International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 2113-2127

https://doi.org/10.70517/ijhsa464181

Research on Adaptive Regulation System and Energy Consumption Optimization Control of Steelmaking Process Parameters in the Steel Industry Based on Reinforcement Learning

Wen Qiu[1,*], Zhenlong Zhao[2], Jian Huang[1], Zhangman Liu[1], Yanlong Zhang[1], Hongyan Luo[1], Weiyi Xing[1] and Baolong Li[1]

[1] Benxi Beiyingshan Iron and Steel (Group) Co., LTD. Steel Plant, Benxi, Liaoning, 117017, China
[2] Nanjing Dongchuang Xintong Internet of Things Research Institute Co., LTD., Nanjing, Jiangsu, 210000, China

Corresponding authors: (e-mail: 13941421170@163.com).

Abstract The traditional steelmaking process is difficult to meet the quality and efficiency of the current users of steel products, for the problem, the steelmaking process parameter adaptive adjustment system and energy consumption optimization control strategy based on the PPO algorithm for the iron and steel industry is proposed. First of all, the steelmaking process parameters are defined, and at the same time, the main problem of this research is determined, and the problem is transformed into a Markov decision-making process, and the PPO algorithm is used to optimize the steelmaking process parameters, and ultimately, the optimal adaptive regulation scheme of the steelmaking process parameters is generated. On this basis, with the help of microcontroller and programming technology, the design task of steelmaking process parameter adaptive adjustment system was completed. It is found that the optimal control of energy consumption of the steelmaking process parameters adaptive adjustment system belongs to multi-objective problem, the maximum completion time of the product and the total energy consumption as the objective function, in addition to the corresponding constraints are given, and the PPO algorithm is used to solve the objective function and get the optimal energy consumption control strategy. Integrate the above theories, the research program of this paper to carry out empirical investigation and analysis. Under the condition of 40mm scrap input, the PPO algorithm is more effective than the traditional PID algorithm in the adaptive adjustment of steelmaking process parameters, which confirms the effectiveness of adaptive adjustment of steelmaking process parameters in the steel industry based on reinforcement learning. In addition, under the same furnace size conditions, compared with the DDPG algorithm and SAC algorithm, the algorithm in this paper is easier to achieve the optimal control scheme of energy consumption, and its maximum completion time and total energy consumption values are 2.581s and 136.615KJ.

Index Terms PPO algorithm, steelmaking process parameters, adaptive regulation system, energy consumption

I. Introduction

In recent years, China's steel industry has faced intense pressure from the international market. To counter the threats posed by the international steel market to the domestic market, it is essential to continuously enhance steel production capacity and quality, accelerate technological upgrades, and align with international standards [1]-[3]. The steelmaking process system is a critical infrastructure for China's steel industry. Improving and optimizing this system, particularly through the adjustment and optimization of process parameters, holds profound significance for the development of China's steel industry [4], [5].

Process parameter adjustment and optimization play a crucial role in the steelmaking process of the steel industry [6]. By reasonably adjusting and optimizing process parameters, product quality and output can be improved, production costs reduced, and production efficiency enhanced [7]-[9]. Process parameters are key indicators in the steel production process, including temperature, pressure, speed, and concentration, which directly impact product quality and output [10], [11]. If process parameters are not accurately adjusted and optimized, it may lead to unstable product quality, failure to meet production targets, or even production accidents [12]-[14]. Therefore, the adjustment and optimization of process parameters are essential steps to ensure the smooth operation of the steelmaking process [15].

With the development of artificial intelligence, the application of optimization algorithms has provided technical support for achieving adaptive regulation and optimization of process parameters in steelmaking systems [16], [17].

Reinforcement learning, as a method of machine learning, enables the system to learn to make optimal choices through automatic trials by providing rewards or penalties, thereby optimizing steelmaking process parameters and energy consumption. The key distinction from other machine learning methods is that it does not require training data but instead continuously interacts with the environment to learn, thereby automatically optimizing the steelmaking process [18]-[21].

Based on the knowledge of steelmaking process and reinforcement learning theory, this paper transforms the problem of adaptive regulation of steelmaking process parameters in the iron and steel industry into a Markov decision-making process, which includes state space, action space, reward function, and designs an adaptive regulation system of steelmaking process parameters with the technical support of single-chip microcomputer and programming software. Combined with the relevant literature and data, it can be seen that the optimal control of energy consumption of the steelmaking process parameters adaptive adjustment system belongs to the multi-objective problem, in this regard, the maximum completion time of the product and the total energy consumption are set as the objective function, and the corresponding constraints are determined, and the PPO algorithm is used for the training and solving of the objective function to realize the optimal control of energy consumption. To synthesize the above, the adaptive adjustment system of steelmaking process parameters based on PPO algorithm and the optimal control of energy consumption are verified and analyzed.

II. Reinforcement Learning Based Exploration of Steelmaking Processes in Iron and Steel Industry

II. A.Steelmaking process

Steelmaking is the process of blowing molten iron through a converter to adjust its chemical composition, reduce carbon content and produce high quality steel. The raw materials required are mainly scrap, molten iron and rare gases. Firstly, the raw materials such as scrap and molten iron from the blast furnace are put into the converter for smelting, removing harmful gases (carbon, phosphorus, sulfur, hydrogen, etc.) and adjusting the content of the key elements to a specific range, raising the temperature of the molten steel until it reaches the temperature of the steel to be made, so as to obtain the molten steel, which is then poured into the ladle that has been prepared beforehand, and then the ladle will be sent to the corresponding refining equipment for refining by means of a traveling car. The steel is poured into the prepared ladle, which is transported to the corresponding refining equipment by means of a traveling car.

II. B.Enhanced learning

II. B. 1) Principles of Reinforcement Learning

Reinforcement learning is regarded as an important machine learning technique, whose core problem is how an intelligent body goes about maximizing the rewards it can obtain in a complex and uncertain environment [22], [23]. Reinforcement learning's unique ability to learn on its own enables it to learn from scratch in a manner similar to the human cognitive and trial-and-error process. Reinforcement learning does not need to know all the details of the problem in advance (i.e., the state transfer matrix and the payoff function are unknown), nor does it need to exhaust all the possible states, which makes reinforcement learning able to cope with complex sequential decision problems. In reinforcement learning algorithms, the core is mainly focused on how the intelligent body interacts with the environment. The intelligent body learns by interacting with the environment and flexibly adjusts its behavioral strategies based on immediate feedback to further optimize the results of adaptive regulation.

II. B. 2) Markov decision-making process

Markov Decision Process (MDP) is the model underlying reinforcement learning [24]. For the interaction process between the intelligence and the environment in reinforcement learning, researchers usually use MDP to model it. Its basic mathematical framework mainly consists of five parts, which can be represented as a quintuple $< S, A, P, R, \gamma >$:

(1) S is the set of environment states, $S = [s_1, s_2, ..., s_i]$, where s_i denotes the environment state at time step t.

(2) A is the set of intelligent body actions, $A = [a_1, a_2, ..., a_i]$.

(3) P is the probability of state transfer, i.e., the probability that the current state $s_i \in S$ is transformed to state $s_t + 1 \in S$ by action $a_i \in A$ at time step t. It is specifically denoted as:

$$P(s_{t+1} \mid s_t) = P(s_{t+1} \mid s_1, s_2, ..., s_t) \tag{1}$$

(4) R is the immediate reward function, which indicates the reward obtained by the intelligent body for taking an action in the current moment state.

(5) γ is the discount factor and $\gamma \in [0,1]$. The closer its value is to 0 indicates that the strategy values current rewards more and the closer its value is to 1 values future rewards more.

The Markov decision process is shown in Fig. 1. In the whole dynamic process, taking state s_0 as the starting point, after observing the environment the intelligent body selects action a_0 according to strategy π, executes action a_0 and interacts with the environment according to the state transfer probability p, and then transfers to the next state s_i while receiving an immediate reward from the environment r_1. After many interactions, a sequence of interaction processes is formed, i.e., a sequence consisting of states, actions, and rewards: $(s_0, a_0, r_1, s_1, a_1, r_2, ...)$. Extracting the reward value for each reward value of each interaction, multiply and weight it with the discount factor to get the following cumulative reward, as shown in Equation (2):

$$G_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + ... = \sum_{k=0}^{N} \gamma^k r_{t+k+1} \tag{2}$$

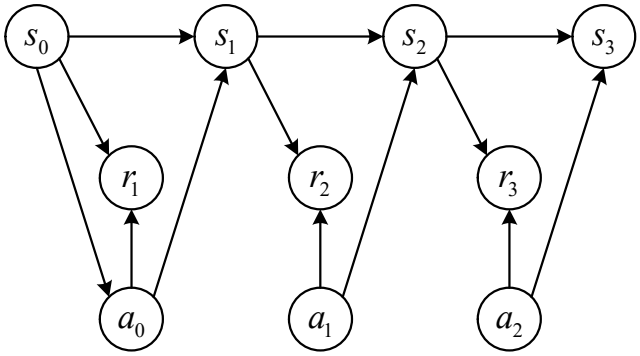


Figure 1: Markov Decision Process

The problem to be solved by this decision-making process is to find an optimal strategy $\pi(s): s \to a$ in order to maximize the cumulative reward G_i. However, throughout the interaction process, a large number of strategy trajectories are generated when the intelligent body explores the optimal strategy, and the existence of diversity in strategy trajectories also makes the value of the cumulative reward function complex and difficult to compute, in order to measure the goodness of a certain state or a certain action under the strategy π, it is proposed to compute the cumulative reward function with mathematical expectation. Equation (3) and equation (4) are known as Bellman's equation. For:

$$\begin{aligned} V^{\pi}(s) &= E^{\pi}\left[\sum_{k=0}^{N} \gamma^k r_{t+k+1} \mid s_t = s\right] \\ &= E^{\pi}[r_{t+1} + \gamma v^{\pi}(s_{t+1}) \mid s_t = s] \end{aligned} \tag{3}$$

$$\begin{aligned} Q^{\pi}(s,a) &= E^{\pi}\left[\sum_{k=0}^{N} \gamma^k r_{t+k+1} \mid s_t = s, a_t = a\right] \\ &= E^{\pi}[r_{t+1} + \gamma Q^{\pi}(s_{t+1}, a_{t+1}) \mid s_t = s, a_t = a] \end{aligned} \tag{4}$$

The mathematical expectation of the cumulative reward function brought about by an intelligent body from state s using under strategy π is denoted as the state-value function $V^{\pi}(s)$, $V^{\pi}(s)$ only with respect to strategy π. Whereas strategy π is composed of a series of actions a, adding the actions to the expectation yields the formulas, i.e., the state-action value function $Q^{\pi}(s,a)$, $Q^{\pi}(s,a)$ denotes the expectation of the long term cumulative rewards for executing in accordance with strategy π after deterministically choosing action a while in state s. Obviously, the following relationship exists between $V^{\pi}(s)$ and $Q^{\pi}(s,a)$:

$$V^{\pi}(s) = \sum_{a \in A} \pi(a \mid s) Q^{\pi}(s,a) \tag{5}$$

$$Q^{\pi}(s,a) = r_{t+1} + \gamma \sum_{s'} p_{ss'}^{a} V^{\pi}(s') \tag{6}$$

Further, maximizing the cumulative reward function yields the optimal policy π^*, which is formulated as shown in Eqs. (7) and (8) and is known as the Bellman optimality equation. The objective of π^* is equivalent to maximizing the state-value function or state-action-value function. For:

$$v^*(s) = \max_{\pi} v^{\pi}(s) \tag{7}$$

$$Q^*(s,a) = \max_{\pi} Q^{\pi}(s,a) \tag{8}$$

Reinforcement learning seeks the optimal value function and hence the optimal policy by optimizing the Bellman equation to find the optimal value function.

II. B. 3) Optimization Algorithm for Near-End Policies

The Proximal Policy Optimization (PPO) algorithm is currently a very popular deep reinforcement learning method and has been adopted by OpenAI as the default algorithm for the reinforcement learning module due to its ease of use and good performance.The PPO algorithm uses Actor-Critic as the basic framework.The traditional Actor-Critic method is online training, but during the training process, each sample data is only trained once, which leads to low data utilization. In order to improve the data utilization, PPO improves the training method by setting up a network of two Actor strategies, old and new, and introducing importance sampling technique. During the training process, the new strategy is constantly updated using the strategy gradient, while the old strategy is responsible for recording the network parameters at the time of data generation, which makes it unnecessary for the optimization of the strategy and the training process to be carried out at the same time. The PPO algorithm is used to solve the problem of adaptive regulation system of steelmaking process parameters and optimal control of energy consumption in the iron and steel industry, and it is necessary to analyze and design each part of the PPO algorithm, such as the state space, the action space, and the reward function, according to the characteristics of the dynamic scheduling problem and optimization objectives, and then complete the solution of the final problem through the training of the algorithm.

II. C.Adaptive adjustment system for steelmaking process parameters

II. C. 1) Parameter definitions

S : Production stage pooling, $S = \{1,2,...,s...,|S|\}$.

J : Furnace stage collection, $J = \{1,2,...,j...,|J|\}$.

G : casting collection, $G = \{1,2,...,g...,|G|\}$.

N_g : the set of furnaces contained in casting g .

L_{gj} : the j th furnace of casting g , $L_{g,1}$ represents the first furnace of casting g , $L_{g,j}$ represents the last furnace of casting g .

$w_{j,s,s+1}$: Waiting time between adjacent processes in the furnace.

t_{pre} : Preparation time between neighboring pours.

t_s : Transportation time between adjacent furnaces in production stages s and $s+1$.

$p_{j,s}$: Processing time of furnace j in production stage s .

$p_{j,s}^{L}$: Minimum processing time of furnace j in production stage s .

$p_{j,s}^{U}$: Maximum processing time of furnace j in production stage s .

M_s : the set of machines in production phase s , $M_s = \{1,2,...,m,....,|M_s|\}$ where $|M_s|$ is the total number of machines included in production phase s and M is the total number of machines.

$c_{j,s}$: the moment of completion of the furnace j in production phase s .

U : a sufficiently large positive number.

Decision variable definition:

$s_{j,s}$: The moment of the start of processing of furnace j in production phase s .

$x_{j,s,m}$: 0-1 variable equal to 1 when furnace order j is scheduled to be processed on machine m during production phase s , and 0 otherwise.

$y_{j_1,j_2,s,m}$: 0-1 variable equal to 1 when furnace number j_1, j_2 is scheduled to be processed on the same machine m during production phase s and furnace number j_1 is processed before furnace number j_2 , 0 otherwise.

II. C. 2) Related issues

(1) Optimization goals:

$$\min f = \sum_{j=1}^{|J|} (s_{j,|s|} - c_{j,1}) \tag{9}$$

(2) Constraints:

$$\sum_{m=1}^{|M_j|} x_{j,s,m} = 1, \forall j \in J, s \in S \tag{10}$$

$$\sum_{j=1}^{|J|} x_{j,s,m} = 1, \forall m \in M_s, s \in S \tag{11}$$

$$s_{j,s+1} \geq s_{j,s} + p_{j,s} + w_{j,s,s+1} + t_s, \forall j \in J, 1 \leq s \leq |S| - 1 \tag{12}$$

$$y_{j_1,j_2,s,m} + y_{j_2,j_1,s,m} = 1, \forall j_1 \neq j_2 \in J, s \in S, m \in M_s \tag{13}$$

$$\begin{gathered} s_{j_2,s} - s_{j_1,s} - p_{j_1,s} - U \cdot (3 - x_{j_1,s,m} - x_{j_2,s,m} - y_{j_1,j_2,s,m}) \geq 0 \\ \forall j_1 \neq j_2 \in J, m \in M_s, 1 \leq s \leq |S| - 1 \end{gathered} \tag{14}$$

$$x_{j_1,s,m} = x_{j_2,s,m}, \forall j_1, j_2 \in N_g, g \in G, m \in M_s, s = |S| \tag{15}$$

$$s_{j+1,s} - s_{j,s} - p_{j,s} = 0, \forall j, j+1 \in N_g, g \in G, s = |S| \tag{16}$$

$$\begin{gathered} s_{L_{\theta_2,1},s} - s_{L_{\theta_1,10},s} - p_{L_{\theta_1,10},s} \geq t_{pre} + U \cdot (3 - x_{L_{\theta_2,1},s,m} - x_{L_{\theta_1,10},s,m} - y_{L_{\theta_1,10},L_{\theta_2,1},s,m}) \\ \forall m \in M_s, g_1 \neq g_2 \in G, s = |S| \end{gathered} \tag{17}$$

$$c_{j,s} = s_{j,s} + p_{j,s}, \forall j \in J, s \in S \tag{18}$$

$$p_{j,s}^L \leq p_{j,s} \leq p_{j,s}^v, \forall j \in J, s \in S \tag{19}$$

$$x_{j,s,m} \in \{0,1\}, \forall j \in J, s \in S, m \in M_s \tag{20}$$

$$y_{j_1,j_2,s,m} \in \{0,1\}, \forall j_1, j_2 \in J, s \in S, m \in M_s \tag{21}$$

Eq. (9) is the objective function, which represents the minimization of the total sojourn time of the furnace. Eqs. (10)-(11) are allocation constraints indicating that furnace times can only be allocated to one machine in each production phase and each machine can only process one furnace time at the same moment, respectively. Eqs. (12)-(13) represent the priority relationship constraints, and Eq. (14) indicates that each furnace finish the previous process before the next process can be processed. Equation (15) indicates that there is a priority relationship between any two different furnaces processed on the same machine. Equation (16) indicates that at all stages (excluding the continuous casting stage), when two furnaces are scheduled to be processed on the same machine, the machine can process the following furnace only if it finishes processing the previous furnace. Formula (17)-(18) for continuous casting constraints, formula (17) that in the continuous casting stage, the same casting all furnaces within the same pouring to be arranged in the same continuous casting machine processing. Equation (18) indicates that in the continuous casting stage, the furnaces belonging to one casting need to be processed closely on the same continuous casting machine. Equation (19) indicates that there is a preparation time when two different pours are processed before and after the continuous casting machine. Equation (20) indicates the completion time of the furnace at each stage. (21) indicates that the processing time is constrained within a certain range of decision variables.

II. C. 3) Problem Transformation

In this paper, the proximal policy optimization algorithm implements the problem of adaptive regulation of steelmaking process parameters, and it is necessary to convert the problem into a Markov decision process. In this

process, the intelligent body selects the appropriate furnace to be processed for each equipment according to the state space, i.e., the current environmental information of the workshop, and the intelligent body is able to obtain the instant reward after scheduling the furnace processing. The reinforcement learning algorithm solves the steelmaking-continuous casting scheduling problem according to Markov theory in two stages: pre-training and online optimization. In the training phase, when orders are released to the shop floor, the intelligent body continuously interacts with the environment to collect historical experience. Next, the intelligent body uses the DRL method and historical decision-making experience in order to optimize the steelmaking process parameters. Finally, the intelligence learns to optimize the target by selecting appropriate scheduling rules based on the real-time state of the shop floor. The online optimization phase can use the trained model to make fast decisions on new orders and generate a better adaptive regulation of steelmaking process parameters.

II. C. 4) State space

Intelligent bodies select actions based on system state information, and changes in state information will affect the action strategies of intelligent bodies. Therefore, the ideal state information should contain information related to the characteristics of the research problem and be able to reflect the changes in the production environment in real time. Processing the data related to decision-making in the production process and integrating them into the state space can improve the decision-making ability of the intelligent body. Then, the information related to the machines and furnaces in the workshop is analyzed and quantitatively described as the state input to the reinforcement learning model, and the state space of the scheduling problem is defined as $F=(f_1,f_2,f_3,f_4)$, as shown below.

(1) State feature $f_i(|J|\times 1)$: the current total sojourn time matrix of each furnace, which describes the sojourn time information of each furnace up to the current decision point.

(2) State feature $f_2(|J|\times|S|)$: Workpiece start time matrix at each production stage, which records the start time of each furnace at each production stage.

(3) State feature $f_3(|J|\times|S|)$: Workpiece end time matrix for each production phase, which records the end time of each furnace in each production phase.

(4) State characteristic $f_4(M\times 1)$: Machine state matrix, defined as follows:

$$f_{m,4}=\begin{cases}1, & \text{It indicates that machine } m \text{ is currently busy}\\ 0, & \text{It indicates that machine } m \text{ is currently idle}\end{cases} \quad m=1,2,...,M \tag{22}$$

II. C. 5) Action space

In the steelmaking process parameter adaptive regulation scheduling problem, action means selecting the appropriate processing furnace for each idle machine according to the current system state. While heuristic rules can achieve prioritization for each task according to the given rules, which is conducive to learning intelligences to communicate and interact with the environment using prior knowledge. Therefore, the action space in this paper consists of several different heuristic scheduling rules.

II. C. 6) Reward functions

The optimization objective is to minimize the total sojourn time of the furnaces, but not all furnaces are scheduled at one time during the scheduling process and the final total sojourn time cannot be obtained, so the effect of the decision point scheduling strategy on the objective must be considered. In this paper, we set d_t to be the total sojourn time at the t decision moment, and the reward function is set to be the opposite of the increment of the total sojourn time of the furnaces in the t decision point system, as shown in the following equation:

$$r(s,,a,)=(d,-d,)/100 \tag{23}$$

where d_t represents the total oven stay time of the current decision point system and d_{t-1} represents the total oven stay time of the previous decision point system. It can be shown that maximizing the cumulative reward is equivalent to minimizing the total sojourn time. For:

$$\begin{aligned}R&=\sum_{i=1}^{T}r(s_i,a_i)=\frac{1}{100}\sum_{i=1}^{T}(d_{i-1}-d_i)\\&=\frac{1}{100}(d_0-d_1+d_1-d_2+...+d_{T-1}-d_T)\\&=\frac{1}{100}(d_0-d_T)=-d_T/100\end{aligned} \tag{24}$$

It can be seen that the cumulative reward is inversely proportional to the total sojourn time, which can be achieved by maximizing the cumulative reward to minimize the optimization objective of the original steelmaking-continuous casting scheduling problem.

II. C. 7) System implementation

The steelmaking process parameter adaptive adjustment system is divided into two parts, the upper computer and the lower computer, the upper computer adopts ordinary PC, while the lower computer adopts MCS-51 microcontroller for the expansion of the interface circuit. The lower computer adopts MCS-51 as the total control chip, which can obtain the real-time conversion of the last moment of stay through the ADC0809, compare it with the pre-set stay time, get the difference, and then according to the PPO algorithm to minimize the total stay time, and the control volume output is handed over to the 74LS138 to execute the control of the switching volume of the intelligent body and other switching volume. It can also control the In-tel8279 chip to display the on-site process parameters of steelmaking and display them on the seven-segment digital tube. Finally, the total control chip sends some steelmaking process parameters (e.g., rise time, overshooting amount, adjustment time) to the PC through the MAXM232 chip, and the upper computer dynamically displays the parameters transmitted from the lower computer as well as the whole process flow on the screen. The upper computer of the system uses the technology of WPF to write the program, WPF is a new generation of Microsoft graphic system, which can run in the window XP environment. The lower computer software is developed using LCAET51 microcontroller language, the lower computer software adopts modular design, which consists of the main control program and a number of subroutine modules.

II. D. Optimized control of energy consumption

II. D. 1) Description of multi-objective optimization

A multi-objective optimization problem (MOP) is defined as an optimization problem that solves for the minimum of multiple conflicting objective functions.The formal description of MOP is as follows:

$$Minf(x) = \{f_1(x), f_2(x), ..., f_q(x)\}, x \in X \tag{25}$$

where $f_1, f_2, ..., f_q$ is the objective function for q conflicts and x is the decision variable in the decision space x.

For different feasible solutions a and b in the MOP, a feasible solution a is defined as a dominated solution (denoted as $a < b$) of a feasible solution b, assuming that any objective function $\forall Q \in \{1, 2, ..., q\}$ satisfies $f_Q(a) \le f_Q(b)$ and there exists an objective function $\exists Q \in \{1, 2, ..., q\}$ satisfying $f_Q^{'}(a) < f_{Q'}(b)$. A feasible solution that is not dominated by any feasible solution is defined as an undominated solution.The Pareto optimal set consists of all the undominated solutions, and the projection of the Pareto optimal set in the objective space forms the optimal Pareto front. In the EDNWFSP-SDST problem, the total energy consumption TEC of workpiece machining is also considered on the basis of the maximum completion time . $C_{\max}$. studied in the traditional scheduling problem. The advantages and disadvantages of both $C_{\max}$ and TEC optimization objectives are related to the machining speed of the workpiece, and the change of the machining speed produces a change of a different nature in the $C_{\max}$ and the TEC, therefore, considering the optimal control of energy consumption for the optimization objectives of the $C_{\max}$ and the TEC The problem is MOP with 2 conflicting objective functions.

II. D. 2) Problem description

In the energy optimization control problem, n workpieces to be machined are assigned to f plants, where each workpiece to be machined can only be machined in the assigned plant. Each plant contains isomorphic flow shops with m machines in each flow shop. Each part is processed in the same order on m machines in different plants. There is a zero-waiting constraint in the optimal control of energy consumption, where each workpiece to be processed has to be continuously processed between machines, and there should not be a situation where the workpiece is processed by the current machine and still has to wait for the next machine to finish processing it. The problem takes into account the preparation time associated with the sequence, i.e., the workpieces need to spend a certain amount of time for preparation before they are formally processed on each machine. Due to the differences in the processes, the preparation time required for the machining of different workpieces is not the same for different workpieces located after the machining. In this problem, two optimization objectives, maximum completion time $C_{\max}$ and total energy consumption TEC, are considered simultaneously.

Assume that each workpiece is machined at a different speed on the machine and that the machining speed remains constant until machining is completed on the same machine. Define the standard machining time of part j on machine i to be $p_{i,j}$ and the machining speed to be $v_{i,j}$, so the actual machining time of this part is $p_{i,j}/v_{i,j}$. The energy consumption per unit of time is related to the machining speed, with faster machining speeds implying shorter machining times and higher machining energy consumption. The problem needs to optimize C_{max} and TEC at the same time to obtain the optimal Pareto solution set, and the machining speed of the workpiece and the change of the workpiece arrangement order have conflicting effects on the objective function, so how to design a reasonable algorithm to optimize the conflicting objectives is the focus of this study.

II. D. 3) Constraints and objectives

According to the optimal control problem of energy consumption in the adaptive regulation system of steelmaking process parameters in the iron and steel industry, a multi-objective optimization mathematical model is constructed. The corresponding parameters are as follows:

j - Index of the workpiece to be machined.

k - Position of the workpiece to be machined in the machining process.

i - Index of machining machines.

l - Index of the machining plant.

c - Machining speed level.

n - Number of workpieces to be machined.

m - Number of machining machines.

f - Number of processing plants.

π - Workpiece processing process.

PI_i - Idle energy consumption per unit time on processing machine i.

SP_i - Preparation energy consumption per unit time on processing machine i.

$PP_{i,c}$ - Processing energy consumption per unit of time when the processing speed level on the processing machine i is c.

$v_{i,j}$ - Speed of machining workpiece j on machining machine i.

$O_{i,j}$ - Operation of machining workpiece j on machine i.

$p_{i,j}$ - Standard machining time of workpiece j on machine i.

$ST_{i,j,j}$ - Preparation time between adjacent machined workpieces j and j' on machine i.

$X_{j,k,l}$ - Binary variable, variable value 1 when workpiece j is assigned to position k on plant l, 0 otherwise.

$Y_{j,i,c}$ - Binary variable with a variable value of 1 when workpiece j is machined at speed grade c on machine i and 0 otherwise.

$tp_{i,j,l}$ - Actual machining time when workpiece i is machined on machine i in factory l.

IEC - Total energy consumption during idle time.

SEC - Total energy consumption for preparation time.

PEC - Total energy consumption for processing time.

TEC - Total energy consumption of the workpiece during machining.

$C_{k,i,l}$ - Machining completion time of the workpiece at the k th position of the i th machine in the l th plant.

$C_{\max}$ - Maximum completion time of all workpieces.

The MILP model for the distributed zero-waiting sequence-dependent flow shop scheduling problem under energy constraints with respect to minimizing C_{max} and TEC is shown in (26) to (42). where (26) is the optimization objective of the model. Constraints (27) and (28) indicate that in all factories, only one machining location can be assigned to each workpiece and only one machining workpiece can be assigned to each machining location. Constraint (28) guarantees a specific processing speed for each job on each machine. Constraint (29) denotes how the completion time of the first position on the first machine in each plant is calculated. Constraint (30) guarantees that the start time of the later processed workpiece on the same machine is later than the completion time of the previous processed workpiece. Constraints (31) to (32) define zero-wait constraints that guarantee that there is no waiting time between two neighboring processes for the same workpiece. Constraint (32) defines that the machining time cannot be negative. Constraint (34) defines the C_{max} optimization objective. Constraints (35) and (36) are the machining energy consumption under idle time. The preparation time energy consumption calculation is

represented by Eq. (37) and the processing time energy consumption calculation is defined as Eq. (38) and Eq. (39). Equation (40) defines the TEC. constraints (41) and (42) are logical representations of binary decision variables.

$$Minimize(C_{\max}, TEC) \tag{26}$$

$$\sum_{l=1}^{f}\sum_{k=1}^{n} X_{j,k,l} = 1, \forall j \tag{27}$$

$$\sum_{l=1}^{f}\sum_{j=1}^{n} X_{j,k,l} = 1, \forall k \tag{28}$$

$$\sum_{c=1}^{e} Y_{j,i,c} = 1, \forall j, \forall i \tag{29}$$

$$C_{1,1,l} = \sum_{j=1}^{n}\sum_{c=1}^{e} Y_{j,1,c} \cdot X_{j,1,l} \cdot (p_{1,j} / v_{1,j} + ST_{1,\pi_0,\pi_1}) = 1, \forall l \tag{30}$$

$$C_{k+1,i,l} \geq C_{k,i,l} + \sum_{j=1}^{n}\sum_{c=1}^{e} Y_{j,i,c} \cdot X_{j,k+1,l} \cdot (p_{i,j} / v_{i,j} + ST_{i+1,\pi_k,\pi_{k+1}}) \quad k = 1,2...n-1, \forall i, l \tag{31}$$

$$C_{k,i+1,l} = C_{k,i,l} + \sum_{c=1}^{e} Y_{j,i+1,c} \cdot X_{j,k,l} \cdot (p_{i+1,j} / v_{i+1,j} + ST_{i+1,\pi_{k-1},\pi_k}) \quad i = 1,2...m-1, \forall k, l \tag{32}$$

$$C_{k,i,l} \geq 0 \quad \forall k, i, l \tag{33}$$

$$C_{\max} = \max C_{n,m,l} \quad \forall l \tag{34}$$

$$IEC = \sum_{l=1}^{f}\sum_{i=1}^{m}\sum_{j=1}^{n} P_l \cdot ti_{j,i,l} \tag{35}$$

$$ti_{i,j,l} = \sum_{j=1}^{n-1}\sum_{c=1}^{e} Y_{j,i,c} \cdot X_{j,k+1,l} \cdot (C_{k+1,i,l} - C_{k,i,l} - p_{i,j} / v_{i,j} - ST_{i,\pi_k,\pi_{k+1}}) \tag{36}$$

$$SEC = \sum_{l=1}^{f}\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{n-1} PS_i \cdot X_{j,k,l} \cdot ST_{i,\pi_k,\pi_{k+1}} \tag{37}$$

$$PEC = \sum_{l=1}^{f}\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{c=1}^{e} PP_{i,c} \cdot Y_{j,i,c} \cdot tp_{i,j,l} \tag{38}$$

$$tp_{j,i,l} = \sum_{l=1}^{f}\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{c=1}^{e} X_{j,k,l} \cdot Y_{j,i,c} \cdot p_{i,j} / v_{i,j} \tag{39}$$

$$TEC = IEC + SEC + PEC \tag{40}$$

$$X_{j,k,l} = \{0,1\}, \forall j, k, l \tag{41}$$

$$Y_{j,i,c} = \{0,1\}, \forall j, i, c \tag{42}$$

II. D. 4) Optimized control strategy based on PPO algorithm

The PPO intelligent body-environment interaction part and the PPO intelligent body training part. During the interaction between the intelligent body and the environment, the environment is defined as a set of job lists with specified devices and processing times as well as corresponding transport times, and in the energy-optimized control environment, the intelligent body randomly takes actions according to the probability distribution in the action space. This is because it is not certain which action is good or bad until the end of training. The stochastic strategy is conducive to fully exploring the action space of the environment and speeding up the convergence of the training decision model. The probability distribution of the action space is obtained according to the actual steelmaking-continuous casting production state, and the action with the highest probability is selected to guide the equipment to perform the production task, i.e., selecting the appropriate furnace and processable equipment, so that the furnaces can be processed in a continuous and orderly manner, and shortening the energy consumption is the main goal in the whole process. After the execution of the action, it receives the reward function given by the environment and the next state of the environment, and schedules the intelligent body to enter a new scheduling time point. In the training part of the intelligent body, for the beginning of the training the intelligent body interacts with the simulated environment in a trial-and-error manner, and continuously trains and optimizes the energy consumption control based on the obtained decision results and rewards. Once the scheduling intelligence learns the optimal energy consumption strategy, it can be applied to the steelmaking-continuous casting production environment for real-time dynamic energy consumption optimization control.

III. Analysis of empirical findings

III. A. Experimental analysis of adaptive regulation system

In the experimental analysis of the steelmaking process adaptive regulation system in the steel industry, the reinforcement learning algorithm in this paper is used to compare with the traditional PID algorithm to verify the advantages of the reinforcement learning algorithm. The traditional PID parameters are measured by experienced experimental personnel through the experimental method, and the reinforcement learning algorithm is deployed in the experimental platform after being stabilized by the real environment training. The practical value of the reinforcement learning-based adaptive adjustment system for steelmaking process parameters in the steel industry is verified in terms of both stability and robustness.

III. A. 1) Stability analysis

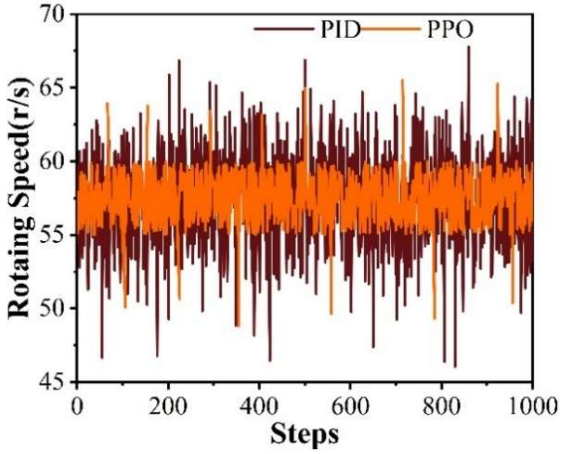
In order to quantitatively evaluate the adaptive regulation system, the time domain performance indicators and comprehensive performance indicators are used to evaluate the control system According to the experience of evaluating the discrete control system, the standard deviation (SD) of the actual value of the output rotational speed is used to measure the fluctuation of the data, and the smaller the standard deviation means that the fluctuation of the system is smaller, and the more stable the working condition is, and the calculation formula is as follows:

$$SD = \sqrt{\frac{\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2}{N-1}} \tag{43}$$

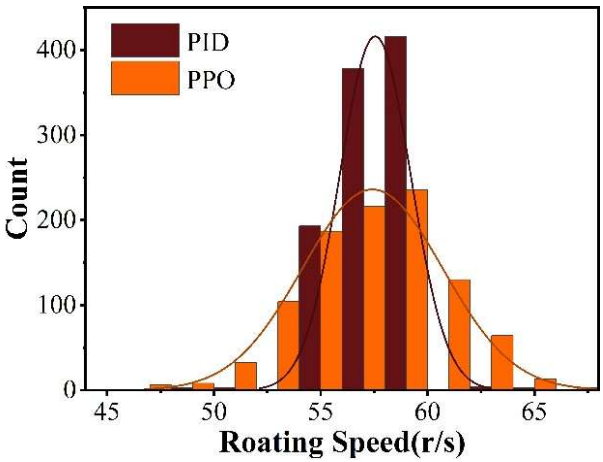
In order to measure the control effect of the reinforcement algorithm, three parameters of the steelmaking process in the steel industry, i.e., rise time, overshooting amount, and regulation time, are used to evaluate the performance of the algorithm.

The experiment is divided into two groups to carry out, respectively, into the 40mm scrap steel and 80mm scrap steel, set the motor speed of 80 revolutions per second, simulate the system load is small and the load is large, the microcontroller to collect the speed of the time period of 20ms, selected 1000 data for processing, in this 1000 cycle of the steelmaking equipment work time is 12s, recorded the motor speed changes with time, the Draw the speed curve graph and speed distribution graph. Calculate the rise time, overshoot, and regulation time of different algorithms, and draw a table for comparison, and the stability analysis results are shown in Fig. 2, where (a) to (d) are the light-load speed curve, light-load speed distribution, heavy-load speed curve, and heavy-load speed distribution, respectively. Figures (a) and (c) illustrate that the suppression effect of the PPO reinforcement learning algorithm on speed fluctuation is significantly improved compared with the traditional PID algorithm. After the load is increased, the traditional PID algorithm is not enough to suppress the huge excitation generated by the load, and the rotational speed generates huge fluctuations after the load is increased, but the distribution of the PPO algorithm is closer to the target value than that of the traditional PID algorithm, which is more similar to that of the performance under the light-loaded condition. Figures (b) and (d) can more intuitively see the distribution of the current speed and the target speed, the PPO algorithm shows more concentrated and stable characteristics in the speed

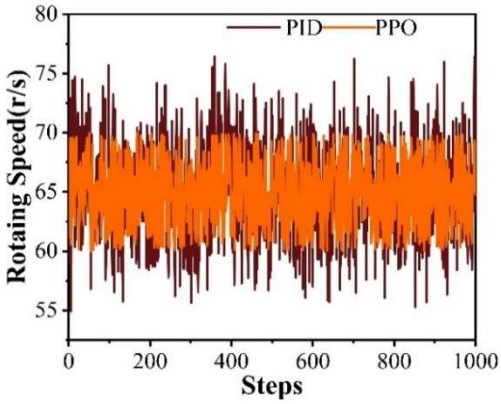
distribution, the speed fluctuation range is smaller, and closer to the desired target speed. Especially after the load increases, it is obviously more concentrated and appears far more frequently near the target speed, which indicates that the PPO algorithm is more adaptable and can show satisfactory control performance in both light load and heavy load environments.



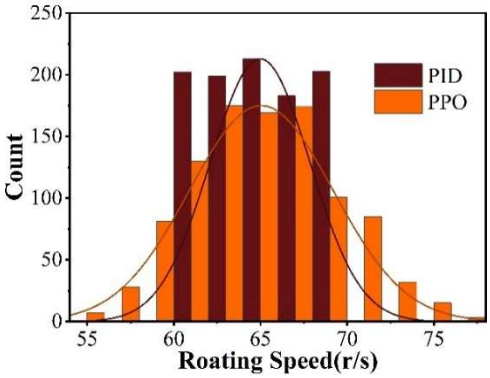
(a) Light-load speed curve



(b) Light-load speed distribution



(c) Heavy-load speed curve



(d) Heavy load speed distribution

Figure 2: Stability analysis results

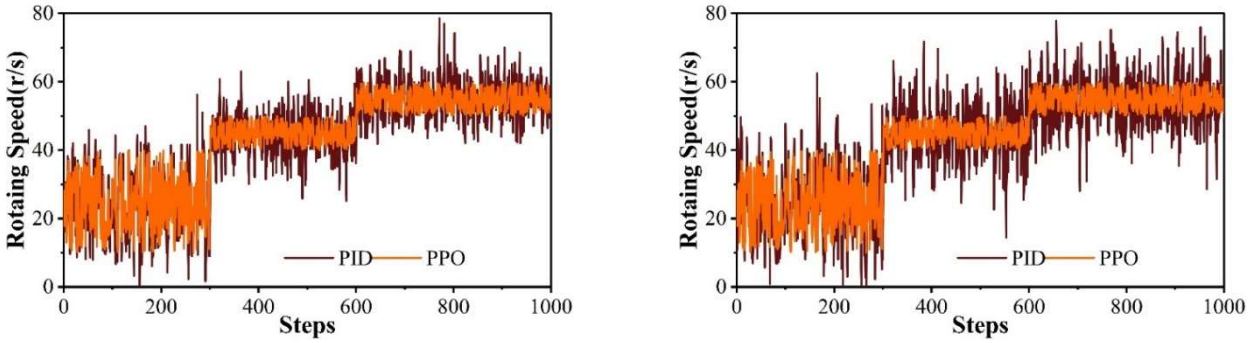
The stability comparison analysis is shown in Table 1. From the table, it can be seen that under light load environment (input 40mm scrap), the PPO algorithm improves 9.41%, 33.33%, and 32.22% in the performance of the three parameters of rise time, overshooting amount, and regulation time, respectively, compared with the traditional PID algorithm. Under heavy load environment (80mm scrap input), the PPO algorithm improves 69.52%, 72.24%, and 52.95% in the performance of the three parameters compared to the traditional PID algorithm. After the load increase, the fixed parameters of the traditional PID are no longer applicable to the current environment, but the PPO algorithm can adaptively adjust the steelmaking process parameters in the steel industry according to the environmental changes, which makes the system work more stable, and verifies the stability of the reinforcement learning-based adaptive adjustment system for steelmaking process parameters in the steel industry.

Table 1: Comparative analysis of stability

Algorithm	Rising time/s	Overshoot	Adjust the time/s
PID (Light)	1.392	16.824	2.132
PPO (Light)	1.261	11.217	1.445
PID (Heavy)	4.177	54.315	3.088
PPO (Heavy)	1.273	15.076	1.453

III. A. 2) Robustness analysis

Robustness refers to the algorithm's ability to tolerate changes in steelmaking process parameters, external disturbances and other uncertainties in the steel industry. In the steel industry steelmaking process parameters adaptive adjustment system, the actual engineering environment may exist parameter changes, external disturbances and other factors, a robust algorithm can be in the system parameter changes or external disturbances, but still maintain a better control performance, so that the steel industry steelmaking process parameters adaptive adjustment system can work more stably. For the steelmaking working environment, robustness is mainly reflected in the motor for the external load can react to sudden changes, such as a sudden change in the set speed, the system can still drive the motor to quickly reach the set speed, or when the external load suddenly increased, the motor will not be jammed because of the sudden increase in resistance. First of all, the speed change, the set speed from 30r / s to 45r / s followed by a sudden change to 55r / s, every 20ms to collect the speed and record, respectively, using the traditional PID algorithm, PPO algorithm for the experiments, drawing speed - time curve as shown in Figure 3, of which (a) ~ (b) were 40mm steel scrap, 80mm steel scrap. It can be seen that, both in the light load environment (into the 40mm scrap) and heavy load environment (into the 80mm scrap), even if the target speed changes within a short time, the PPO algorithm can still keep the system running stably, and the inhibition effect of speed fluctuation is better than that of the traditional PID algorithm.



(a) 40mm scrap steel (b) 80mm scrap steel

Figure 3: Rotational speed - time curve

In order to show the experimental results more intuitively, the standard deviation of the control speed of each algorithm at different rotational speeds is calculated separately, and the specific results are shown in Table 2. It can be seen that, in the light load environment (input 40mm scrap), with the speed from 30r/s to 55r/s, the standard deviation of the PPO algorithm to control the rotational speed from 3.945 to 3.053, while the PID algorithm is from 5.283 to 6.164. In the heavy load environment (input 80mm scrap), the PPO algorithm has a higher priority than the PID algorithm, which fully proves that the PPO algorithm based on the PID algorithm can control the rotational speed of steelmaking in the steel industry. It fully proves the robustness of the adaptive adjustment system of steelmaking process parameters in steel industry based on PPO algorithm, and provides solid theoretical support for the following research on optimal control of energy consumption.

Table 2: Comparison of variable speed standard deviations

Algorithm	40mm			80mm		
	30r/s	45r/s	55 r/s	30r/s	45r/s	55 r/s
PID	5.283	6.172	6.164	8.112	9.056	9.284
PPO	3.945	3.351	3.053	5.227	5.184	4.527

III. B. Empirical analysis of optimal control of energy consumption

The analysis above shows that the effectiveness of the adaptive adjustment system of steelmaking process parameters in the steel industry based on the PPO algorithm ensures the smooth implementation of the empirical analysis of the optimal control of energy consumption. This subsection will explore the optimal control of energy consumption of the system based on reinforcement learning from two aspects of the same furnace size problem and different furnace size problem.

III. B. 1) Optimized control of energy consumption with the same furnace size

In order to verify that different algorithms solve the energy consumption optimal control problem under the same furnace size conditions, simulation experiments are conducted, the raw data of three springs and ten furnace schedules are shown in Table 3, and the PPO algorithm (proximal policy optimization algorithm), the DDPG algorithm (deep deterministic policy gradient), and SAC (maximum entropy reinforcement learning) are used to solve the energy consumption optimal control problem, respectively. In the first step, we compare the three algorithms and summarize the experimental results to demonstrate the advantages of the PPO algorithm in solving the optimal energy consumption control problem. In the second step, in order to avoid the chance of the results of one experiment, we run the PPO algorithm (Proximal Policy Optimization Algorithm), DDPG algorithm (Deep Deterministic Policy Gradient), and SAC (Maximum Entropy Reinforcement Learning) independently for 10 times, and then we compare and analyze the differences of the data of the 10 experiments and summarize the results of the experiments, so as to further reflect the priority of the PPO algorithm (Proximal Policy Optimization Algorithm) in solving the energy consumption optimal control problem. control problem.

Table 3: The original data of the plan for three pouring times and ten furnace times

Pouring number	Furnace Number	Steel grade	Manufacture command number	Casting destination	Scheduled watering time
1	1	DV3943D1	115739	1#CC	6:20
	2	DT0912D1	115540	1#CC	
	3	DV3948D1	115041	1#CC	
	4	DT0138D1	115044	1#CC	
2	5	AP0740D5	118238	2#CC	6:10
	6	AP0740D5	118239	2#CC	
	7	DV3943D1	118241	2#CC	
3	8	DV3943D1	461324	3#CC	6:30
	9	DV3943D1	461325	3#CC	
	10	DV3943D1	461326	3#CC	

In order to gain a deeper understanding of the differences between the three algorithms in solving the energy consumption optimal control problem, and also to avoid the chance of the results of one experiment. Therefore, each algorithm was run independently for 10 times against the experimental data in Table 3, and the experimental results were summarized, and the analysis of the optimal control of energy consumption under the same scale is shown in Table 4. The results show that in the same furnace scale conditions, the PPO algorithm (maximum completion time: 2.581s, total energy consumption: 136.615KJ) compared with the DDPG algorithm (maximum completion time: 6.492s, total energy consumption: 261.152KJ), SAC (maximum completion time: 6.482s, total energy consumption: 229.604KJ), can get the most optimal energy consumption control scheme.

Table 4: Analysis of energy consumption optimization control under the same scale

No.	Maximum completion time/s			Total energy consumption/KJ		
	DDPG	SAC	PPO	DDPG	SAC	PPO
1	5.23	5.69	2.98	269.64	249.9	140.36
2	5.72	7.27	2.38	261.59	200.2	149.3
3	7.94	7.92	2.27	296.59	255.9	121.14
4	5.25	5.37	2.57	251.77	240.1	117.98
5	5.12	7.06	2.86	299.96	244.46	148.88
6	6.25	6.59	3.56	288.83	226.33	127.01
7	6.66	5.17	3.71	223.01	229.36	130.88
8	7.74	7.05	1.56	270.54	208.46	133.84
9	7.95	7.53	2.39	232.89	203.01	149.46
10	7.06	5.17	1.53	216.7	238.32	147.3
Mean	6.492	6.482	2.581	261.152	229.604	136.615

III. B. 2) Different furnace sizes

In this subsection, three algorithms are used to conduct comparative experiments on energy consumption optimal control problems with different furnace sizes, and the experimental results are analyzed to further prove the

superiority of the PPO algorithm in energy consumption optimal control problems. In order to further improve the experiment and increase the persuasive power of the experiment, three different algorithms are used to solve the energy consumption optimal control problems of seven, ten and twelve furnaces, and each algorithm is run independently for ten times, and then the experimental results are recorded and averaged for the comparison of the results, and the results of the three algorithms are shown in Tables 5, 6, and 7 for the simulation experiments on the different furnaces, respectively. Table 5, Table 6 and Table 7 show that under the condition of the same furnace size, with the increase of the number of tests, the average values of the running speed and the start time of the DDPG algorithm and the SAC algorithm are increased, while the average values of the maximum completion time and the total energy consumption of the PPO algorithm are less affected by the number of tests. And with the gradual increase of furnace size, the maximum completion time and total energy consumption of the PPO algorithm are smaller than that of the DDPG algorithm and SAC algorithm, which proves the priority of the PPO algorithm in the optimal control problem of energy consumption.

Table 5: Optimal energy consumption control for seven cycles

No.	Maximum completion time/s			Total energy consumption/KJ		
	DDPG	SAC	PPO	DDPG	SAC	PPO
1	6.743	5.885	3.027	270.881	222.444	124.009
2	6.195	5.715	3.041	274.428	242.880	109.567
3	7.061	5.970	2.695	267.796	269.171	87.173
Mean	6.666	5.857	2.921	271.035	244.832	106.916

Table 6: Optimal energy consumption control for ten cycles

No.	Maximum completion time/s			Total energy consumption/KJ		
	DDPG	SAC	PPO	DDPG	SAC	PPO
1	6.170	6.158	2.135	271.704	208.631	147.006
2	6.177	5.308	2.818	221.164	152.266	115.826
3	5.946	4.476	3.662	218.136	153.596	156.297
Mean	6.098	5.314	2.872	237.001	171.498	139.710

Table 7: Optimal energy consumption control for twelve furnace cycles

No.	Maximum completion time/s			Total energy consumption/KJ		
	DDPG	SAC	PPO	DDPG	SAC	PPO
1	5.292	5.704	3.129	232.739	250.956	160.411
2	6.490	5.564	2.724	227.884	131.246	200.837
3	6.702	5.334	3.314	183.533	205.422	148.249
Mean	6.162	5.534	3.056	214.719	195.875	169.832

IV. Conclusion

In this paper, under the guidance of reinforcement learning and steelmaking process theory, an adaptive regulation system of steelmaking process parameters and energy consumption optimization control strategy based on PPO algorithm in the steel industry is designed, and the validity of this paper's research is verified from several aspects. It is found that under light load environment, compared with the traditional PID algorithm, the performance of this paper's algorithm is especially outstanding in the three parameters of rise time (1.261s), overshooting amount (11.217), and regulation time (1.445s), which verifies the stability of the steelmaking process parameter adaptive regulation system based on the PPO algorithm. In addition, in terms of the energy consumption optimization control strategy, the priority of this paper's algorithm (maximum completion time: 2.581s, total energy consumption: 136.615KJ) is much higher than that of the DDPG algorithm (maximum completion time: 6.492s, total energy consumption: 261.152KJ) and SAC algorithm (maximum completion time: 6.482s, total energy consumption: 229.604KJ), which verifies the stability of the energy consumption optimization system based on PPO algorithm. The practical application value of the energy consumption optimization control strategy based on PPO algorithm.

References

[1] He, K., Wang, L., & Li, X. (2020). Review of the energy consumption and production structure of China's steel industry: Current situation and future development. Metals, 10(3), 302.

[2] Zhang, J., Shen, J., Xu, L., & Zhang, Q. (2023). The CO2 emission reduction path towards carbon neutrality in the Chinese steel industry: A review. Environmental Impact Assessment Review, 99, 107017.
[3] Dong, H., Liu, Y., Wang, L., Li, X., Tian, Z., Huang, Y., & McDonald, C. (2019). Roadmap of China steel industry in the past 70 years. Ironmaking & Steelmaking, 46(10), 922-927.
[4] Long, J. Y., Zheng, Z., Gao, X. Q., & Gong, Y. M. (2014). Production planning system for the whole steelmaking process of Panzhihua Iron and Steel Corporation. Journal of Iron and Steel Research, International, 21, 44-50.
[5] Cui, J., Meng, G., Zhang, K., Zuo, Z., Song, X., Zhao, Y., & Luo, S. (2025). Research progress on energy-saving technologies and methods for steel metallurgy process systems—A review. Energies, 18(10), 2473.
[6] Clijsters, S., Craeghs, T., & Kruth, J. P. (2012). A priori process parameter adjustment for SLM process optimization. Innovative developments on virtual and physical prototyping, 553-560.
[7] Zhang, X. W. (2021). Optimization and Adjustment of Production Process Parameters in Order to Reduce Production Costs and Increase Quality. Industrial Engineering & Management Systems, 20(4), 621-628.
[8] Zhao, L., Diao, G., & Yao, Y. (2015). A dynamic process adjustment method based on residual prediction for quality improvement. IEEE Transactions on Industrial Informatics, 12(1), 41-50.
[9] Wang, Y., Liu, J., Zhou, L., Cong, L., & Sutherland, J. W. (2024). Integrated operation planning and process adjustment for optimum cost with attention to manufacturing quality and waste. Journal of Manufacturing Systems, 73, 241-255.
[10] Ahmed, N., Barsoum, I., Haidemenopoulos, G., & Al-Rub, R. A. (2022). Process parameter selection and optimization of laser powder bed fusion for 316L stainless steel: A review. Journal of Manufacturing Processes, 75, 415-434.
[11] Hajnys, J., Pagac, M., KOTERA, O., PETRU, J., & Scholz, S. (2019). INFLUENCE OF BASIC PROCESS PARAMETERS ON MECHANICAL AND INTERNAL PROPERTIES OF 316L STEEL IN SLM PROCESS FOR RENISHAW AM400. MM Science Journal.
[12] Backman, J., Kyllönen, V., & Helaakoski, H. (2019). Methods and tools of improving steel manufacturing processes: Current state and future methods. IFAC-PapersOnLine, 52(13), 1174-1179.
[13] Wang, Q., Zhang, Z., Tong, X., Dong, S., Cui, Z., Wang, X., & Ren, L. (2020). Effects of process parameters on the microstructure and mechanical properties of 24CrNiMo steel fabricated by selective laser melting. Optics & Laser Technology, 128, 106262.
[14] Mudhaffar, M. A., Saleh, N. A., & Aassy, A. (2017). Influence of hot clad rolling process parameters on life cycle of reinforced bar of stainless steel carbon steel bars. Procedia Manufacturing, 8, 353-360.
[15] Yan, Y., & Lv, Z. (2021). A novel multi-objective process parameter interval optimization method for steel production. Metals, 11(10), 1642.
[16] Muthuram, N., & Frank, F. C. (2021). Optimization of machining parameters using artificial Intelligence techniques. Materials Today: Proceedings, 46, 8097-8102.
[17] Altuğ, M., & Söyler, H. (2023). Optimization with artificial intelligence of the machinability of Hardox steel, which is exposed to different processes. Scientific Reports, 13(1), 14100.
[18] Liu, C., Tang, L., & Zhao, C. (2023). A novel dynamic operation optimization method based on multiobjective deep reinforcement learning for steelmaking process. IEEE Transactions on Neural Networks and Learning Systems, 35(3), 3325-3339.
[19] Dharmadhikari, S., Menon, N., & Basak, A. (2023). A reinforcement learning approach for process parameter optimization in additive manufacturing. Additive Manufacturing, 71, 103556.
[20] Andreiana, D. S., Acevedo Galicia, L. E., Ollila, S., Leyva Guerrero, C., Ojeda Roldán, Á., Dorado Navas, F., & del Real Torres, A. (2022). Steelmaking process optimised through a decision support system aided by self-learning machine learning. Processes, 10(3), 434.
[21] Deng, J., Sierla, S., Sun, J., & Vyatkin, V. (2023). Offline reinforcement learning for industrial process control: A case study from steel industry. Information Sciences, 632, 221-231.
[22] Pan Ruilin,Wang Qiong,Cao Jianhua & Zhou Chunliu. (2024). Deep reinforcement learning for solving steelmaking-continuous casting scheduling problems under time-of-use tariffs. International Journal of Production Research,62(1-2),404-420.
[23] Andreiana Doru Stefan,Acevedo Galicia Luis Enrique,Ollila Seppo,Leyva Guerrero Carlos,Ojeda Roldán Álvaro,Dorado Navas Fernando & del Real Torres Alejandro. (2022). Steelmaking Process Optimised through a Decision Support System Aided by Self-Learning Machine Learning. Processes, 10(3),434-434.
[24] Ståhl Niclas,Mathiason Gunnar & Alcacoas Dellainey. (2021). Using Reinforcement Learning for Generating Polynomial Models to Explain Complex Data. SN Computer Science,2(2).