Advances in Computer, Signals and Systems (2026)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2026.100207
ISSN 2371-8838 Vol. 10 Num. 2

Deep Reinforcement Learning-Based Adaptive PID Temperature Control Method for Scrap Aluminum Induction Heating System

Zhuoyue Xu[1,a], Jingwei Wang[1,b,*]

[1]*TUT Maritime College, Tianjin University of Technology, Tianjin, China*
[a]*15932413633@163.com*, [b]*wjw249146691@163.com*
**Corresponding author*

Keywords: Proximal Policy Optimization; Deep Reinforcement Learning; PID Control; Induction Heating

Abstract: To address the issues of temperature response lag, difficulty in real-time adjustment of control parameters, and insufficient system stability under complex operating conditions during induction heating of scrap aluminum, this paper proposes a proportional-integral-derivative (PID) temperature control method based on Proximal Policy Optimization (PPO) deep reinforcement learning. First, a first-order inertial pure time-delay mathematical model of the induction heating system is established to describe the thermal inertia and time-delay dynamic characteristics during the heating process. Second, a near-end strategy optimization algorithm is introduced into the proportional-integral-derivative (PID) controller parameter adjustment process. A reinforcement learning state space is constructed using the system temperature error, error rate of change, control output, and actual temperature. The proportional, integral, and derivative parameter corrections are used as action outputs to achieve online adaptive updating of the control parameters. Subsequently, a co-simulation experimental environment is built on the MATLAB/Simulink platform, and comparative analyses are conducted with traditional proportional-integral-derivative (PID) control, fuzzy proportional-integral-derivative (Fuzzy PID) control, and Sparrow Search Algorithm-optimized proportional-integral-derivative (SSA-PID) control method. Experimental results show that the proposed method can effectively reduce system temperature overshoot, shorten system rise time and settling time, and quickly recover the system's stable state under external disturbances, exhibiting good dynamic tracking performance and disturbance rejection capability. This research provides a new solution for intelligent temperature control of complex industrial induction heating systems.

1. Introduction

With the rapid development of industries such as aerospace manufacturing, automotive, and rail transportation, aluminum alloys are widely used in modern industrial production due to their advantages such as low density, high strength, good corrosion resistance, and recyclability [1-2]. In recent years, the amount of waste aluminum has been continuously increasing, and how to achieve efficient recycling and resource utilization of waste aluminum has become an important research direction in the fields of green manufacturing and energy conservation and emission reduction [3-4]. In the process of waste aluminum recycling, induction heating melting technology is widely used in the heat treatment and metal smelting of waste aluminum due to its advantages such as high heating efficiency [5], fast heating rate, high thermal energy utilization rate, and low environmental pollution. Compared with traditional gas heating methods, induction heating can achieve non-contact rapid heating and has better temperature regulation performance. Therefore, high-precision temperature control during the induction heating process is of great significance for improving the quality of aluminum recycling, reducing energy consumption, and ensuring stable operation of industrial production [6].

However, waste aluminum induction heating systems typically exhibit complex dynamic characteristics such as significant nonlinearity, large time delay, and strong coupling [7]. In actual heating processes, the system temperature is easily affected by factors such as thermal inertia, changes in material state, and external environmental disturbances, resulting in large fluctuations in the system's dynamic response. Traditional proportional-integral-derivative (PID) controllers are widely used in industrial temperature control systems due to their simple structure [8], ease of engineering implementation, and good stability [9]. However, the parameters of traditional PID controllers are usually tuned manually based on experience, and their control parameters remain fixed during operation, lacking online adaptive capability. When the system operating conditions change, fixed-parameter PID controllers are prone to problems such as large overshoot, slow response speed, and decreased disturbance rejection performance. This is especially true in induction heating systems with large time delays, where traditional PID control methods struggle to meet the requirements of high-precision dynamic temperature control [10-11].

To address the temperature control problems in complex industrial processes, researchers both domestically and internationally have proposed various intelligent optimization control methods to improve the performance of traditional PID control. For example, fuzzy PID control methods achieve online parameter adjustment through fuzzy rules, thereby improving system robustness; swarm intelligence optimization methods such as particle swarm optimization (PSO)[12], genetic algorithms (GA)[13], and sparrow search algorithms (SSA) achieve PID parameter optimization through global search, improving the system's dynamic response performance and steady-state control accuracy to some extent. In recent years, with the development of intelligent control theory, PID control methods based on swarm intelligence optimization algorithms have been widely studied in the field of industrial process control. However, most intelligent optimization PID methods still belong to offline parameter optimization mode, and their optimization results usually depend on fixed operating conditions, lacking real-time online learning capabilities. When the system is disturbed or the operating environment changes, traditional intelligent optimization algorithms struggle to quickly and dynamically adjust control parameters. Furthermore, some swarm intelligence optimization algorithms suffer from slow convergence speed, susceptibility to local optima, and insufficient generalization ability under complex operating conditions, thus limiting their application in complex nonlinear industrial control systems.

To address these issues, this paper proposes a deep reinforcement learning-based adaptive temperature control method for induction heating, combining the Proximal Policy Optimization (PPO) algorithm with a PID controller to construct a PPO-DRL-PID intelligent control framework. This method utilizes the interaction mechanism between the reinforcement learning agent and the industrial environment to achieve online dynamic adjustment of PID parameters, thereby improving the control system's adaptability to complex dynamic conditions. Meanwhile, considering the dynamic control characteristics of the induction heating system, a reward function is designed that comprehensively takes into account tracking error, overshoot, and control energy consumption to enhance the stability and robustness of the controller. A model of the induction heating control system for scrap aluminum is established on the MATLAB/Simulink platform. Comparative experiments with traditional PID control methods and intelligent optimized PID methods verify the effectiveness and superiority of the proposed method in terms of temperature tracking performance, dynamic response speed, and disturbance rejection capability.

2. Induction Heating System Modeling

To achieve high-precision temperature control during the induction heating process of scrap aluminum, it is necessary to first analyze the overall structure and dynamic characteristics of the induction heating system. Due to the significant thermal inertia, time delay, and nonlinear coupling characteristics of the induction heating process, establishing a reasonable mathematical model of the system is crucial for subsequent controller design and reinforcement learning strategy optimization. This chapter first introduces the composition and structure of the induction heating system, then establishes a first-order inertial pure time delay mathematical model suitable for temperature control analysis, and combines it with the PID control principle to lay the foundation for the subsequent PPO reinforcement learning controller design.

2.1. Induction Heating System Structure

The induction heating system for scrap aluminum mainly consists of a power supply module, an induction coil, a temperature detection module, a controller, and the object being heated. Its overall structure is shown in Figure 1.

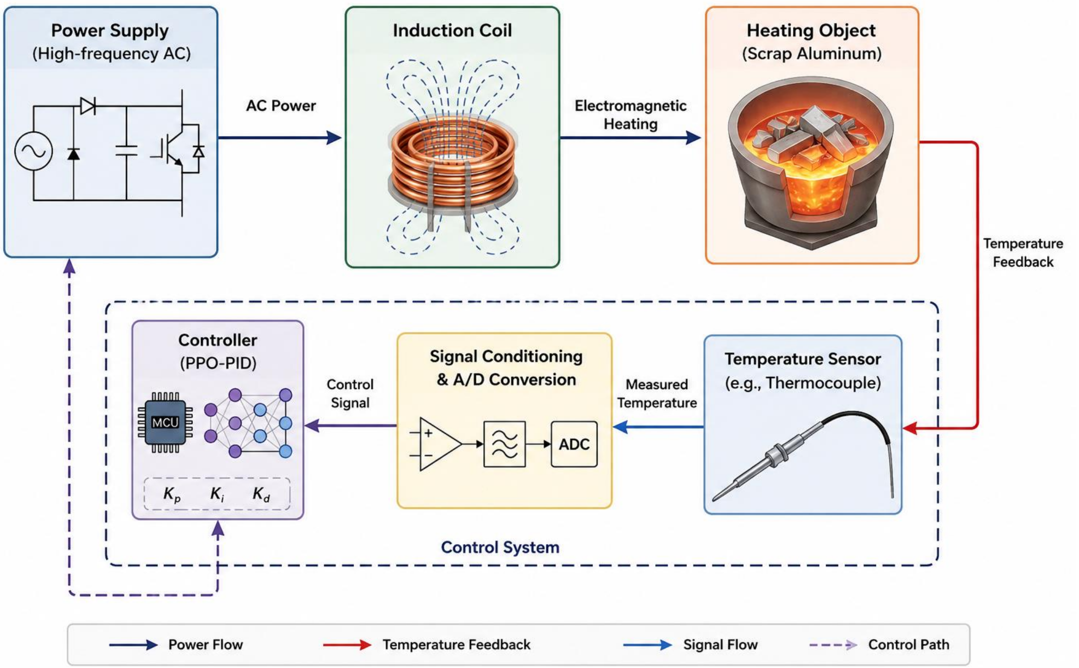

Figure 1: Structure of induction heating temperature control system.

During system operation, a high-frequency AC power supply provides alternating current to the induction coil, generating an alternating magnetic field inside the coil. This induces a current within the metal material, producing Joule heating and achieving rapid heating of the scrap aluminum. To ensure temperature stability during heating, the system uses a temperature sensor to collect the surface temperature of the heated object in real time and feeds the temperature signal back to the controller. The controller dynamically adjusts the input power based on the deviation between the target temperature and the actual temperature, thus achieving closed-loop temperature control.

Traditional induction heating systems typically use fixed-parameter PID controllers for regulation. However, due to the system's strong nonlinearity and thermal inertia, fixed-parameter control is difficult to adapt to complex operating conditions. Therefore, this paper introduces a PPO reinforcement learning algorithm into the traditional PID control framework. A reinforcement learning agent is used to achieve online dynamic adjustment of the PID parameters, improving the system's dynamic response and anti-disturbance performance.

2.2. Mathematical Modeling

The induction heating temperature control system is essentially a typical thermal process control object with significant time delay characteristics. During practical operation, due to heat conduction, material thermal inertia, and environmental heat dissipation, the system dynamic response generally exhibits obvious lag characteristics. To facilitate subsequent controller design and reinforcement learning algorithm training, this study adopts a first-order inertia plus dead-time model to describe the induction heating system. The transfer function of the model is expressed as follows:

$$G(s) = \frac{K}{Ts+1} e^{-\tau s} \tag{1}$$

where: K represents the system gain; T denotes the system time constant; τ represents the pure time delay; s is the Laplace operator.

The above model can effectively describe the dynamic temperature response of the induction heating system under varying input power conditions. Specifically, the system gain K determines the amplitude of temperature variation, the time constant T reflects the thermal inertia of the system, and the pure time delay τ describes the delay caused by heat transfer and measurement feedback processes. Owing to the strong thermal inertia in the induction heating process, the system usually exhibits obvious large-delay characteristics, which makes it difficult for traditional fixed-parameter control methods to achieve satisfactory control performance.

To realize real-time dynamic temperature regulation, a PID controller is adopted as the basic control structure in this study. The controller output can be expressed as follows:

$$u(t) = K_p e(t) + K_i \int e(t)dt + K_d \frac{de(t)}{dt} \tag{2}$$

where: $u(t)$ denotes the controller output; $e(t)$ represents the system error; K_p, K_i, and K_d denote the proportional, integral, and derivative coefficients, respectively.

Traditional PID controllers typically have fixed parameters and cannot be dynamically adjusted in real time according to system changes. When the operating environment changes or external disturbances occur, PID controllers with fixed parameters often exhibit problems such as increased overshoot, reduced response speed, and increased steady-state error.

3. PPO-DRL-PID Adaptive Control Method

3.1. PPO Reinforcement Learning Control Framework

Deep reinforcement learning is an intelligent decision-making method that enables an agent to learn optimal control strategies through continuous interaction with the environment. Its core objective is to maximize the long-term cumulative reward through iterative trial-and-error learning.

In the induction heating control system proposed in this study, the PPO reinforcement learning agent dynamically adjusts the PID parameters by observing the system operating state in real time, thereby realizing adaptive temperature control of the system [14-15]. During the control process, the reinforcement learning agent outputs PID parameter adjustment values, and the system generates control signals according to the updated PID parameters. Subsequently, the environment feeds back the new system state and reward value to complete the control policy update.

The system state transition process in the reinforcement learning framework can be expressed as follows:

$$s\{t+1\}=f(st,at) \tag{3}$$

where st denotes the current system state, at represents the action taken by the agent under the current state, $f(.)$ denotes the system dynamic transition function, and $s\{t+1\}$ represents the system state at the next time step.

To improve the stability of policy updates, the PPO algorithm adopts a clipped probability ratio mechanism to constrain the policy update range. The objective function is defined as follows:

$$L^{CLIP}(\theta) = \hat{E}_t\left[\min\left(r_t(\theta)\hat{A}_t, \mathrm{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t\right)\right] \tag{4}$$

where θ denotes the policy network parameters, $r_t(\theta)$ represents the probability ratio between the new and old policies, $\hat{A}_t$ denotes the estimated advantage function, ϵ is the clipping coefficient, and $\mathrm{clip}(.)$ denotes the clipping operation.

3.2. State Space Design

To enable the reinforcement learning agent to accurately perceive the dynamic characteristics of the induction heating system, it is necessary to construct an appropriate system state space. The design of the state space directly affects the learning capability of the reinforcement learning model regarding system dynamic features. Therefore, this study comprehensively considers the system temperature error, error variation rate, and control output to construct the following state vector:

$$s_t = [e(t), \Delta e(t), u(t), T(t)] \tag{5}$$

where $e(t)$ denotes the temperature error between the target temperature and the actual temperature, $\Delta e(t)$ represents the error variation rate, $u(t)$ denotes the current controller output signal, and $T(t)$ represents the current actual system temperature.

To further describe the error variation characteristics of the system, the system error is defined as follows:

$$e(t)=T\{ref\}(t)-T(t) \tag{6}$$

where $T\{ref\}(t)$ denotes the target reference temperature and $T(t)$ represents the actual system temperature.

The error variation rate can be expressed as follows:

$$\Delta e(t) = e(t) - e(t-1) \tag{7}$$

where $e(t)$ denotes the current error and $e(t-1)$ denotes the error at the previous time step.

3.3. Action Space Design

In the reinforcement learning control process, the action space is used to describe the regulation behavior adopted by the intelligent agent for the control system. In this study, a continuous action space is adopted, where the PPO agent dynamically outputs PID parameter adjustment values in real time to realize online adaptive tuning of the PID controller parameters. The action vector is defined as follows:

$$a_t = [\Delta K_p, \Delta K_i, \Delta K_d] \tag{8}$$

where ΔK_p, ΔK_i, and ΔK_d represent the adjustment values of the proportional, integral, and derivative parameters, respectively.

After being output by the reinforcement learning agent, the PID parameter update process is expressed as follows:

$$K_p^{t+1} = K_p^t + \Delta K_p \tag{9}$$

where K_p^t denotes the current proportional coefficient, ΔK_p denotes the proportional parameter adjustment value generated by the reinforcement learning agent, and K_p^{t+1} represents the updated proportional coefficient.

Similarly, the integral and derivative parameter update equations are defined as follows:

$$K_i^{t+1} = K_i^t + \Delta K_i \tag{10}$$

where K_i^t denotes the current integral coefficient, ΔK_i denotes the integral parameter adjustment value, and K_i^{t+1} represents the updated integral coefficient.

$$K_d^{t+1} = K_d^t + \Delta K_d \tag{11}$$

where K_d^t denotes the current derivative coefficient, ΔK_d denotes the derivative parameter adjustment value, and K_d^{t+1} represents the updated derivative coefficient.

3.4. Reward Function Design

The reward function is the core component that enables reinforcement learning algorithms to optimize control strategies, and its design directly affects the learning performance of the intelligent agent. To improve the dynamic control performance of the system, this study comprehensively considers the system tracking error, overshoot, and control energy consumption to construct the following reward function:

$$R_t = -(\alpha e(t)^2 + \beta M_p + \gamma u(t)^2) \tag{12}$$

where R_t denotes the reward value at the current time step, $e(t)$ denotes the system temperature error, M_p represents the system overshoot, $u(t)$ denotes the control output signal, and α, β, and γ denote the weighting coefficients corresponding to the error term, overshoot term, and control energy consumption term, respectively.

To further reduce the steady-state error of the system, the cumulative error is constrained using the following integral error index:

$$J = \int_0^T |e(t)| \, dt \tag{13}$$

where J denotes the cumulative absolute error index, $e(t)$ represents the system error, and

T denotes the system operating time.

3.5. PPO Network Structure Design

To improve the learning capability of the reinforcement learning agent for complex industrial dynamic features, this study adopts an Actor–Critic structure to construct the PPO control network. The Actor network is responsible for generating the control policy, while the Critic network evaluates the value of the current policy. The system states are first input into a fully connected neural network, where nonlinear activation functions are employed to extract dynamic system features. Subsequently, the action probability distribution and state value function are output separately.

The Actor network policy output can be expressed as follows:

$$a_t \sim \pi_\theta(a_t \mid s_t) \tag{14}$$

where π_θ denotes the policy network parameterized by θ, s_t represents the system state, and a_t denotes the action generated by the intelligent agent.

The Critic network outputs the state value function as follows:

$$V_\phi(s_t) = E\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k}\right] \tag{15}$$

where $V_\phi(s_t)$ denotes the state value function, ϕ represents the value network parameters, γ denotes the discount factor, and R_{t+k} represents the future reward value.

4. Simulation Platform and Experimental Settings

4.1. MATLAB/Simulink Co-simulation Platform

This paper uses MATLAB R2024a and Simulink to jointly construct an experimental platform for induction heating temperature control. Simulink is used to establish the dynamic control model for induction heating, while the PPO reinforcement learning algorithm is implemented based on the MATLAB Reinforcement Learning Toolbox.

During system operation, the PPO reinforcement learning agent dynamically outputs PID parameter corrections based on the real-time system state and inputs the updated PID parameters into the controller. Subsequently, the controller generates control signals that act on the induction heating system, and the controlled object outputs a real-time temperature response. The system state and reward value are fed back through a temperature sensor, thus completing the reinforcement learning policy update. Through this co-simulation mechanism, real-time closed-loop interaction between the reinforcement learning controller and the industrial dynamic system can be achieved.

To more intuitively describe the co-simulation process established in this paper, its overall structure is shown in Figure 2.

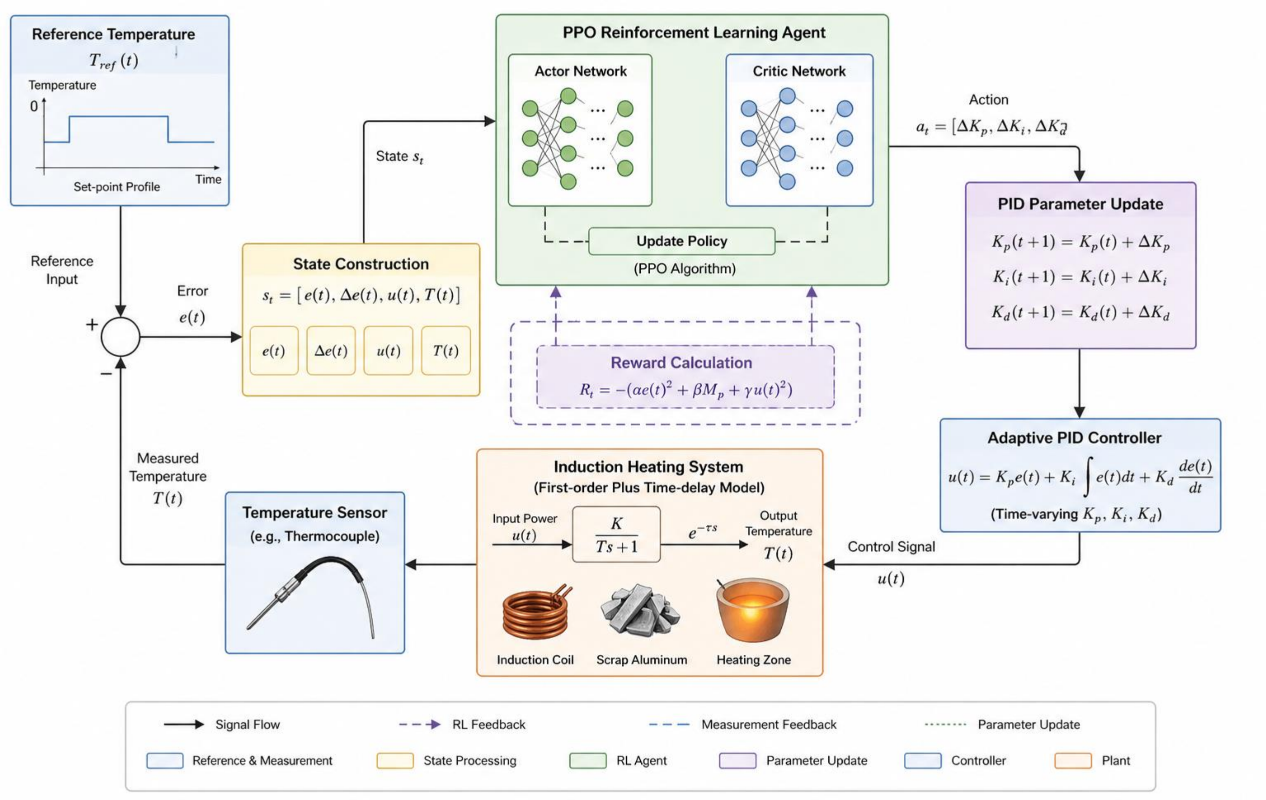

Figure 2: MATLAB/Simulink co-simulation framework of PPO-DRL-PID control system.

4.2. Experimental Parameter Settings

To ensure the stability of the reinforcement learning training process and the dynamic response performance of the control system, this paper standardizes the parameters of the PPO reinforcement learning network and the PID controller. The reinforcement learning part mainly includes parameters such as learning rate, batch size, discount factor, and the number of PPO iterations; the PID controller part includes initial proportional, integral, and derivative parameter settings. All experiments are conducted in the same simulation environment to ensure the fairness of the comparison results between different control algorithms.

During the reinforcement learning training process, the discount factor is used to control the importance of future rewards to the current policy update, and its expression is as follows:

$$G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k} \tag{16}$$

where G_t denotes the cumulative discounted reward, γ represents the discount factor, and R_{t+k} denotes the future reward value.

To improve the stability of advantage function estimation, the Generalized Advantage Estimation (GAE) method is adopted in this study, which can be expressed as follows:

$$\hat{A}_t = \sum_{l=0}^{\infty} (\gamma\lambda)^l \delta_{t+l} \tag{17}$$

where $\hat{A}_t$ denotes the estimated advantage function value, γ represents the discount factor, λ denotes the GAE decay coefficient, and δ_{t+l} denotes the temporal difference error.

Table 1: Experimental Parameter Settings of PPO-DRL-PID.

Parameter Category	Parameter Name	Parameter Value
Reinforcement Learning	Learning Rate	0.0003
	Batch Size	128
	PPO Epochs	10
	Discount Factor γ	0.99
	GAE Coefficient λ	0.95
	Clip Ratio ε	0.2
	Replay Horizon	2048
Network Structure	Hidden Layer Dimension	128
	Activation Function	ReLU
PID Parameters	Initial Kp	2.0
	Initial Ki	0.5
	Initial Kd	0.1
Simulation Parameters	Simulation Time	100 s
	Sampling Time	0.01 s

4.3. Comparative Control Algorithms

To verify the effectiveness and superiority of the proposed PPO-DRL-PID control method, this paper selects traditional PID controllers, fuzzy PID controllers, and SSA-optimized PID controllers as comparative algorithms, and conducts experimental analysis under the same simulation environment and parameter conditions.

Among them, Traditional PID adopts a fixed-parameter control method, which has a simple control structure but lacks online adaptive capability; Fuzzy-PID dynamically adjusts PID parameters through fuzzy rules, improving system robustness to a certain extent; SSA-PID uses a sparrow search algorithm to globally optimize PID parameters, improving the dynamic response performance of traditional PID controllers; the proposed PPO-DRL-PID method is based on deep reinforcement learning to achieve online dynamic updating of PID parameters, thereby further improving the system's control accuracy and disturbance rejection capability.

5. Results and Discussion

5.1. Reward Convergence Analysis

This paper statistically analyzes the convergence of the cumulative reward value with the number of training rounds during reinforcement learning training. The results are shown in Figure 3. In the figure, the horizontal axis represents the training round (Episode), and the vertical axis represents the cumulative reward value (Reward).

In the early stages of reinforcement learning training, because the agent has not yet formed a stable control strategy, the system output fluctuates significantly, resulting in a generally low and highly variable reward value. As the number of training rounds increases, the PPO agent gradually learns a more reasonable dynamic adjustment strategy for PID parameters, the system control error and overshoot gradually decrease, and the cumulative reward value continuously increases and gradually stabilizes.

The PPO reinforcement learning controller exhibits a relatively fast learning speed in the early stages. After approximately 200 training rounds, the reward value begins to increase significantly. As the number of training rounds further increases, the reward curve gradually smooths out and eventually converges stably. This indicates that the proposed PPO-DRL-PID control method can effectively learn the dynamic control law of the induction heating system and obtain a superior temperature control strategy.

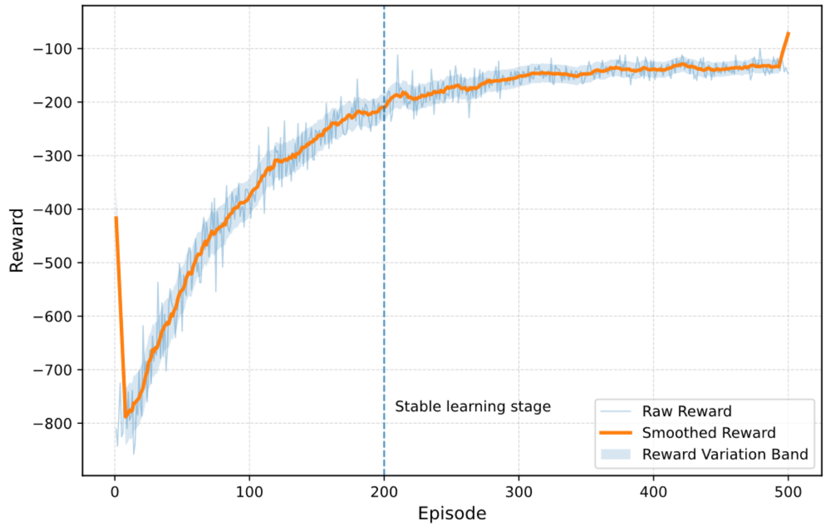

Figure 3: PPO reward convergence curve.

Furthermore, compared to traditional swarm intelligence optimization algorithms that can only perform offline parameter optimization, the PPO reinforcement learning method can continuously interact with the environment to update the policy online, thereby improving the system's adaptability and generalization performance under complex dynamic conditions. The control policy obtained after training can not only effectively reduce system temperature error, but also reduce parameter oscillations during the control process, improving the overall operational stability of the system.

5.2. Temperature Response Analysis

To further verify the dynamic response performance of the proposed control method in an induction heating system, temperature tracking experiments were conducted using three control methods: Traditional PID, SSA-PID, and PPO-DRL-PID. The system temperature response results are shown in Figure 4.

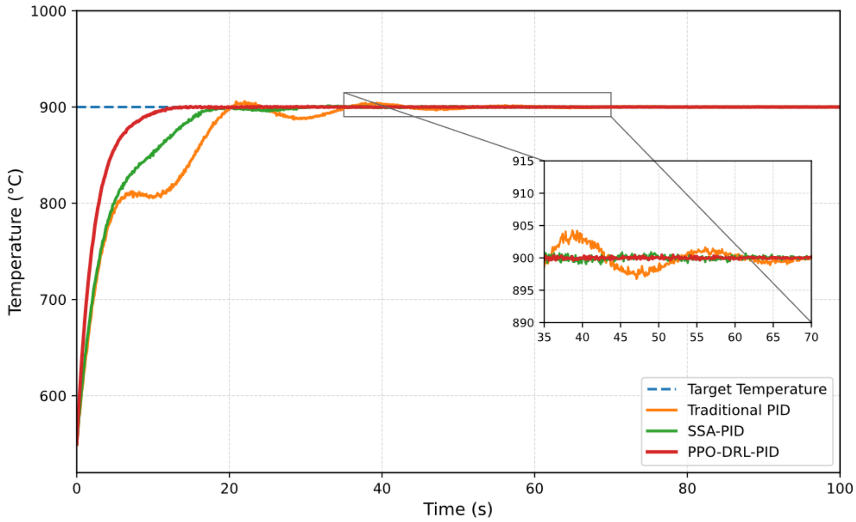

Figure 4: Temperature response curves of different control algorithms.

The Traditional PID controller, due to its fixed parameter control, exhibits significant overshoot during the system's heating phase and requires a long time to reach steady state. When the system is affected by thermal inertia and time delay, the traditional PID controller struggles to quickly adjust parameters, resulting in a slow system response and relatively large steady-state error.

In contrast, the SSA-PID controller uses a sparrow search algorithm to globally optimize the PID parameters, improving the system's dynamic response performance to some extent. Its temperature rise rate is significantly faster than that of Traditional PID, and the system overshoot is reduced. However, since SSA optimization is essentially an offline parameter optimization method, its control parameters lack real-time online adjustment capability when the system's dynamic characteristics change, thus still exhibiting some steady-state fluctuations.

The PPO-DRL-PID control method proposed in this paper exhibits superior control performance. During system heating, the PPO reinforcement learning agent dynamically adjusts the PID parameters based on the real-time system state, achieving a smoother and faster temperature tracking process. Compared to Traditional PID and SSA-PID, PPO-DRL-PID has smaller overshoot, shorter rise time, and lower steady-state error, enabling the system to reach the target temperature more quickly and maintain stable operation.

5.3. Disturbance Rejection Performance Analysis

To further verify the anti-disturbance capabilities of different control algorithms in complex industrial environments, this paper introduces an external disturbance signal at t=40 s after the system has stabilized. A comparative analysis of three control methods—Traditional PID, SSA-PID, and PPO-DRL-PID—is conducted, and the temperature response results are shown in Figure 5.

As can be seen from Figure 5, when the system is subjected to an external disturbance, the Traditional PID controller exhibits significant temperature fluctuations, with the system output deviating considerably from the target temperature, and requiring a long recovery time to reach steady state. This is mainly because the fixed-parameter PID controller lacks dynamic adaptive capability; when the system's dynamic characteristics change, its control parameters cannot be adjusted in a timely manner, thus easily generating large oscillations and steady-state deviations.

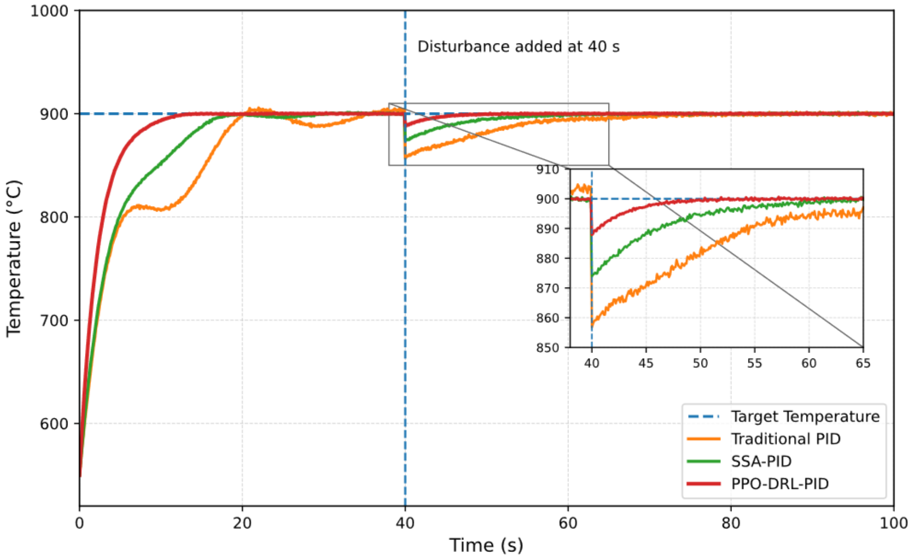

Figure 5: Disturbance rejection response curves of different control algorithms.

In contrast, the SSA-PID controller demonstrates a certain degree of robustness under disturbances, with its temperature fluctuation amplitude significantly smaller than that of the Traditional PID, and the system recovery speed is also improved. This indicates that the PID parameters optimized by the sparrow search algorithm can enhance the dynamic stability of the system to a certain extent. However, since SSA-PID essentially still uses offline parameter optimization, it still exhibits some control lag in complex dynamic environments.

The PPO-DRL-PID control method proposed in this paper demonstrates optimal disturbance rejection performance. When an external disturbance is introduced, the PPO reinforcement learning agent can quickly adjust the PID parameters based on the real-time system state, thereby effectively suppressing temperature fluctuations. Compared to Traditional PID and SSA-PID, PPO-DRL-PID has a smaller disturbance offset and a faster recovery speed, enabling the system to return to the target temperature and maintain stable operation in a shorter time.

5.4. Dynamic PID Parameter Analysis

To further investigate the online adaptive capability of the PPO reinforcement learning controller in the induction heating system, the dynamic variation of PID parameters during the control process is visualized and analyzed, as shown in Figure 6 it illustrates the dynamic variation curves of the proportional coefficient K_P, integral coefficient K_i, and derivative coefficient K_d during system operation.

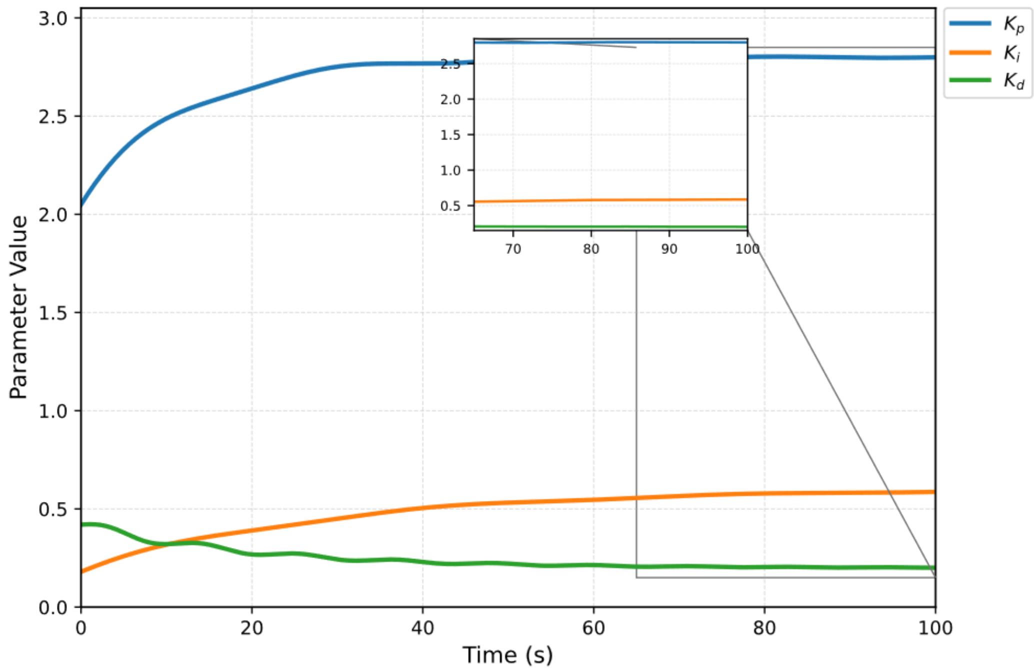

Figure 6: Dynamic variation curves of PID parameters in PPO-DRL-PID control.

As observed from Figure 6, during the initial stage of the system operation, the system temperature error is relatively large, and the PPO reinforcement learning agent rapidly adjusts the proportional parameter K_P to improve the response speed and accelerate the temperature rise process. As the system gradually approaches the target temperature, the variation of the proportional parameter becomes smoother, thereby avoiding excessive overshoot.

Meanwhile, the integral parameter K_i gradually increases after the system enters the steady-state stage, which helps further reduce the steady-state error and improve the temperature tracking accuracy. The derivative parameter K_d, on the other hand, changes significantly during the dynamic fluctuation stage of the system and is mainly used to suppress temperature oscillations and enhance system stability.

Compared with traditional fixed-parameter PID controllers, the proposed PPO-DRL-PID method can dynamically adjust control parameters according to the real-time operating state of the system, thereby realizing true online adaptive control. This dynamic parameter adjustment mechanism enables the controller to better adapt to the nonlinear and time-delay characteristics in complex industrial systems, thus effectively improving the dynamic response performance and robustness of the system.

5.5. Time-Domain Performance Comparison

To further quantify the dynamic control performance of different control algorithms, this paper comprehensively compares and analyzes three control methods—Traditional PID, SSA-PID, and PPO-DRL-PID—based on indicators such as system overshoot, rise time, settling time, and root mean square error (RMSE). The results are shown in Table 2.

Table 2 shows that the Traditional PID controller, due to its fixed-parameter control method, has the largest overshoot and the longest time required for the system to reach steady state. It also has a relatively high RMSE, indicating lower overall control accuracy. SSA-PID, through an intelligent optimization algorithm, globally optimizes the PID parameters, improving the system's dynamic response performance to some extent. Its overshoot and settling time are superior to Traditional PID.

Table 2: Comparison of time-domain performance of different control algorithms.

Method	Overshoot (%)	Rise Time (s)	Settling Time (s)	RMSE
Traditional PID	12.84	14.32	28.75	18.64
SSA-PID	6.27	10.48	18.36	9.12
PPO-DRL-PID	2.91	7.35	11.42	4.37

In contrast, the proposed PPO-DRL-PID control method achieves the best results across all evaluation indicators. Specifically, it further reduces system overshoot, significantly shortens rise time and settling time, and has the lowest RMSE, indicating that this method can achieve faster, more stable, and higher-precision temperature control.

6. Conclusion

To address the issues of nonlinearity, large time delay, and insufficient dynamic adjustment capability of traditional PID controllers in induction heating systems for scrap aluminum, this paper proposes a PPO-DRL-PID adaptive temperature control method based on deep reinforcement learning. This method combines the PPO reinforcement learning algorithm with the traditional PID control structure. Through real-time interaction between the reinforcement learning agent and the industrial environment, it achieves online dynamic optimization of PID parameters, thereby improving the system's dynamic response performance and disturbance rejection capability under complex operating conditions.

First, this paper establishes a first-order inertial pure time delay mathematical model of the induction heating system and constructs a co-simulation environment based on the MATLAB/Simulink platform. Based on this, a reinforcement learning state space, action space, and a reward function that comprehensively considers temperature error, overshoot, and control energy consumption are designed. Through reinforcement learning training, the PPO agent can gradually learn the dynamic control law of the system and form a stable and effective online PID parameter update strategy.

Experimental results show that compared with Traditional PID and SSA-PID control methods, the proposed PPO-DRL-PID method achieves superior results in terms of system overshoot, rise time, settling time, and RMSE. Meanwhile, under external disturbances, the PPO-DRL-PID controller exhibits stronger robustness and faster recovery capability, effectively suppressing system temperature fluctuations. Furthermore, the dynamic variation results of the PID parameters further validate the proposed method's excellent online adaptive control capability.

References

[1] Majumdar S. An insight view of evolution of advanced aluminum alloy for aerospace and automotive industry: current status and future prospects[J]. Journal of The Institution of Engineers (India): Series D, 2024: 1-18.

[2] Trzepieciński T, Najm S M. Current trends in metallic materials for body panels and structural members used in the automotive industry[J]. Materials, 2024, 17(3): 590.

[3] Samberger S. New strategies to improve Recycling and reduce CO2-emission of Aluminum production and processing[J]. Journal of light metal welding, 2024, 62(3): 119-132.

[4] Yang Y. Supply potential, carbon emission reduction, energy conservation, and sustainable pathways for aluminum recycling in China[J]. Sustainable Production and Consumption, 2024, 50: 239-252.

[5] Yakubov V. Recycled aluminium feedstock in metal additive manufacturing: A state of the art review[J]. Heliyon, 2024, 10(5).

[6] Esteve V, Bellido J L, Jordán J. State of the art and future trends in monitoring for industrial induction heating applications[J]. Electronics, 2024, 13(13): 2591.

[7] Chen X. Research status and applications of dual-frequency induction heating power supply[J]. Electronics, 2024, 13(24): 4913.

[8] Chanda S. Comparative analysis of proportional, proportional–integral, and proportional–integral–derivative controllers for thermal regulation in a cylindrical cooling system with multiple O-Ring heat sources[J]. Physics of Fluids, 2024, 36(8).

[9] Lakman N A, Mokhtar Z A. Mobile application for practical structural engineering learning: A USE questionnaire-based student evaluation on usefulness, satisfaction and ease of use[J]. International Journal of Business Studies and Education, 2024, 2(1).

[10] Emara R. Artificial Intelligence Based Controller for a Temperature Control System[J]. Faculty of Engineering, The British University in Egypt, 2024.

[11] Jabari M. An advanced PID tuning method for temperature control in electric furnaces using the artificial rabbits optimization algorithm: M. Jabari et al[J]. International Journal of Dynamics and Control, 2025, 13(5): 175.

[12] Abualigah L. Particle swarm optimization algorithm: review and applications[J]. Metaheuristic optimization algorithms, 2024: 1-14.

[13] Priyadarshi R, Kumar R R. Evolution of swarm intelligence: a systematic review of particle swarm and ant colony optimization approaches in modern research[J]. Archives of Computational Methods in Engineering, 2025, 32(6): 3609-3650.

[14] Lee S, Chau N K, Choi S. Deep Reinforcement Learning-Based Real-time Temperature Control in Thermoelectric Heat Exchange System[J]. Case Studies in Thermal Engineering, 2026: 107633.

[15] Gharbi A. Intelligent HVAC Control: Comparative Simulation of Reinforcement Learning and PID Strategies for Energy Efficiency and Comfort Optimization[J]. Mathematics, 2025, 13(14): 2311.