DDPG-based continuous thickness and tension coupling control for the unsteady cold rolling process

Wenying Zeng · Jinkuan Wang · Yan Zhang · Yinghua Han · Qiang Zhao

Received: 11 December 2021 / Accepted: 18 April 2022 / Published online: 25 April 2022
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

Abstract
Cold rolling is an important part of the iron and steel industry, and the unsteady rolling process of cold rolling usually brings significant influences on the stability of product quality. In the unsteady rolling process, various disturbances and uncertainties such as variable lubrication state, variable equipment working conditions lead to difficulties in the establishment of state space model of thickness and tension, which has become a thorny problem in thickness and tension control. In this paper, we present a model-free controller based on Deep Deterministic Policy Gradient (DDPG), which can continuously control the thickness tension of the unsteady rolling process without the mathematical model. We first formulate the thickness and tension control problem to Markov Decision Process (MDP). We apply strategies such as dividing state space variables with the mechanism model, defining reward function and state normalization, the random disturbance and complex uncertainties of the unsteady cold rolling process are coped with by utilizing the DDPG controller. In addition, these strategies also ensure the learning performance and stability of DDPG controller under random disturbance. Simulations and experiments show that the proposed the DDPG controller does not require any prior knowledge of uncertain parameters and can operate without knowing unsteady rolling mathematical models, which has better accuracy, stability, and rapidity for thickness and tension in the unsteady rolling process than proportional integral (PI) controller. The artificial intelligence–based controller brings both product quality improvement and intelligence to cold rolling.

Keywords  Deep reinforcement learning · Cold rolling · Neural network · Thickness and tension control · Unsteady rolling process

1 Introduction
Cold rolling strip is one of the important parts of strip products, which has been widely used in the automotive industry, aerospace, shipbuilding, bridge, architecture, electronic, and household appliance industries. The cold rolling strip is produced by the cold rolling production line composed of stands in which the hot strip passes through each stand in turn, and each stand applies rolling force to the strip through pairs of work rolls, the thickness of the strip reduces continuously.

Figure 1 is the equipment distribution diagram of a five-stand cold rolling line. The thickness of the strip is an important quality indicator of cold rolling strip. In recent years, the demand for the cold rolling strip is constantly increasing, which brings new challenges to the production quality and the production efficiency of the cold rolling strip. Once the tension of the strip is out of range, it will cause the strip fracture which shows that the tension control of the strip is the key to guaranteeing the stable production [1]. For the cold rolling process, there is a strong interaction between strip thickness and tension. Generally, the control strategy of the thickness control and the tension is that roll gap controls tension and roll speed adjusts thickness.

The unsteady rolling process is a special part of the cold rolling process which includes acceleration, deceleration, and over weld process, and so on. With the advancement of the industrialization process, the control performance of the steady-state rolling process has reached a very high
level for the cold rolling process. However, in the unsteady rolling process, various disturbances and uncertainties such as variable lubrication state, variable equipment working conditions lead to difficulties in the establishment of the mathematical model, which cause the thickness and tension control performance in the unsteady rolling process are still inadequate. In addition, the coupling of thickness and tension in the unsteady rolling process also affects the quality of the finished strip.

The widely used Automatic Gauge Control (AGC) and Automatic Tension Control (ATC) in the cold rolling process based on PI controller have reached a high level. PI controller has achieved excellent control performance in Single Input Single Output (SISO) control system, but it also has many shortcomings such as inability to solve multivariable coupling problems, lack of flexibility, and poor adaptability. In the actual production process of cold rolling strip, there is a strong correlation and interaction between the thickness control system and tension control system.

To better adapt to the actual production process and improve the production quality, many researchers try to use different control theories to control the thickness and tension of cold rolling strips [2]. Wang et al. [3] constructed the mathematical model which considered random actuator failure and designed the fault-tolerant controller for the AGC system to control the thickness of the cold rolling process. Friedel et al. [4] established the rolling process model of a single stand and applied predictive functional control to thickness tension control to improve the accuracy of thickness and tension control. Some researchers focus on decoupling control of thickness and tension. Li et al. [5] designed a decoupling controller based on PID algorithm with the help of a diagonal recurrent neural network and achieved certain results on the proposed dynamic coupling model. In view of the coupling relationship of the thickness and tension, a decoupling compensator based on invariance principle was designed and improve the accuracy of thickness and tension [6]. An [7] designed the multivariable decoupling control system of thickness and tension of cold rolling based on the neural network with the genetic algorithm which can resist small disturbances in the system. However, the above methods did not consider the random disturbance in the unsteady rolling process, which led to the insufficient thickness tension control in the unsteady rolling process.

To improve control performance and enhance system stability, many researchers have established mathematical models considering disturbance terms and applied them to the thickness tension control. Hu et al. [8, 9] introduced a new state space model of the cold rolling process, proposed the multivariable optimization strategy based on inverse linear quadratic form and the control method based on receding horizon control strategy respectively to optimize the control performance of thickness and tension. Koofigar et al. [10] established a mathematical model considering the uncertainty of cold rolling process, and used the model to design a robust controller based on H∞ control theory. Ogasahara et al. [11] designed a controller based on explicit model predictive control, which obtained good control performance in the acceleration process with a small disturbance. In the unsteady rolling process with random disturbance, the performance of the controller is still insufficient. Hu et al. [12] established a state space model with constraints and proposed the distributed model predictive control online optimal control strategy which has good tracking performance compared with PI controller. Ozaki et al. [13] proposed a nonlinear mathematical model to accurately control the tension and thickness of the cold rolling system during the acceleration process and deceleration process. In order to analyze the characteristics of the unsteady rolling process, Cao et al. [14] proposed an unsteady lubrication model and developed a dynamic chatter model, and analyzed the effects of the main process parameters on the stability of rolling stand. The methods mentioned above can be categorized as model-based methods, and the parameters are constant in the unsteady rolling process. However, accurate model parameters cannot be obtained because of many unexpected situations, such as insufficient data samples [15]. In addition, when the system state changes, the controller based on
mathematical model needs to change the parameters affected in the model to achieve optimal performance.

Meanwhile, Reinforcement Learning (RL) has developed significantly in recent years and has been applied in many fields, from wireless Internet of Things, manufacturing to industrial process control [16–18]. In recent years, Deep Reinforcement Learning (DRL) has attracted wide attention in solving high-dimensional control with high complexity [19]. DDPG is one of the model-free DRL methods for continuous action spaces and has been widely applied in many fields, such as robotic control [20], manipulator control [21], and wireless sensors [22].

The stability of the control system with neural networks was proved by the Lyapunov method in [23, 24], and DRL for control has made great progress in recent years [25]. Many researchers applied DDPG to industrial process control which has characteristics of large disturbance and strong coupling and achieved a series of results [18]. Gao et al. [26] applied the DDPG method in heating, ventilation, and air conditioning (HAVC) to achieve continuous and accurate thermal control of the system. Ma et al. [27] applied DDPG to a polymerization system with strong nonlinearity, large time delay, noise tolerance. The DDPG controller learned the optimal strategy as the optimal control strategy for controlling nonlinear valves in [28]. Spielberg et al. [29] extended the DDPG to process control problems, and evaluated their approach on SISO Systems, Multi Input Multi Output (MIMO) systems, and tested it under various scenarios as well. Notice that the DDPG method can operate without knowing the mathematical models and increase stability against uncertainty in the thickness and tension control problem of the unsteady rolling process.

Therefore, the main contributions are listed as follow:

1. In this paper, the model-free DDPG algorithm is applied in the thickness and tension control problem of the unsteady rolling process with random disturbance and complex uncertainty for the first time, which overcomes the inaccurate modeling caused by random disturbance and coupling relationship of thickness and tension.

2. We map the thickness and tension control problem of the unsteady rolling process to MDP by taking random disturbance in the unsteady rolling process into consideration, where action space of the agent and state of the environment were divided by state space model of steady-state rolling process and reward function was designed to overcome the absence of the mathematical model of unsteady rolling process and improve the control performance of the unsteady rolling process.

3. Compared with the traditional PI controller, simulation results show that the DDPG controller has better accuracy, stability, and rapidity for the thickness tension in the unsteady rolling process. Compared with the maximum tension error of PI controller of 26.13%, the maximum tension error of the DDPG controller is 11.58%. The controller based on artificial intelligence improves the quality of the cold rolling strip and realizes intelligence.

The rest of the paper is organized as follows. In Sect. 2, we describe the unsteady rolling process and mapping thickness and tension control to Markov decision process. In Sect. 3, we set up the environmental network and apply the DDPG controller to solve the problem of thickness tension control in the unsteady rolling process. Section 4 provides training results and testing results which show that the DDPG controller has better stability and rapidity compared with a PI controller. Finally, some conclusions are shown in Sect. 5.

2 Problem description

2.1 The unsteady rolling process

The unsteady rolling process is a special part of the cold rolling process which includes acceleration, deceleration and over weld process, and so on. How to control the unsteady rolling process stably and accurately has become an urgent problem in the cold rolling process. The factors that make the unsteady rolling process hard to control are listed as follows:

Due to the nonlinearity, time-varying in the unsteady rolling process coupled with various uncertainties such as variable lubrication state, variable equipment working condition, etc., it is difficult to establish an accurate mathematical model to describe the unsteady rolling process. Despite having the steady-state mathematical model for reference, the constant parameter mathematical model used in the actual control process is difficult to meet the requirements of continuous change which leads to low model matching. In addition, there are deviations in key parameters such as rolling force and forward slip in the actual production process which also leads to inaccurate results when using the existing model.

Figure 2 depicts the curve of thickness and tension control in the steady-state rolling process and the unsteady rolling process. It is noticed that the thickness control curve and tension control curve have good tracking performance to the target curve in the range of constant speed. However, in the range of velocity variation, the thickness control curve and tension control curve deviate from the target curve, which indicates that the control accuracy of thickness and tension in the acceleration and deceleration process with the PI controller is insufficient.
In our work, we propose a controller based on DRL to alleviate the insufficient control accuracy of thickness and tension in the acceleration and deceleration process. In order to deal with the difficulty of accurately describing the unsteady rolling process by the mathematical model, we adopt the strategy of constructing a deep neural network to represent the mathematical model. The specific methods will be described in Sect. 3.

2.2 State space model of thickness and tension

According to the bounce equation of the rack, the incremental equation of the output thickness can be expressed as Eq. (1). The subscripts in, out, 0 and i in the formula respectively represent entry of stand, exit of stand, i-th stand, and initial value.

\[
\Delta h_{out,i} = \Delta S_i + \frac{\Delta P_i}{M_i}
\]  

(1)

where \( h, S, P, M_i \) represent strip thickness, roll gap, rolling force, the mill stiffness, respectively.

The rolling force is affected by input thickness, output thickness, front tension, and back tension.

\[
\Delta P_i = \alpha_{i1} \cdot \Delta h_{in,i} + \alpha_{i2} \cdot \Delta h_{out,i} + \alpha_{i3} \cdot \Delta T_{in,i} + \alpha_{i4} \cdot \Delta T_{out,i}
\]  

(2)

where \( T \) is the tension of strip between adjacent stand, \( \alpha_{i1}, \alpha_{i2}, \alpha_{i3}, \alpha_{i4} \) are coefficients depending on the rolling force, roll gap, rolling speed, and stand stiffness.

According to Eqs. (1) and (2), the formulation of output thickness increment model can be expressed as

\[
\Delta h_{out,i} = e_{i1} \cdot \Delta S_i + e_{i2} \cdot \Delta h_{in,i} + e_{i3} \cdot \Delta T_{in,i} + e_{i4} \cdot \Delta T_{out,i}
\]  

(3)

where

\[
e_{i1} = \frac{M_i}{M_i - \alpha_{i2}}, \quad e_{i2} = \frac{\alpha_{i1}}{M_i - \alpha_{i2}}, \quad e_{i3} = \frac{\alpha_{i3}}{M_i - \alpha_{i2}}, \quad e_{i4} = \frac{\alpha_{i4}}{M_i - \alpha_{i2}}
\]

In the rolling process, rolling mill exit forward slip \( S_h \) and back slip \( S_H \). The forward slip coefficient \( f_i \) and the backward slip coefficient \( b_i \) are defined as Eqs. (6) and (7):

\[
S_h = \frac{V_{out} - V}{V}
\]  

(4)

\[
S_H = \frac{V - V_{in}}{V}
\]  

(5)

\[
f_i = 1 + S_h
\]  

(6)

\[
b_i = 1 - S_H
\]  

(7)

where \( V \) represents rolling speed.
\( V_{in}h_{in} = V_{out}h_{out} \)  \( (8) \)

The backward slip coefficient is defined as the following,

\[
\frac{d(\Delta T_{in,i})}{dt} = \frac{EBh_{in,i}}{L} \cdot (\Delta V_{in,i} - \Delta V_{out,i-1}) \quad (14)
\]

Combined with the above formula, the change rate of tension increment is deduced as follows,

\[
\frac{d(\Delta T_{in,i})}{dt} = \frac{EBh_{in,i}}{L} \cdot \left\{ -V_{i-1,o} \cdot (\beta_{2i-1} \cdot e_{3i-1} + \beta_{3i-1}) \cdot \Delta T_{in,i-1} 
+ \left[ \frac{V_{io}}{h_{in,io}} (h_{out,io} \cdot \beta_{2i} \cdot e_{3i} + h_{out,io} \cdot \beta_{3i} + f_{io} \cdot e_{3i}) - V_{i-1,o} \cdot \beta_{2i-1} \cdot e_{4i-1} - V_{i-1,o} \cdot \beta_{4i-1} \right] \cdot \Delta T_{in,i} 
+ \frac{V_{io}}{h_{in,io}} (h_{out,io} \cdot \beta_{2i} \cdot e_{4i} + h_{out,io} \cdot \beta_{4i} + f_{io} \cdot e_{4i}) \right\} 
+ \frac{V_{io}}{h_{in,io}} (h_{out,io} \cdot \beta_{2i} + h_{out,io} \cdot \beta_{4i} + f_{io} \cdot e_{21} - f_{io} \cdot h_{out,io} / h_{in,io}) \cdot \Delta h_{in,i} 
- f_{i-1,o} \cdot \Delta V_{i-1} + b_{io} \cdot \Delta V_{i} - V_{i-1,o} \cdot \beta_{2i-1} \cdot e_{1i} \cdot \Delta S_{i-1} + V_{io} \cdot e_{1i} \right\} \right\} 
+ \frac{V_{io}}{h_{in,io}} (h_{out,io} \cdot \beta_{2i} + f_{io} \cdot \Delta S_{i}) \]

\[
= \xi_{1} \cdot \Delta T_{in,i-1} + \xi_{2} \cdot \Delta T_{in,i} + \xi_{3} \cdot \Delta T_{in,i+1} + \xi_{4} \cdot \Delta h_{in,i-1} + \xi_{5} \cdot \Delta h_{in,i} 
+ \xi_{6} \cdot \Delta V_{i-1} + \xi_{7} \cdot \Delta V_{i} + \xi_{8} \cdot \Delta S_{i-1} + \xi_{9} \cdot \Delta S_{i}
\]

\( b_{i} = \frac{f_{i} \cdot h_{out,i}}{h_{in,i}} \quad (9) \)

Forward slip is related to the friction coefficient, deformation resistance, and strip thickness and tension between stands, we defined the increment of the forward slip coefficient as the following:

\[
\Delta f_{i} = \beta_{1i} \cdot \Delta h_{in,i} + \beta_{2i} \cdot \Delta h_{out,i} + \beta_{3i} \cdot \Delta T_{in,i} + \beta_{4i} \cdot \Delta T_{out,i} \quad (10)
\]

where

\[
\beta_{1i} = \frac{\partial f_{i}}{\partial h_{in,i}}, \quad \beta_{2i} = \frac{\partial f_{i}}{\partial h_{out,i}}, \quad \beta_{3i} = \frac{\partial f_{i}}{\partial T_{in,i}}, \quad \beta_{4i} = \frac{\partial f_{i}}{\partial T_{out,i}}
\]

The increment of rolling speed and backward slip coefficient can be expressed as the following:

\[
\Delta V_{in,i} = \Delta V_{i} \cdot b_{io} + \Delta b_{i} \cdot V_{io} \quad (11)
\]

\[
\Delta V_{out,i-1} = \Delta V_{i-1} \cdot f_{i-1,o} + \Delta f_{i-1} \cdot V_{i-1,o} \quad (12)
\]

\[
\Delta b_{i} = \frac{\Delta f_{i} \cdot h_{out,i}}{h_{in,i}} + \frac{f_{io} \cdot \Delta h_{out,i}}{h_{in,i}} - \frac{f_{io} \cdot h_{out,i}^{2}}{h_{in,i}^{2}} \cdot \Delta h_{in,i} \quad (13)
\]

According to Hooke’s law and the tension formula between frames, the change rate of tension increment between frames can be written as:
Based on the field data and rolling experience, the parameter change rate of the actuator can be defined as follows:

\[
d\frac{\Delta S_i}{dt} = \frac{\Delta U_{Si}}{\tau_S} - \Delta S_i, \quad S_i(0) = 0
\]

(16)

\[
d\frac{\Delta V_i}{dt} = \frac{\Delta U_{Vi}}{\tau_V} - \Delta V_i, \quad V_i(0) = 0
\]

(17)

where \(\Delta U_{Si}\) and \(\Delta U_{Vi}\) are the roll gap target value and roll speed target value of the stand \(i\) respectively. \(\tau_S\) is the time constant of ASR system and \(\tau_V\) is the time constant of HGC system. By adjusting the control signal \(U_{Si}\) and \(U_{Vi}\), the thickness of strip and the inter-stand tension can track the reference value.

The control structure of stand 5 is shown in Fig. 3. Based on the above formula and state space theory, the state space model of tension and thickness between adjacent stand can be defined as the following:

\[
\begin{aligned}
&dx_i(t) = \tilde{A}_i x_i(t) + \tilde{B}_i u_i(t) + \tilde{W}_{i1} d_i(t) \\
&y_i(t) = \tilde{C}_i x_i(t) + \tilde{W}_{i2} d_i(t)
\end{aligned}
\]

(18)

where

\[
\begin{aligned}
x_i(t) &= [\Delta T_{in,i} \Delta S_i \Delta V_{i-1}]^T, \quad u_i(t) = [\Delta U_{Si} \Delta U_{Vi-1}]^T, \\
&y_i(t) = [\Delta h_{out,i} \Delta T_{in,i}]^T.
\end{aligned}
\]

3 Optimal control with deep reinforcement learning

3.1 RL and DDPG

RL is an important branch of machine learning. RL algorithm maps the environment state detected by the agent to the action of the agent. Through a trial-and-error learning process in the environment, the agent takes the action of maximizing the long-term discount reward in Eq. (19) where the discount factor \(0 < \gamma < 1\).

\[
R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k} = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \cdots
\]

(19)

Many RL algorithms, such as Q-learning, SARSA, DQN, and so on, rely on the action-value function estimated from
experience which is called the value-based RL algorithm. Trial-and-error search and delayed rewards are prominent features of RL and provide a solid foundation for RL. In essence, the action-value function estimates the total reward \( R_t \) from a state \( s \) by taking a certain action follow policy \( \pi \). The action-value function is expressed by Eq. (20)

\[
Q^\pi = E_r [R_t | s, a] = E_r \left[ \sum_{k=0}^{\infty} \gamma^k r_{t+k} | s, a \right]
\]  

(20)

Once the action-value function \( Q^\pi \) converges to the optimal policy with sufficient experience, the action taken in each state will change the current state to the state with the highest action-value function estimate.

Usually, the action-value function is represented by nonlinear function approximators. If action-value functions are not generalized, the architecture of the action-value function will no longer be applicable [30]. DQN and DDPG update the weights \( \theta^Q \) of the Q-network iteratively with minimizing the loss function \( L \).

\[
L = \frac{1}{N} \sum_i (y_i - Q(s_i, a_i | \theta^Q))^2
\]  

(21)

RL was initially proposed to deal with the discrete action-space problem, but there are many situations that require continuous action [31]. An algorithm called actor-critic based on policy gradient for continuous space was proposed by Sutton et al. [32]. The actor-critic consists of two parts: one is the actor network updated by the policy gradient, the other is the critic network estimating the action-value function \( Q^\pi \).

Figure 4 shows the schematic diagram of the actor-critic algorithm.

However, the actor-critic algorithm needs a long training cycle, and it is difficult to converge to the optimal result. Lillicrap developed DDPG based on learning in mini-batch from DQN to ensure convergence and improve training speed [33]. Replay buffer is used in DQN to store historical samples \( [s_i, a_i, r_i, s_{i+1}] \). In the training period, the network’s weight is updated by the mini-batch tuple data that were sampled randomly.

Referring to the process from DQN to DDQN, DDPG also adopts the concepts of the target network and current network. The weight updating speed of the target network is slow, which greatly improves the stability of the network [34]. DDPG is a deterministic policy, different from the actor-critic algorithm, its actor policy is updated by the sampled gradient \( \nabla_{\theta^\pi} \pi_{\theta^\pi}(s) |_{s_i} \)

\[
\nabla_{\theta^\pi} \pi_{\theta^\pi} |_{s_i} \approx \frac{1}{N} \sum_i \nabla_{\theta^Q} Q(s, a | \theta^Q) |_{s_i} \nabla_{\theta^\pi} \pi_{\theta^\pi}(s) |_{s_i}
\]  

(22)

The target network is slowly updated by Eqs. (23) and (24).

\[
\theta^Q \leftarrow \tau \theta^Q + (1-\tau) \theta^Q'
\]  

(23)

\[
\theta^\pi \leftarrow \tau \theta^\pi + (1-\tau) \theta^\pi'
\]  

(24)

### 3.2 Mapping thickness and tension control to MDP

In this subsection, we map the thickness and tension control problem of the unsteady rolling process as a Markov Decision Process (MDP), which is then solved by a model-free DRL algorithm in Sect. 3.3.

MDP is a mathematical model of the sequential decision in RL [34]. It describes the completely observable environment, including four basic elements: state \( s \), action \( a \), reward \( r \), state transition probability \( p \). In MDP, the agent constantly interacts with its environment, chooses actions \( a \) according to policy, and the environment responds to it by presenting a new situation \( s_{i+1} \), and the environment also gives feedback reward \( r \) to the agent.

According to the state space model of single stand in Eq. (18), the thickness and tension of strip steel at the current time are only affected by the previous state parameters and not by the state parameters at any other time.

In the thickness and tension control problem of the unsteady rolling process, the four basic elements are described as follows:

- State \( s \): (1) current entry tension of \((i-1)\)th stand \( T'_{m-1} \); (2) current entry tension of \((i+1)\)th stand \( T'_{m+1} \); (3) current entry tension of \(i\)th stand \( T'_{m} \); (4) current roll gap of \((i-1)\)th stand \( S'_{m-1} \); (5) current roll gap of \(i\)th stand \( S'_{m} \); (6) current

![Fig. 4 Schematic diagram of actor-critic algorithm](image_url)
rolling speed of \((i-1)\)th stand \(V^r_{i-1}\); (7) current rolling speed of \(i\)th stand \(V^r_{i}\); (8) current entry thickness of \(i\)th stand \(h^r_{in,i}\); (9) current entry thickness of \(i\)th stand \(h^r_{in,i-1}\); (10) current exit thickness of \(i\)th stand \(h^r_{out,i}\).

Notice that the state parameters include disturbance term \(d_i(t)\), input term \(x_i(t)\), and output term \(y_i(t)\) of state space model.

Action \(a_i\): (1) The set-point of roll gap \(U^r_{Si}\); (2) The set-point of Roll speed \(U^r_{Si}\); Action includes control term \(u_i(t)\) of state space model which \(U^r_{Si}\) and \(U^r_{Si-1}\) are continuous variable.

We define tension error \(\Delta T^i_{in,i}\) and thickness error \(\Delta h^i_{out,i}\) as follows:

\[
\Delta T^i_{in,i} = T^i_{in,i} - T_{ref}
\]
\[
\Delta h^i_{out,i} = h^i_{out,i} - h_{ref}
\]

Equation (25)

Reward: We design the reward function by comprehensively considering the accuracy of thickness and tension of rolling stand and the convergence efficiency of the agent training process, which is defined as follows:

\[
r = \begin{cases} 
- (\Delta T^2_{in,i} + \Delta h^2_{out,i}) + 0.08, & \text{if} \ (\Delta T^2_{in,i} + \Delta h^2_{out,i}) \leq 0.1 \\
- (\Delta T^2_{in,i} + \Delta h^2_{out,i}) + 0.05, & \text{if} \ 0.1 \leq (\Delta T^2_{in,i} + \Delta h^2_{out,i}) \leq 0.2 \\
- (\Delta T^2_{in,i} + \Delta h^2_{out,i}), & \text{else}
\end{cases}
\]

Equation (26)

We define the reward function as the negative number of the square sum of tension error and thickness error. This form of the reward function can avoid the excessive thickness error or tension error, and the thickness and tension can be controlled at the reference value to maintain stable control performance. For the DDPG algorithm, if the average value of the reward is closer to 0, the optimal strategy will be easier to learn as well as the better convergence speed of agent training. When the square sum of tension error and thickness error is between 10 and 20%, the agent will be given an additional reward of 0.05. When the square sum of tension error and thickness error is less than 10%, the agent will be given an additional reward of 0.08. According to the definition in Eq. (26), the higher the reward value is, the more the thickness and tension curves fit the set-point, which also represents that the controller has better control performance. The MDP of the thickness and tension control problem of the unsteady rolling process is illustrated in Fig. 5.

For the above MDP, the state transition probability \(p\) is not defined. The state transition probability is the probability distribution of state transition to other states after the agents take certain actions \(U^r_{Si}\) and \(U^r_{Si-1}\) in a certain state. When the state transition probability is known, the MDP is completely observable, the cumulative reward can be solved by some iterative methods. In the thickness and tension control problem of unsteady rolling process, obtaining accurate state transition probability models is a complex task, because an accurate mathematical model is difficult to establish as mentioned in Sect. 2.

Based on the above considerations, we adopted the model-free DDPG algorithm to overcome the uncertainties of the unsteady rolling process and cross-coupling between stands. Without prior knowledge of the environment or state transition, the model-free DDPG algorithm learns the control strategy of thickness tension from the environment with the growth of the training episode by interacting with the environment.

3.3 Thickness and tension control framework

In the thickness and tension control problem of the unsteady rolling process, the simulation starts from a random initial state. In each step of each episode, the action \(a_i\) is generated by the deterministic policy of the actor network with random OU noise for exploration [31]. The noise decreases with the increase of episodes. The neural network simulating the rolling state will return a set...
of training data to the replay buffer at each time step, in which the reward value is calculated according to the formula in Eq. (26).

After storing enough experience in the replay buffer, the optimal strategy is learned by random sampling in small batches. The update of the critic online network is updated by Eq. (21), and the action network is updated by Eq. (22). After each training step, the target critic network and the target action network are slowly updated by Eqs. (23) and (24).

The specific process is shown in Fig. 6.

The pseudocode of DDPG algorithm for thickness and tension control in cold rolling is included as following.

The flowchart of DDPG Algorithm for thickness and tension control in cold rolling is included as following.

The advantage of the DDPG controller is that it can carry out continuous control, and improves the training efficiency and stability compared with the controller based on actor-critic algorithm [31]. The DDPG controller has become one of the ideal candidates in the industrial process control field.

4 Simulation result

The training and test results of thickness and tension control of the DDPG controller of unsteady rolling process are given in this section. We compare the control effect in thickness and tension control of the PI controller and the DDPG controller to further illustrate the superiority of the DDPG controller. The results show that the DDPG controller has better stability and rapidity.

4.1 Simulation environment based on depth neural network

In this subsection, we define the simulation environment of the thickness and tension control problem of the unsteady rolling process.

According to the theory of MDP mentioned in the above section, the simulation environment should accept the current state and action to present a new situation of next time interval and reward. We use a data-driven neural network to simulate this process, and train the neural network as the simulation environment with real data. The action and the current state are the input vectors of the environment network, and the state at the next time interval is defined as the output of the environment network to simulate the rolling stand in the thickness and tension control problem of unsteady rolling process. Considering the faster training speed and better convergence, different types of output data and input data of the environment network are normalized to $[0, 1]$.

In the environment network with 12 inputs and 10 outputs, we tried the maximum neurons of 128 with five fully connected layers, but the accuracy of the environment network in the test set only reached about 75%. Considering the complexity of the thickness and tension control problem of unsteady rolling process, we increase the maximum number of neurons and the fully connected layers to better simulate the unsteady rolling process. Ultimately, an environment network with seven fully connected layers is used in the simulation, which has 512 and 256 nodes in the third and fourth hidden layers, respectively. And to normalize the

Fig. 6 The framework of thickness and tension control
output data to $[0, 1]$, we adopt sigmoid as the activation function of the output layer. In the thickness and tension control problem of unsteady rolling process, the output thickness of stand 5 directly affects the product quality, so this paper simulates the control of stand 5 in the deceleration process and acceleration process.

In order to better evaluate the performance of the environment network, 10% of the deceleration process data and acceleration process data are extracted as the test set separately. Figure 8 shows the convergence trend of accuracy over episode of deceleration process and acceleration process. The accuracy of the deceleration process environment network reaches 0.8 after 200 episodes and 0.9 after 700 episodes is shown in Fig. 8a. At last, the training accuracy was maintained at 0.93 and stabilized. Similarly, in Fig. 8b, the accuracy of the acceleration process environment network reaches 0.8 at about 500 episodes and 0.9 at about 1000 episodes. Finally, the accuracy was maintained at 0.92 and stabilized.

The accuracy of the trained deceleration and acceleration process environment network is in the test set 92.44% and 91.82% separately, which indicates that the environment network has good generalization and can well characterize the unsteady rolling process of the rolling mill in the deceleration process and acceleration process.

### 4.2 Training result

We construct a framework of thickness tension control in the unsteady rolling process on Python platform based on tensorflow package [35]. Considering that action network and critic network only have 2 and 1 outputs, respectively, we reduce the maximum neurons and the fully connected layer compared with the environment network. We set the action network as five fully connected layers with 32, 64, 256, 128, and 32 neurons respectively. Similarly, we set the critic network as five fully connected layers, and the number of neurons is 64, 256, 128, 16, and 1 respectively. For 3000 episodes of training, it takes about 11 h to complete in Intel Core i5-4200 h CPU.

We set the time interval of each episode to 5 s, each step to 0.1 s, and each episode contains 50 steps. Figure 9a shows the total reward of each episode vs. training episodes in the deceleration process. Figure 9b shows the total reward of each episode vs. training episodes in the acceleration process. We sampled every five episodes of 3000 episodes in Fig. 9a. In 200 episodes of random exploration, the total reward of each episode is concentrated...
between − 22 and − 3. With the training start, the DDPG controller keeps learning better control strategies and obtaining the higher total reward. As mentioned in the definition of the reward function, the higher the reward is, the better the coupling control performance of thickness and tension is. The total reward converges to a maximum of about 2.5 in about 1600 episodes, and it takes about 5 h for the DDPG controller to learn the optimal control strategy.

In Fig. 9b, we also provide the total reward of each episode vs. training episodes in the acceleration process. Similarly, we sampled every five episodes of 3500 episodes. The reward of the DDPG controller decreases once in 1000–1500 episodes, and finally converges to a maximum of about 3.9 in 2000 episodes, which indicates that the DDPG controller learns the optimal control strategy in the acceleration process. It takes about 7 h for the DDPG controller to learn the optimal control strategy.

### 4.3 Test result

In this subsection, we compare the control performance in thickness and tension control of the PI controller and the DDPG controller to further illustrate the superiority of the DDPG controller. Figure 10a shows the curve of rolling speed of stand 5, which represents that the simulation of rolling stand is in the deceleration process. The results in Fig. 10b, c show that the actions the DDPG controller takes are quite different from the PI controller in the deceleration process, in which the PI controller takes a smooth control action. The resulting trajectories of the
Fig. 10 Control performance comparison between DDPG controller and PI controller in the decelerated rolling process
Fig. 11  Control performance comparison between DDPG controller and PI controller in the accelerated rolling process
DDPG controller follow closely to the target curve in the thickness and tension control problem of unsteady rolling process in Fig. 10d, e, but the curve of the PI controller fluctuates up and down and does not track the target curve well. Adopting historical data for pre-training can reduce the training episodes in the practical application stage. It takes 1600 training episodes for the controller to learn the optimal strategy as shown in Fig. 10. It is important to note that the training period for learning the optimal strategy of thickness tension control is not fixed, which is jointly determined by the data used for training environment network and the initial observation state. The strategy adopted by the DDPG controller is to increase the exit thickness of stand 5 and reduce the entry tension of stand 5 to resist the subsequent random disturbance. From the comparison of the control results in Fig. 10d, e, the maximum tension error of the DDPG controller is 11.65% compared with 26.48% to the PI controller, and the set-point of tension is 18.62kN.

The control performance of thickness and tension control in the acceleration process is shown as Fig. 11 below. Different from the thickness tension control in the deceleration process, the controller achieved the control performance as shown in Fig. 11 after about 1800 episodes in the acceleration process. The difference of training periods for learning optimal strategies in the deceleration and the acceleration process are caused by different data used for training environment network and the selection of initial observation state. Figure 11a represents the simulation of rolling stand in the acceleration process. Figure 11b, c also show the actions taken by DDPG controller are quite different from the smooth control actions of the PI controller in the acceleration process. Figure 11d, e show that the control performance of thickness and tension in the acceleration process, which DDPG controller tracks the target thickness and tension curve well but PI controller cannot. Compared with the control strategy of the PI controller, the optimal strategy adopted by DDPG controller reduces the maximum error of tension from 22.33 to 1.65%, and the set-point of tension is 23.52 kN in the acceleration process.

In dealing with the unsteady rolling process with random disturbance, DDPG controller has better stability and accuracy than the traditional PI controller, and the effect of thickness and tension control is well improved. Notice that the DDPG controller learns the control strategy without any prior knowledge to resist the random disturbance in the control system by using the strategy of sampling the replay buffer in mini-batch and using target network and current network to update the actor network.

5 Conclusion

This paper develops the DDPG controller for the thickness and tension control problem of the unsteady rolling process. The controller realizes the coupling control of thickness and tension in the deceleration process and acceleration process. By comparing, the DDPG controller has better control performance for thickness and tension in the unsteady rolling process than the PI controller. DDPG improves control accuracy and stability of the process by maximizing the value of the long-term reward function. We also prove that the DDPG controller has strong adaptability in complex process control fields with random disturbance, strong coupling, and nonlinearity. In addition, because the DDPG algorithm is a model-free RL algorithm, the control performance of the controller based on the DDPG algorithm will not be affected by the change of rolling environment. In summary, the controller based on the DDPG algorithm can control the thickness and tension of the unsteady rolling process without prior knowledge. DRL can be used to solve complex industrial problems and is different from other control methods, which indicates that the controller based on the DRL method has a great application prospect in industrial process control.

In future work, we will use parallel actors in a distributed system to comprehensively control the whole cold rolling thickness tension control system. We may also consider the prioritized replay mechanism to improve the training efficiency in the next research.

Author contribution Wenying Zeng: conceptualization, methodology, investigation, data curation, software, formal analysis, experiment, and writing of the manuscript. Jinkuan Wang: conceptualization, resources, funding acquisition, supervision, project administration, and review. Yan Zhang: data collection and curation, writing review and editing. Yinghua Han: methodology, supervision, and writing including review and editing. Qiang Zhao: methodology, supervision, and writing including review and editing.

Funding This work was supported by the National Natural Science Foundation of China (U21A20475, U1908213), Colleges and Universities in Hebei Province Science Research Program (QN2020504), The Fundamental Research Funds for the Central Universities (N2223001).

Data availability The data are available on reasonable demand.

Code availability The code is available on reasonable demand.

Declarations

Ethics approval and consent to participate All authors understand and approve the ethical responsibilities of the authors. The authors consent to participate.
Consent for publication  The authors consent to transfer the copyright of the article to publish.

Conflict of interest All authors declare no competing interests.

References

1. Zhang X, Zhang Q, Sun C (2009) Gauge and tension control in unsteady state of cold rolling using mixed H2/Hoo control. In IEEE International Conference on Control and Automation Christchurch, New Zealand, pp 9–11
2. Seung-Ho S, Seung-Ki S (2000) A new tension controller for continuous strip processing line. IEEE Trans Ind Appl 36:2. https://doi.org/10.1109/28.833782
3. Wang Y, Xia J, Wang Z, Shen H (2020) Design of a fault-tolerant output-feedback controller for thickness control in cold rolling mills. Appl Mat Comput. https://doi.org/10.1016/j.amc.2019.124841
4. Frieleb T, Zabet K, Haber R, Jelali M (2015) Predictive functional control of tandem cold metal rolling. IEEE Conf Control Appl 324–329
5. Li B, Fan X, Jiang C, Jiang G (2014) Decoupling control of thickness and tension based on DRNN-PID in cold-rolling. In Proceeding of the 11th World Congress on Intelligent Control and Automation, pp 1180–1184
6. Tan S, Wang L, Liu J (2014) Research on decoupling method of thickness and tension control in rolling process. In Proceeding of the 11th World Congress on Intelligent Control and Automation Shenyang, pp. 4715–4717
7. An S (2016) The decoupling control of tandem cold rolling tension and gauge. In 2016 3rd International Conference on Information Science and Control Engineering (ICISCE), pp. 1154–1158
8. Hu Y-J, Sun J, Wang Q-L, Yin F-C, Zhang D-H (2018) Characteristic analysis and optimal control of the thickness and tension system on tandem cold rolling. Int J Adv Manuf Technol. https://doi.org/10.1007/s00170-018-03088-1
9. Hu Y, Sun J, Chen SZ, Zhang X, Peng W, Zhang D (2019) Optimal control of tension and thickness for tandem cold rolling process based on receding horizon control. Ironmak Steelmak 1–11. https://doi.org/10.1080/03019233.2019.1615813
10. Koofigar HR, Sheikholeslams F, Hosseinia N (2011) Unified gauge-tension control in cold rolling mills: a robust regulation technique. Int J Precis Eng Manuf 12(3):393–403. https://doi.org/10.1007/s12541-011-0051-6
11. Ogashara T, Hovd M, Asano K (2016) Explicit model predictive controller design for thickness and tension control in a cold rolling mill. IFAC-PapersOnLine 49(20):126–131
12. Hu Y, Sun J, Peng W, Zhang D (2021) Nash equilibrium-based distributed predictive control strategy for thickness and tension control on tandem cold rolling system. J Process Control 97:92–102. https://doi.org/10.1016/j.jprocont.2020.11.014
13. Ozaki K, Ohtsuka T, Fujimoto K, Kitamura A, Nakayama M (2010) Nonlinear receding horizon control of thickness and tension in a tandem cold mill with a variable rolling speed. Tetsu-to-Hagane 96(7):459–467. https://doi.org/10.2355/tetsutohagane.96.459
14. Cao L, Li X, Wang Q, Zhang D (2021) Vibration analysis and numerical simulation of rolling interface during cold rolling with unsteady lubrication. Tribol Int. https://doi.org/10.1016/j.triboint.2020.106604
15. Sun B, He M, Wang Y, Gui W, Yang C, Zhu Q (2018) A data-driven optimal control approach for solution purification process. J Process Control 68:171–185. https://doi.org/10.1016/j.jprocont.2018.06.005
16. Frieha MS, Gammar SM, Lahmadi A, Andrey L (2021) Reinforcement and deep reinforcement learning for wireless Internet of Things: a survey. Comput Commun 178:98–113. https://doi.org/10.1016/j.comcom.2021.07.014
17. Vihauros ZJ, Jakab R (2021) Reinforcement learning for statistical process control in manufacturing. Measurement. https://doi.org/10.1016/j.measurement.2021.109616
18. Nian R, Liu J, Huang B (2020) A review on reinforcement learning: introduction and applications in industrial process control. Comput Chem Eng. https://doi.org/10.1016/j.compchemeng.2020.106886
19. Du Y, Zandi H, Kotlevska O, Kurte K, Munk J, Amasyali K, Makee E, Li F (2021) Intelligent multi-zone residential HVAC control strategy based on deep reinforcement learning. Appl Energy. https://doi.org/10.1016/j.apenergy.2020.116111
20. Gu S, Ethan H, Timothy L, Sergey L (2017) Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), pp 3389–3396
21. Liu L, Chen E, Gao Z, Wang Y (2019) Research on motion planning of seven degree of freedom manipulator based on DDPG. In Wang K., Wang Y., Strandhagen J., Yu T. (eds) Advanced Manufacturing and Automation VIII. IWAMA 2018. Lecture Notes in Electrical Engineering, vol 484. Springer, Singapore. https://doi.org/10.1007/978-981-13-2375-1_44
22. Qiu C, Hu Y, Chen Y, Zeng B (2019) Deep deterministic policy gradient (DDPG)-based energy harvesting wireless communications. IEEE Internet Things J 6(5):8577–8588. https://doi.org/10.1109/jiot.2019.2921159
23. Wang Y, Shen H, Duan D (2017) On stabilization of quantized sampled-data neural-network-based control systems. IEEE Transactions on Cybernetics 47(10):3124–3135. https://doi.org/10.1109/tcyb.2016.2581220
24. Qi Z, Peng S, Honghai L, Shengyuan X (2012) Neural-network-based decentralized adaptive output-feedback control for large-scale stochastic nonlinear systems. IEEE Trans Syst Man Cybern Part B (Cybern) 42(6):1608–1619. https://doi.org/10.1109/tsmc.2012.2196432
25. Buşoniu L, de Bruin T, Tolić D, Koher J, Palunko I (2018) Reinforcement learning for control: performance, stability, and deep approximators. Annu Rev Control 46:8–28. https://doi.org/10.1016/j.arcontrol.2018.09.005
26. Gao G, Li J, Wen Y (2020) DeepComfort: energy-efficient thermal comfort control in buildings via reinforcement learning. IEEE Internet Things J 7(9):8472–8484. https://doi.org/10.1109/jiot.2020.2992117
27. Ma Y, Zhu W, Benton MG, Romagnoli J (2019) Continuous control of a polymerization system with deep reinforcement learning. J Process Control 75:40–47. https://doi.org/10.1016/j.jprocont.2018.11.004
28. Siraskar R (2021) Reinforcement learning for control of valves. Mach Learn Appl. https://doi.org/10.1016/j.mlwa.2021.100030
29. Spielberg S, Gopaluni RB, Loewen PD (2017) Deep reinforcement learning approaches for process control. In 2017 6th International Symposium on Advanced Control of Industrial Processes (AdCONIP), pp. 28–31
30. Mnih V, Kavukcuoglu K, Silver D, Graves A, Antonoglou I, Wierstra D, Riedmiller M (2013) Playing atari with deep reinforcement learning. arXiv:1312.5602
31. Lillicrap TP, Hunt JJ, Pritzel A, Heess N, Erez T, Tassa Y, Silver D, Wierstra D (2015) Continuous control with deep reinforcement learning. arXiv:1509.02971
32. Sutton RS, McAllester D, Singh S, Mansour Y (2000) Policy gradient methods for reinforcement learning with function approximation. Adv Neural Inf Process Syst 1057–1063

33. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G, Petersen S, Beattie C, Sadik A, Antonoglou I, King H, Kumaran D, Wierstra D, Legg S, Hassabis D (2015) Human-level control through deep reinforcement learning. Nature 518(7540):529–533. https://doi.org/10.1038/nature14236

34. Fang X, Han Y, Wang J, Zhao Q (2019) A cognitive control approach for microgrid performance optimization in unstable wireless communication. Neurocomputing 355:168–182. https://doi.org/10.1016/j.neucom.2019.04.048

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.