{s,j}/T_s \in [v_{s\min}, v_{s\max}], \\ \Delta V_{c,k}/T_s \in [v_{c\min}, v_{c\max}], \\ \Delta A_f/T_s \in [a_{f\min}, a_{f\max}]. \end{cases} \\ -20, & else \end{cases} \tag{8}$$

In **Formula (8)**, $d_{\mathrm{AIRmin}}, d_{\mathrm{AIRmax}}, d_{AIR\min} andd_{AIR\max}$ are the lower limit and upper limit of the coal quantity adjustment rate; $v_{\mathrm{smin}}, v_{\mathrm{smax}}, v_{\mathrm{fmin}}, v_{\mathrm{fmax}}, v_{\mathrm{cmin}}, andv_{\mathrm{cmax}}$ are the lower limit and upper limit of the adjusting rate of primary damper opening, secondary damper opening, and burnt out damper opening; $a_{fmin} anda_{fmax}$ are the lower limit and upper limit of the burner swing angle adjustment rate, when the three conditions are met at the same time, no punishment is given.

Auxiliary Tasks

The ultimate goal of combustion system optimization should improve boiler operation efficiency as much as possible and meet environmental protection requirements. Because combustion system optimization is a complex problem of multi-objective optimization, auxiliary tasks are added through weighted deviations, and when the optimization structure begins to gradually become better, and continuously increases the reward value. At the same time, when the maximum value of the boiler inner wall temperature is close to the over-temperature value, that is, when the safety margin ΔT_P is small, a penalty should be given to avoid this situation as much as possible, so the auxiliary task reward is

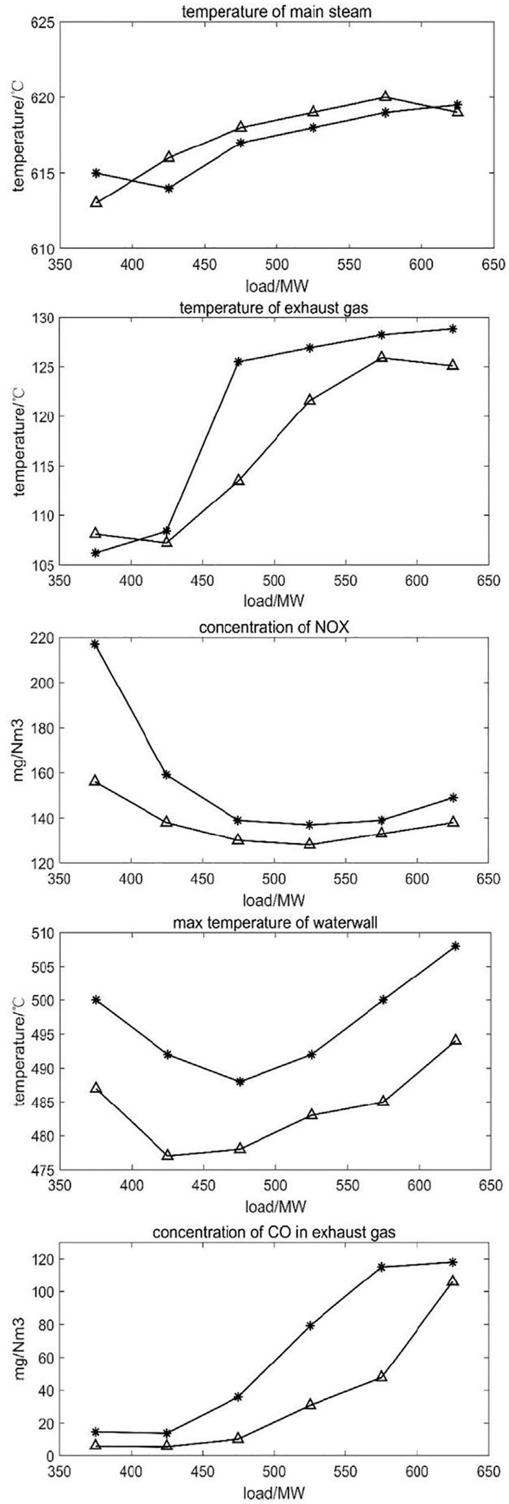


FIGURE 4 | Change in optimized variables from 350 to 650 MW ("*" represents effects before applying the method; "Δ" represents effects after applying the method.).

TABLE 2 | Comparison of the boiler efficiency before and after optimization.

Power range/MW	Boiler efficiency (before)	Boiler efficiency (after) (%)
300–400	91.55	91.89
400–500	92.32	92.44
500–600	92.45	92.76
600–700	92.88	92.92
700–800	93.12	93.22
800–1000	93.56	93.67

$$r3 = \begin{cases} 50, & if\,(|e_t| < 5), \\ \dfrac{5}{e_t^2 + 1}, & if\,(|e_t| < 10), \\ -10, & if\left(\Delta T_p < 5\right), \\ 0, & else. \end{cases} \tag{9}$$

So the final reward value for the boiler combustion system is.

$$r = r_1 + r_2 + r_3. \tag{10}$$

Simulation Experiment Research

Training Process

Taking the model of a 1000 MW boiler combustion system as an example, there are six layers of burners, six layers of primary air, 24 layers of secondary air, and eight layers of exhaust air, so n = 6, i = 6, j = 24, and k = 8, the rest of the set value and the range of the adjustment amount are set according to the boiler design manual. In the neural network structure, both the actor network and the critic network in the algorithm are designed as a 9-layer fully connected neural network structure. A total of 120 nodes were present, the output layer contains 1 node, the input layer of the actor network contains state information, and the output of the critic network, a total of 17 nodes were present, the middle hidden layer is the same as the critic network, the output layer has five nodes, and the output control amount is incremented. The rest of the relevant training parameters are shown in the **Table 1**.

Figure 3 shows the changing trend of the algorithm learning total reward value and each sub-item reward value, where r_1 is the continuous reward item, r_2 is the control amount change rate limit item, and r_3 is the auxiliary task item. As the number of learning increases, the total reward value of the system begins to increase rapidly after 200 episodes, and the algorithm basically converges around 600 episodes.

Simulation Experiment Under Deep Peak Shaving Conditions

After the training is completed, the trained algorithm is used to simulate the model under the condition of deep peak regulation. The load variation range is 350–650 MW, and the steady-state values of various indicators before and after training are observed.

As shown in **Figure 4** that after the optimization, the exhaust gas temperature of the boiler has decreased, and the main steam temperature has increased compared with that before the optimization, which indicates that the boiler efficiency has been improved throughout the load range of the simulation experiment, mainly because the adjustment after optimization, the air distribution method reduces the temperature of the inner wall of the furnace, which is about 15K lower than the maximum value of the optimized front wall temperature, leaving a sufficient safety margin for increasing the temperature of the main steam. The ratio of heat absorption is more reasonable. At the same time, the concentrations of CO and NOx have also decreased, indicating that the combustion in the furnace is more sufficient after optimization (Yang et al., 2019), and the NOx concentration is reduced by means of staged air distribution and oxygen-enriched combustion, which not only improves the economy of the boiler combustion system, but also improves the environmental performance.

As presented in **Table 2**, obviously the boiler efficiency after optimization is larger than before. Considering the pollutant emission constraints, the average efficiency of boiler is increased by increasing the temperature of main steam and reheat steam by improving the combustion quality.

CONCLUSION

This article studies the combustion optimization system under deep peak shaving. Because the boiler combustion system has complex characteristics such as nonlinearity and multi-variables, this study proposed the DW-Dyna-A3C method to study, train, and simulate the combustion system. The DW-Dyna-A3C method is a reward evaluation system that takes into account both the control and the controlled state so that it can meet the requirements of multivariable nonlinear system control.

The simulation results show that this method can effectively improve the boiler efficiency, reduce pollutant emissions, and obtain a better ratio effect under the working conditions of deep peak regulation.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

TW-J studied the method and wrote three sections of the manuscript. WH-T provided a simulation platform. LP-J wrote one section of the manuscript. QF-L provided running data.

FUNDING

This work is supported by China Energy Engineering Group Jiangsu Power Design Institute Co., Ltd. its project number is 32-JK-2022-020.

REFERENCES

Adams, D., Oh, D.-H., Kim, D.-W., Lee, C.-H., and Oh, M. (2021). Deep Reinforcement Learning Optimization Framework for a Power Generation Plant Considering Performance and Environmental Issues. *J. Clean. Prod.* 291, 125915. doi:10.1016/j.jclepro.2021.125915

Bouhamed, O., Ghazzai, H., Besbes, H., and Massoud, Y. (2020). "Autonomous UAV Navigation: A DDPG-Based Deep Reinforcement Learning Approach," in Proceeding of the 2020 IEEE International Symposium on Circuits and Systems (ISCAS) (IEEE), 1–5. doi:10.1109/iscas45731.2020.9181245

Cheng, Y., Huang, Y., Pang, B., and Zhang, W. (2018). ThermalNet: A Deep Reinforcement Learning-Based Combustion Optimization System for Coal-Fired Boiler. *Eng. Appl. Artif. Intell.* 74, 303–311. doi:10.1016/j.engappai.2018.07.003

Dal Secco, S., Juan, O., Louis-Louisy, M., Lucas, J.-Y., Plion, P., and Porcheron, L. (2015). Using a Genetic Algorithm and CFD to Identify Low NOx Configurations in an Industrial Boiler. *Fuel* 158, 672–683. doi:10.1016/j.fuel.2015.06.021

Dzikuć, M., Kuryło, P., Dudziak, R., Szufa, S., Dzikuć, M., and Godzisz, K. (2020). Selected Aspects of Combustion Optimization of Coal in Power Plants. *Energies* 13 (9), 2208. doi:10.3390/en13092208

Fang, Y., Qin, X., and Fang, Y. (2012). "Optimization of Power Station Boiler Coal Mill Output Based on the Particle Swarm Algorithm," in Proceeding of the 2012 IEEE International Conference on Industrial Engineering and Engineering Management(IEEE), 612–616. doi:10.1109/ieem.2012.6837812

Han, Z., Li, J., Hossain, M. M., Qi, Q., Zhang, B., and Xu, C. (2022). An Ensemble Deep Learning Model for Exhaust Emissions Prediction of Heavy Oil-Fired Boiler Combustion. *Fuel* 308, 121975. doi:10.1016/j.fuel.2021.121975

Li, G., and Niu, P. (2016). Combustion Optimization of a Coal-Fired Boiler with Double Linear Fast Learning Network. *Soft Comput.* 20 (1), 149–156. doi:10.1007/s00500-014-1486-3

Liu, X., Zhang, H., Long, K., Nallanathan, A., and Leung, V. C. (2021). Deep Dyna-Reinforcement Learning Based on Random Access Control in LEO Satellite IoT Networks. *IEEE Internet Things J* 103, 312–327. doi:10.1109/jiot.2021.3112907

Pan, H., Zhong, W., Wang, Z., and Wang, G. (2018). Optimization of Industrial Boiler Combustion Control System Based on Genetic Algorithm. *Comput. Electr. Eng.* 70, 987–997. doi:10.1016/j.compeleceng.2018.03.003

Sanaye, S., and Hajabdollahi, H. (2015). Thermo-economic Optimization of Solar CCHP Using Both Genetic and Particle Swarm Algorithms. *J. Sol. Energy Eng.* 137 (1). doi:10.1115/1.4027932

Shi, Y., Zhong, W., Chen, X., Yu, A. B., and Li, J. (2019). Combustion Optimization of Ultra Supercritical Boiler Based on Artificial Intelligence. *Energy* 170, 804–817. doi:10.1016/j.energy.2018.12.172

Tong, C., Niu, W., Xiang, Y., Bai, X., and Gang, L. (2019). Gradient band-based adversarial training for generalized attack immunity of A3C path finding. arXiv e-prints 1807, 6752. doi:10.48550/arXiv.1807.06752

Wang, C., Liu, Y., Zheng, S., and Jiang, A. (2018). Optimizing Combustion of Coal Fired Boilers for Reducing NOx Emission Using Gaussian Process. *Energy* 153, 149–158. doi:10.1016/j.energy.2018.01.003

Xu, Q., Yang, J., and Yang, Y. (2008). "Identification and Control of Boiler Combustion System Based on Neural Networks and Ant Colony Optimization Algorithm," in Proceeding of the 2008 7th World Congress on Intelligent Control and Automation (IEEE), 765–768. doi:10.1109/wcica.2008.4593018

Xu, X., Chen, Q., Ren, M., Cheng, L., and Xie, J. (2019). Combustion Optimization for Coal Fired Power Plant Boilers Based on Improved Distributed ELM and Distributed PSO. *Energies* 12 (6), 1036. doi:10.3390/en12061036

Yang, W., Wang, B., Lei, S., Wang, K., Chen, T., Song, Z., et al. (2019). Combustion Optimization and NOx Reduction of a 600 MWe Down-Fired Boiler by Rearrangement of Swirl Burner and Introduction of Separated Over-fire Air. *J. Clean. Prod.* 210, 1120–1130. doi:10.1016/j.jclepro.2018.11.077

Ye, Z., Zhang, D., Wu, Z. G., and Yan, H. (2021). A3C-based Intelligent Event-Triggering Control of Networked Nonlinear Unmanned Marine Vehicles Subject to Hybrid Attacks. *IEEE Trans. Intelligent Transp. Syst.* 75, 165–178. doi:10.1109/tits.2021.3118648

Zou, J., Hao, T., Yu, C., and Jin, H. (2020). A3C-Do: A Regional Resource Scheduling Framework Based on Deep Reinforcement Learning in Edge Scenario. *IEEE Trans. Comput.* 70 (2), 228–239. doi:10.1109/TC.2020.2987567

Conflict of Interest: TW-J, WH-T, LP-J, and QF-L were employed by the company China Energy Engineering Group Jiangsu Power Design Institute Co., Ltd.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.