Article

Reinforcement Learning Based Speed Control with Creep Rate Constraints for Autonomous Driving of Mining Electric Locomotives

Ying Li , Zhencai Zhu * , Xiaoqiang Li

Posted Date: 23 April 2024

doi: 10.20944/preprints202404.1512.v1

Keywords: autonomous driving; creep rate; mining electric locomotive; reinforcement learning; 15 speed control

Article

Reinforcement Learning Based Speed Control with Creep Rate Constraints for Autonomous Driving of Mining Electric Locomotives

Ying Li [1], Zhencai Zhu [1,*] and Xiaoqiang Li [2]

[1] School of Mechanical and Electrical Engineering, China University of Mining and Technology, Xuzhou, China 221116; lycumt@cumt.edu.cn (Y.L.)

[2] School of Electrical Engineering;, China University of Mining and Technology, Xuzhou, China 221116; xiaoqiangli@cumt.edu.cn (X.L.)

* Correspondence: zhuzhencaijs@163.com

Abstract: The working environment of mining electric locomotives is wet and muddy coal mine roadway. There may be idling or slipping between the wheels and rails of mining electric locomotives due to low friction between the wheel and rail and insufficient utilization of creep rate. Therefore, it is necessary to control the creep rate within a reasonable range. In this paper, the autonomous control algorithm for mining electric locomotives based on improved ε-greedy is theoretically proven to be convergent and effective firstly. Secondly, after analyzing the contact state between the wheel and rail under wet and slippery road conditions, it was concluded that the value of creep rate is an important factor affecting the autonomous driving of mining electric locomotives. Therefore, the autonomous control method for mining electric locomotives based on creep control is proposed in this paper. Finally, the effectiveness of the proposed method was verified through simulation. The problem of wheel slipping and idling caused by insufficient friction of mining electric locomotives in coal mining environments is effectively suppressed. Autonomous operation of vehicles with optimal driving efficiency can be achieved through quantitative control and utilization of the creep rate between wheels and rails.

Keywords: autonomous driving; creep rate; mining electric locomotive; reinforcement learning; speed control

1. Introduction

Mining electric locomotives have the transportation function of materials, equipment, and people in roadway. Safe driving of mining electric locomotives is crucial. However, the method of underground mining is often used in the Chinese coal mining industry [1]. In deep underground confined spaces, there are unfavorable conditions for driving in coal mine roadway, such as slippery, muddy, dusty, foggy, and complex human behavior, which lead to frequent accidents. Reducing personnel participation and ensuring the safe operation of production and transportation equipment are necessary to ensure the safety production of coal mines. At present, the intelligent development of coal mine equipment has become an inevitable trend [2]. Unmanned transportation of coal mine equipment can fundamentally solve the problem of personnel participation and reduce casualties in the event of inevitable accidents. We have conducted relevant research on autonomous mining electric locomotives [3]. The mining electric locomotive has achieved functions based on RL and improved ε-greedy such as autonomous and efficient operation on speed limited sections, maintaining a safe distance from vehicle in front, and avoiding obstacles.

However, problems such as slipping of autonomous vehicles and wheel idling caused by slippery roadway in deep mines have not been addressed in a targeted manner. Manually driven mining electric locomotives rely on the driver's experience to sprinkle sand to increase wheel rail friction. But this method is a remedial measure taken when slipping/idling occurs during driving. And there is no quantitative evaluation standard for actions judged by human subjectivity. Moreover, actions based on human subjective judgment cannot be used as a quantifiable evaluation criterion applicable to machine autonomous decision-making. In the autonomous driving control process of mining electric locomotives, it is necessary to make decisions that can actively control the interaction between wheels and rails to prevent slipping, idling, and other phenomena caused by external environment.

For rail vehicles, the destruction of the adhesion between the wheel and rail is the basic reason of wheel slip/idling [4]. One of the reasons for the traction and braking forces of wheelsets is the presence of contact and friction between the wheel and rail [5]. This kind of wheel rail interaction force is not pure friction, but a phenomenon called creep caused by deformation after wheel rail contact [6]. Reasonably utilizing the creep rate can improve the efficiency and safety of vehicle operation [7,8]. Nowadays, to prevent train wheel slipping or idling, scholars have conducted research on algorithms related to adhesion control. According to different control methods, adhesion control algorithms are generally divided into re-adhesion control and optimized adhesion control [9]. Re-adhesion control is a method that can quickly adjust the motor torque to achieve balance with the current adhesion conditions, and avoid wheel idling when determining whether the locomotive's wheels are idling/slipping. The research on this method is relatively mature [10–13], but it belongs to passive adhesion control method, with low adhesion utilization rate and long algorithm response time. The purpose of optimizing adhesion control is to achieve the peak point of the optimal adhesion utilization, which is of great significance for the operation control of trains under wet, muddy and emergency braking conditions. Mehmet Ali Çimen et al. [14] analyzed the input and output phase shift dynamic characteristics of the traction system. They also proposed an adaptive control method that effectively controls the adhesion utilization rate of the traction system. Song Wang et al. [15] proposed an adhesion control method based on optimal torque search for high-speed trains, which can achieve stable operation of trains in the optimal adhesion state under changes of track surface and high-speed driving conditions, effectively reducing the idle rate of the wheel and improving train adhesion utilization. Shuai Zhang et al. [16] proposed a sliding mode control method, which used recursive least squares method based on enhanced forgetting factor to solve the problem of wheel anti lock on heavy-duty trains. This algorithm can obtain the optimal creep rate and construct a PI closed-loop observer to estimate the unmeasured adhesion torque, enabling the locomotive to adjust to the optimal creep rate when the contact of wheel and rail changes. Although the optimized adhesion control algorithm can suppress the sliding and idling of the wheel, the maximum adhesion point in the contact area of wheel and rail is at the junction of creep and sliding, making it difficult to ensure that the contact state of wheel and rail is always creep during the actual control process. Therefore, this method still belongs to passive adhesion control. In addition, there are certain differences in the utilization of adhesion capacity among the axles of the train. The existing methods have insufficient control accuracy, as most of the adhesive characteristic curves relied on are empirical formulas obtained through experiments.

Therefore, this paper proposes a creep control method for mining electric locomotives based on RL. This method converts the impact of complex driving environments on mining electric locomotives into reward feedback values obtained by the electric locomotives from the environment. The impact of creep conditions that are difficult to quantify and evaluate on vehicles has been particularly considered in this method. The optimal range of creep rate is tried to achieve after training the mining electric locomotive to adjust the driving torque of the wheelset in this method. To achieve maximum driving efficiency under safe operating conditions is the goal of this method.

The overall structure of this paper is shown in Figure 1. Firstly, the autonomous control method of mining electric locomotives based on improved ε-greedy is analyzed from the theoretical perspective. By changing environmental variables and replicating algorithms, the important factor affecting the operation of mining electric locomotive on wet and slippery tracks is identified. Secondly, a mining electric locomotives autonomous control algorithm based on creep control is proposed, which can effectively improve the safety and reliability of autonomous driving of mining electric locomotives. Finally, based on the three-dimensional dynamic model of mining electric locomotives modeled using Simpack, the feasibility of the proposed algorithm is verified using a multi software co-simulation platform.

Compared to [3], this paper identifies the creep rate, a key factor affecting the autonomous operation of mining electric locomotives, through theoretical verification of the algorithm and practical analysis of operating conditions. This paper also proposes a creep control method suitable for mining electric locomotives, which improves the accuracy of autonomous operation control for mining electric locomotives.

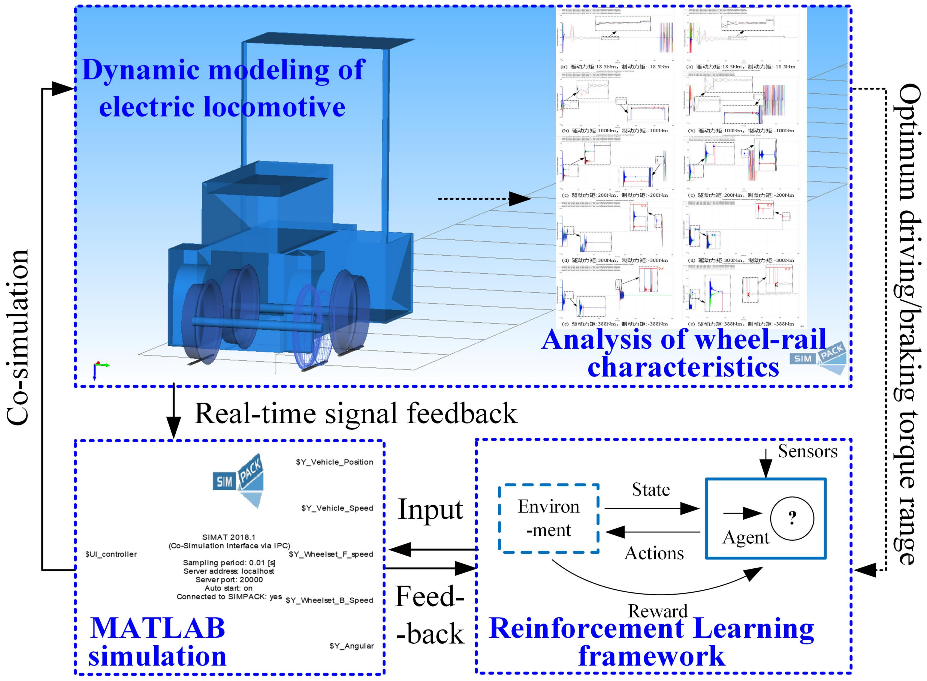

Dynamic modeling of electric locomotive
Analysis of wheel-rail characteristics
Co-simulation
Optimum driving/braking torque range
Real-time signal feedback
$Y_Vehicle_Position
$Y_Vehicle_Speed
$UI_controller
SIMAT 2018.1
(Co-Simulation Interface via IPC)
Sampling period: 0.01 [s]
Server address: localhost
Server port: 20000
Auto start: on
Connected to SIMPACK: yes
$Y_Wheelset_F_speed
$Y_Wheelset_B_Speed
$Y_Angular
MATLAB simulation
Input
Feed-
-back
Sensors
Environ
-ment
State
Actions
Agent
?
Reward
Reinforcement Learning framework

Figure 1. The framework of creep control method for autonomous driving of the mining electric locomotive.

2. Analysis of Autonomous Control Algorithm for Mining Electric Locomotives

In this section, an important factor is found through theoretical analysis and simulation, which was not particularly considered in the previous algorithm design but affects the autonomous driving of mining electric locomotives. First, in reproducing the autonomous control algorithm of mining electric locomotives based on RL, we found that changing the friction between the wheels and rails can lead to poor control performance. Second, this algorithm has been theoretically proven to be reasonable and convergent. Mining electric locomotives can stably achieve autonomous driving under the control of this algorithm. Third, the relationship between creep rate and adhesion coefficient of rail vehicles is analyzed. It is necessary to take the creep rate between the wheel and rail into the autonomous control algorithm. Finally, we conclude that the improved control objective is to achieve the optimal control range of creep rate.

2.1. Introducing Problems

In reference [3], RL was adopted to solve the autonomous control problem of mining electric locomotives, and the ε-greedy strategy was improved to balance the relationship of exploration and exploitation better. This control method is reproduced in this paper on the condition that the friction of wheel and rail is adjusted from 0.4 to 0.3. Under this working condition, the autonomous control state curve of the mining electric locomotive shown in Figure 2 is obtained. It can be seen that mining electric locomotive operates safely and efficiently on speed limited sections, and can maintain a safe distance from obstacles when using the autonomous control method based on the improved ε-greedy strategy. However, it was found that mining electric locomotive is unable to control the acceleration duration correctly when accelerating and reaching the maximum speed limit, and the running speed of the electric locomotive would slightly exceed the maximum speed limit. Moreover, the mining electric locomotive cannot control the deceleration duration to reach the maximum speed on the next speed limit section when decelerating.

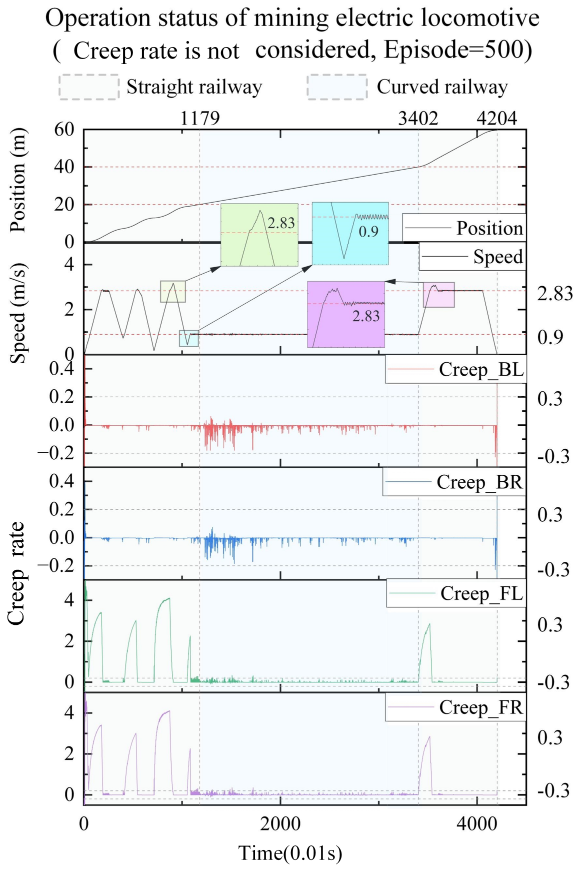


Figure 2. The autonomous control state curve of the mining electric locomotive without creep control.

2.2. Theoretical Analysis

In order to find out the reasons for the problem above, we first verify whether the algorithm structure is designed reasonably from the perspective of algorithm theory. If the structural design is reasonable, the results of the algorithm will tend to converge with the training process of the agent. This section will demonstrate the convergence of the algorithm from two aspects: RL and improved ε-greedy strategy.

2.2.1. Q-Learning

RL is a method for studying how an agent maximizes its reward in a complex and uncertain environment [17,18]. Figure 3 shows the process of RL. The agent interacts with the environment, obtains the current state S_t and reward R_t at moment t, and takes action A_t based on certain strategies. After the agent takes action A_t, the environment obtains the latest state S_{t+1} and reward R_{t+1} at moment t+1, and passes them to the agent. This interaction process can be represented by the Markov Decision Process (MDP).

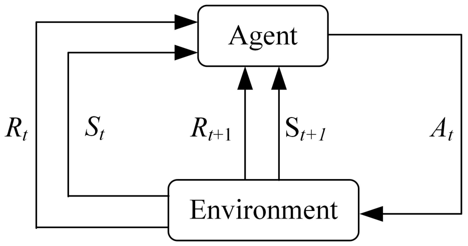
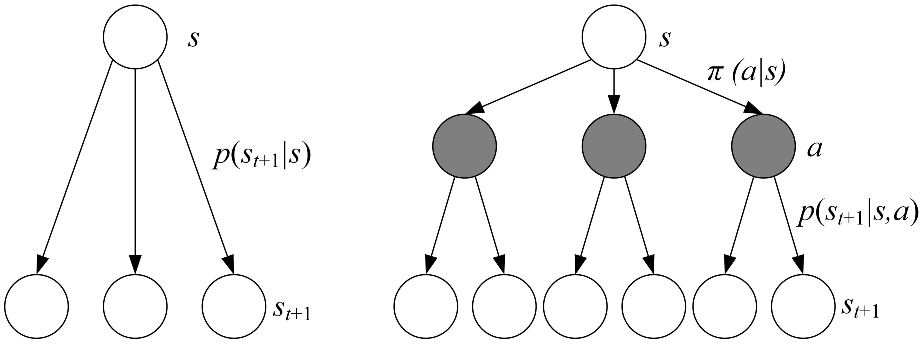


Figure 3. Reinforcement Learning.

The MDP adds a decision layer to the commonly used Markov Process / Markov Reward Process, which is the action a shown in the Figure 4. This means that when the agent is in the state S_t at time t, it must first decide on a specific action a to take in order to reach the middle layer, which is the black node in the diagram. After reaching the black node, the state of the agent at the moment t+1 also depends on the probability distribution.



(a) Markov Decision Process

(b) Markov process / Markov reward process

Figure 4. The Markov Decision Process and Markov Process / Markov Reward Process.

RL, as a trial and error learning method, continuously repeats the above interaction process between agents and the environment to find a mapping that can maximize the cumulative sum G_t of benefits R_t over time.

$$G_t = \sum_{i=0}^{T-i-1} \gamma^i R_{t+i+1}. \tag{1}$$

where, T is the total time, γ is the discount factor.

However, the sum of cumulative returns, also known as value return G_t, is not easily obtained. To solve this problem, researchers propose two value functions for estimating the sum of cumulative returns: the state value function $V_\pi(s)$ and the action value function $Q_\pi(s,a)$. The state value function $V_\pi(s)$ is the expected value of the sum of the reward functions under π strategy and s state. The action value function $Q_\pi(s,a)$ represents the expected value of the sum of the benefit functions of action a under the π strategy and s state.

$$V_\pi(s) = \mathbb{E}_\pi[G_t|S_t = s] = \sum_{a\in\mathcal{A}} \pi(a|s)\cdot Q_\pi(s,a). \tag{2}$$

$$Q_\pi(s,a) = \mathrm{E}_\pi[G_t|S_t = s, A_t = a] = \sum_{r,s_{t+1}} p(s_{t+1}, r|s,a)\cdot[R + \gamma V_\pi(s_{t+1})]. \tag{3}$$

where, $\mathcal{A}$ is the set of action a.

According to the value function, the Bellman equation is derived as follows:

$$V(s) = R(s) + \gamma \sum_{s_{t+1}\in\mathcal{S}} p(s_{t+1}|s)\cdot V(s_{t+1}). \tag{4}$$

where, $\mathcal{S}$ is the set of state s.

The Bellman equation for the Q-function is as follows:

$$Q(s,a) = R(s,a) + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}). \tag{5}$$

We adopt Theorem-Cauchy's convergence Test to verify the convergence of RL we design to apply in the field of autonomous control of mining electric locomotives.

Theorem-Cauchy's convergence Test: The necessary and sufficient condition for the convergence of a sequence is that for any positive number ε, there must be a positive integer N that satisfies the following equation under the condition of $n > N$ and $m > N$:

$$|x_n - x_m| < \varepsilon. \tag{6}$$

We refer to $\{x_n\}$ that satisfies this condition as a Cauchy sequence. So the theorem can be expressed as: the necessary and sufficient condition for the convergence of sequence $\{x_n\}$ is that the sequence is a Cauchy sequence.

Define H as the contraction operator, the following formula will be easily obtained:

$$HQ(s,a) = R(s,a) + \gamma \mathrm{E}_{s_{t+1} \sim p(\cdot|s,a)} [\max_{a_{t+1}} Q(s_{t+1}, a_{t+1})]. \tag{7}$$

Assuming that the optimal value function Q^* is a fixed point of H, it means:

$$Q^* = HQ^*. \tag{8}$$

That is, Q^* remains at its original value after any multiplication with the H operator.

$$(Hq)(s,a) = \sum_{s_{t+1} \in \mathcal{S}} p_a(s, s_{t+1}) [r(s, a, s_{t+1}) + \gamma \max_{a_{t+1} \in \mathcal{A}} q(s_{t+1}, a_{t+1})]. \tag{9}$$

Next, the convergence of the Q function is proved:

$$\begin{aligned} ||Hq_1 - Hq_2||_\infty &= \max_{s,a} \gamma | \sum_{s_{t+1} \in \mathcal{S}} p_a(s, s_{t+1}) [\max_{a_{t+1} \in \mathcal{A}} q_1(s_{t+1}, a_{t+1}) - \max_{a_{t+1} \in \mathcal{A}} q_2(s_{t+1}, a_{t+1})]| \\ &\leq \max_{s,a} \gamma \sum_{s_{t+1} \in \mathcal{S}} p_a(s, s_{t+1}) | \max_{a_{t+1} \in \mathcal{A}} q_1(s_{t+1}, a_{t+1}) - \max_{a_{t+1} \in \mathcal{A}} q_2(s_{t+1}, a_{t+1})| \\ &\leq \max_{s,a} \gamma \sum_{s_{t+1} \in \mathcal{S}} p_a(s, s_{t+1}) \max_{a_{t+1} \in \mathcal{A}} |q_1(s_{t+1}, a_{t+1}) - q_2(s_{t+1}, a_{t+1})| \\ &= \gamma ||q_1 - q_2||_\infty. \end{aligned} \tag{10}$$

According to the **Theorem-Cauchy's convergence Test**, any q_1 function can ultimately converge to Q^* when q_2 is the optimal function of Q^*.

2.2.2. Improved ε-Greedy

In reference [3], in order to balance the relationship between exploration and exploitation, the traditional ε-greedy algorithm was improved by changing the value of ε in formula 11 to the value of ε in formula 12.

$$\varepsilon_1 = (\varepsilon_{\text{initial}} - \varepsilon_{\text{final}}) \cdot (1 - episode/max\_episodes) \tag{11}$$

$$\varepsilon_2 = \{ \begin{array}{l} (\varepsilon_{\text{initial}} - \varepsilon_{\text{final}}) \cdot (0.5 + \sqrt{0.25 - (episode/max\_episodes)^2}),\ (episode/max\_episodes) \in [0, 0.5] \\ (\varepsilon_{\text{initial}} - \varepsilon_{\text{final}}) \cdot (0.5 - \sqrt{0.25 - (1 - episode/max\_episodes)^2}),\ (episode/max\_episodes) \in (0.5, 1] \end{array} \tag{12}$$

In order to evaluate the balance effect of exploration and exploitation, regret l_t is defined to represent the average possible loss at each step:

$$l_t = \mathbb{E}[V^* - Q(a_t)] \tag{13}$$

where,$V^* = Q^*(a^*) = \max_{a \in \mathcal{A}} Q^*(a)$ represents the expected return value corresponding to the optimal action.

Total regret L_t is defined to represent the total loss:

$$L_t = \mathbb{E}[\sum_{\tau=1}^{t} V^* - Q(a_\tau)] = \sum_{a \in \mathcal{A}} \mathbb{E}[N_t(a)](V^* - Q(a)) = \sum_{a \in \mathcal{A}} \mathbb{E}[N_t(a)]\Delta_a \tag{14}$$

where, $\Delta_a = (V^* - Q(a))$ represents the gap between the value of the action and the optimal action, and $N_t(a)$ represents the number of times that the action a has been selected. The goal of maximizing cumulative returns is actually equivalent to minimizing total regret, and an excellent algorithm is hoped to reduce the number of choices for actions with large gaps. For such an evaluation system, the most crucial issue is that the gap in practical problems is difficult to obtain, as V^* is unknown and $Q^*(a)$ also needs to be estimated. However, this evaluation system can still have guiding significance for the evaluation of strategies. It is usually assumed that Δ_a is a known constant value when discussing.

If the total regret L_t increases linearly with the increase of iteration, it indicates that the probability of selecting each action in the algorithm does not change at each time interval. That is, the regret l_t at each step does not change, and the information obtained from exploration is not better utilized. A strategy that can better balance exploration and exploitation should have a sublinear total regret. As the iteration progresses, the regret l_t for each time interval gradually decreases, and the selection of algorithms will gradually abandon those that are likely to achieve lower returns.

Lai and Robbins has proven that for all total regrets that may become an optimal strategy, a asymptotic lower bound in the form of logarithmic growth is required [19–21]:

$$\lim_{t \to \infty} L_t \leq 8 \ln t \sum_{a|\Delta_a} \frac{\Delta_a}{KL(R^a||R^{a^*})} \tag{15}$$

The strategy of ε-greedy strategy can ensure that all actions are sampled infinitely as the number of iterations increases. Thereby the convergence of $Q(a)$ is ensured. This means that the probability of selecting the optimal action can converge to greater than 1 - ε, which is close to certainty. However, the probability of selecting the optimal action is only approaching a value, and its effectiveness cannot be guaranteed in practice. Obviously, the probability of each choice of action by ε-greedy is more than $\frac{\varepsilon}{\mathcal{A}}$. So the regret l_t is more than $\frac{\varepsilon}{\mathcal{A}} \sum_{a \in \mathcal{A}} \Delta_a$, which is a linear total regret. However, although ε-greedy is not the optimal method in theory, Kuleshov and Precup [22] have demonstrated through extensive experimental data that ε-greedy can often achieve better results than some complex methods in practice.

For the improved ε-greedy, the selection of actions follows the following rules:

$$a_t = \begin{cases} \underset{a}{\text{argmax}} Q(s_{t+1}, a), for \quad probability \quad 1 - \varepsilon_t \\ random \quad from \quad \mathcal{A}, for \quad probability \quad \varepsilon_t \end{cases} \tag{16}$$

We can obtain the probability distribution of whether an action is the optimal action:

$$\pi(a|s) = \begin{cases} 1-\varepsilon+\frac{\varepsilon}{|\mathcal{A}(s)|},\ if\ a = \underset{a}{\operatorname{argmax}} Q(s,a) \\ \frac{\varepsilon}{|\mathcal{A}(s)|},\ if\ a \neq \underset{a}{\operatorname{argmax}} Q(s,a) \end{cases} \tag{17}$$

Set $episode = t$, $max\_episodes = T$. We will write the value of ε in the improved ε-greedy as follows:

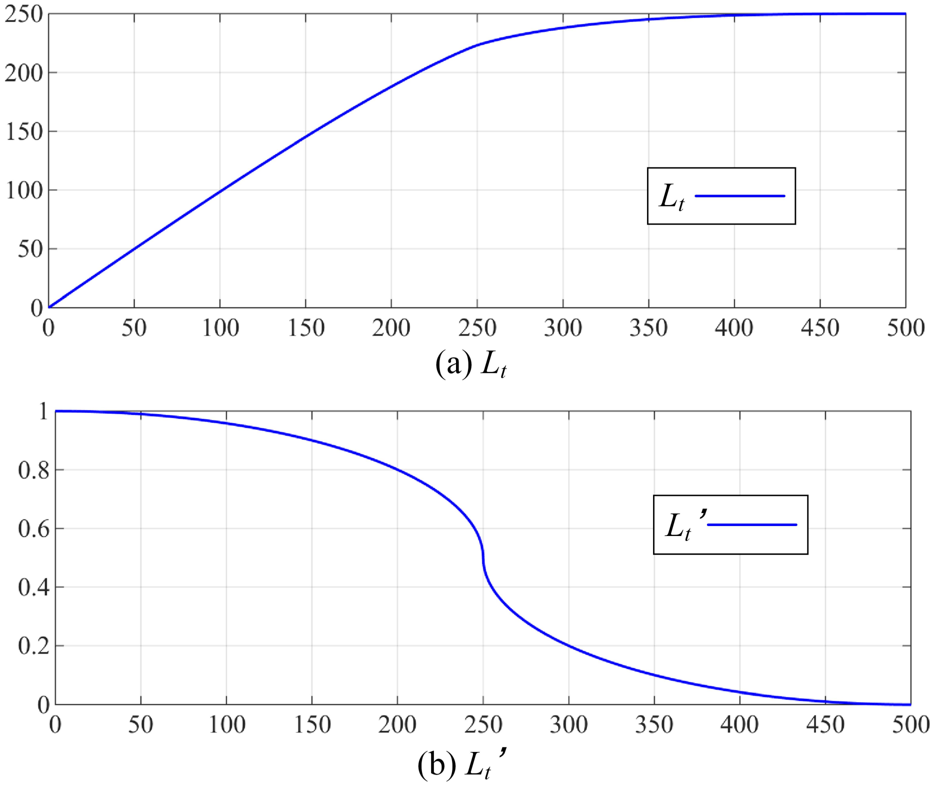
$$\varepsilon_t = \begin{cases} (0.5+\sqrt{0.25-(t/T)^2}),\ (t/T) \in [0, 0.5] \\ (0.5-\sqrt{0.25-(1-t/T)^2}),\ (t/T) \in (0.5, 1] \end{cases} \tag{18}$$

To demonstrate the convergence of the improved ε-greedy strategy, we calculate the regret of the strategy:

$$l_t = \pi(a|s)\sum_{a\in\mathcal{A}}\Delta_a = \begin{cases} (1-\varepsilon+\frac{\varepsilon}{|\mathcal{A}(s)|})\sum\limits_{a\in\mathcal{A}}\Delta_a,\ if\ a = \underset{a}{\operatorname{argmax}} Q(s,a) \\ \frac{\varepsilon}{|\mathcal{A}(s)|}\sum\limits_{a\in\mathcal{A}}\Delta_a,\ if\ a \neq \underset{a}{\operatorname{argmax}} Q(s,a) \end{cases} \tag{19}$$

$$\begin{aligned} L_t &\approx \int\limits_t l_t \mathrm{d}t = \int\limits_t \frac{\varepsilon}{|\mathcal{A}|}\sum_{a\in\mathcal{A}}\Delta_a \mathrm{d}t \\ &= \begin{cases} \int\limits_t \frac{\sum\limits_{a\in\mathcal{A}}\Delta_a}{|\mathcal{A}|}(0.5+\sqrt{0.25-(t/T)^2})\mathrm{d}t,\ (t/T) \in [0, 0.5] \\ \int\limits_t \frac{\sum\limits_{a\in\mathcal{A}}\Delta_a}{|\mathcal{A}|}(0.5-\sqrt{0.25-(1-t/T)^2})\mathrm{d}t,\ (t/T) \in (0.5, 1] \end{cases} \\ &\rightarrow \begin{cases} \frac{\sum\limits_{a\in\mathcal{A}}\Delta_a}{|\mathcal{A}|}(\frac{t}{2}+\frac{1}{8}T\arcsin(\frac{2t}{T})+\frac{t}{4}\sqrt{1-4(\frac{t}{T})^2}),\ (t/T) \in [0, 0.5] \\ \frac{\sum\limits_{a\in\mathcal{A}}\Delta_a}{|\mathcal{A}|}(\frac{t}{2}+\frac{1}{8}T\arcsin(2(1-\frac{t}{T}))+\frac{T}{4}(1-\frac{t}{T})\sqrt{1-4(1-\frac{t}{T})^2}),\ (t/T) \in (0.5, 1] \end{cases} \end{aligned} \tag{20}$$

Plot the trends of L_{t} and its derivative L_{t}' as the number of iterations increases. In Figure 5, it can be seen that L_{t}' is 0 in the later stage of iteration, and the value of L_{t} no longer increases, indicating that the improved εgreedy tends to converge.



(a) L_t

(b) L_t'

Figure 5. The trend of total regret L_t and its derivative L_t'.

In summary, RL and improved ε-greedy strategy applied to autonomous control of mining electric locomotives is convergent. The autonomous control method for mining electric locomotives based on the improved ε-greedy strategy is feasible.

2.3. Adhesion and Creep

After the analysis, we will focus on the constraints considered in the algorithm design process. When designing the algorithm, conditions were taken into consideration, such as speed limits, road obstacles, and safe following of mining electric locomotives . However, the road surface of the coal mine roadway is very slippery and muddy, and the contact condition between the wheels and rails is also one of the factors that needs to be carefully considered.

Two conditions must be met simultaneously to enable rail vehicles to run along the track [23]. One is that the rotating torque is applied to the moving wheel, and the other is that the contact between the moving wheel and the steel rail has frictional effect. As shown in Figure 6, with the wheel as the force analysis object, the moving wheel is in contact with the track due to the vertical force P applied by the vehicle body, and generates a motion trend under the torque M transmitted by the transmission device [24]. Assuming that there is static friction between the wheel and rail, the force F' generated by the moving wheel on the track is equal in magnitude to the force F generated by the track on the moving wheel, but in opposite directions. For the mining electric locomotive, rim traction force F is the driving traction force. But the wheels of rail vehicles mostly have conical tread, which can cause elastic-plastic deformation when the wheels come into contact with the rail. The mining electric locomotive is subjected to shock and vibration during operation. When the wheel set rolls on the rail, it is accompanied by longitudinal and transverse sliding. There is no pure static friction state between the wheel and rail, but rather 'slight movement in stillness' and 'slight sliding in rolling'. This phenomenon is creep, also known as adhesion between wheel and rail. The maximum longitudinal horizontal force in the adhesive state is the adhesive force, and the ratio of adhesive force F_μ to vertical load P is the adhesive coefficient μ. According to Hertz elasticity contact theory [25], the contact area between the wheel and rail is approximately elliptical. The area where the wheel and rail are relatively stationary and do not slide under the action of positive pressure is called the adhesion zone. And the area where slight sliding occurs is called the creep zone. As the driving torque M increases, the creep zone becomes larger and the adhesion zone decreases to 0, and the wheels are in a sliding state. Therefore, when the vehicle is in traction or braking conditions, if the traction or braking force is greater than the adhesion between the wheels and rails, the wheels will idling or slip. This phenomenon can cause damage to the wheel rail tread and even affect the safety of rail vehicles.

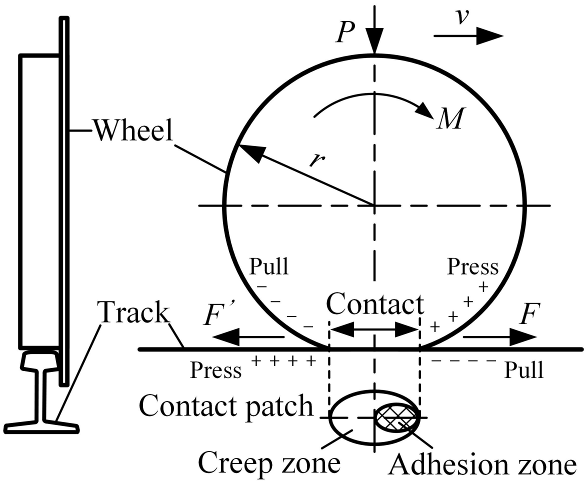


Figure 6. Schematic diagram of adhesion between wheels and rails.

According to the International Union of Railways (UIC) definition of creep rate, the creep rate is represented as follows:

$$\begin{cases} CR_{longitudinal} = \frac{(VW_{longitudinal} - VT_{longitudinal})_{At\ the\ contact\ patch}}{Nominal forwardvelocity} \\ CR_{lateral} = \frac{(Wheel\ lateral\ velocity - Track\ lateral\ velocity)_{At\ the\ contact\ patch}}{Nominal forwardvelocity} \\ CR_{spin} = \frac{(Wheel\ angular\ velocity - Track\ Wheelangularvelocity\ velocity)_{At\ the\ contact\ patch}}{Nominal forwardvelocity} \end{cases} \tag{21}$$

Based on the operating results when the friction between the wheel and rail is 0.3 (Figure 2), it can be observed that there is a sudden change in the interaction of wheel and rail during vehicle acceleration, and the creep rate value is abnormal.

2.4. Control Objectives

As the main auxiliary transportation tool in the coal mine roadway, the mining electric locomotive is greatly affected by the humid, muddy and other harsh environment during driving. When there are pedestrians or obstacles, the tram can't bypass to avoid them, but only take emergency braking. Wet and slippery road surface may lead to insufficient wheel/rail adhesion of mining electric locomotive. When the vehicle is in traction condition and the traction force is greater than the wheel/rail adhesion, the wheel will idling. When the vehicle is in braking condition and the braking force is greater than the adhesion, the wheels will slip. The occurrence of these situation will cause the wheel tread and rail surface to be scratched, which will seriously affect the safety and stability of the vehicle. Therefore, the most critical control factor for the safe running of mining electric locomotive on the track is the value of the driving/braking torque applied on the axle. In this way, we can directly control the driving acceleration and speed of the vehicle by controlling the torque to avoid driving accidents such as vehicle braking failure and vehicle rollover.

As shown by the adhesion characteristic curves under dry and wet rail surface conditions in Figure 7, when the creep rate is within the range of -10% to 10%, the absolute value of the adhesion coefficient between the wheel and rail rapidly increases under braking/traction conditions [26]. When the creep rate is within the range of -30% to -10% and 10% to 30%, the absolute value of the adhesion coefficient between the wheel and rail reaches its optimal value under braking/traction conditions. When the creep rate is around ±20%, the wheel rail adhesion coefficient will reach its peak, and at this time, the vehicle can obtain maximum ground braking/traction force, which can minimize the braking distance. Therefore, controlling the creep rate within the range of ±10% to ±30% is an ideal control goal.

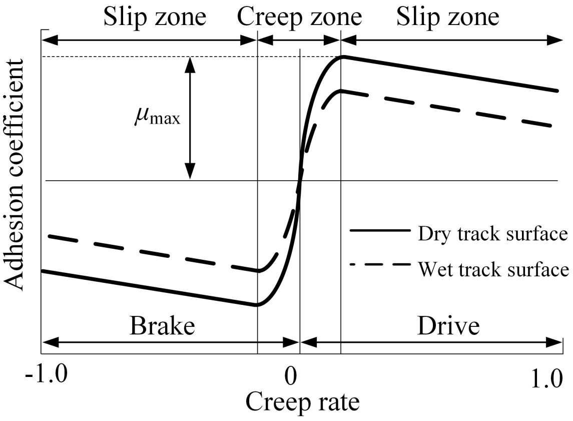


Figure 7. The adhesion characteristic curves under dry and wet rail surface conditions.

3. Creep Controller Model

In this section, an RL speed control method for mining electric locomotives based on the optimal creep rate is proposed. For mining electric locomotives, there are many turns and narrow sections in coal mine roadway, and there are many emergencies during the operation of the locomotives. This method can fully utilize the advantage of RL that does not rely on models. The impact of complex mine environments on vehicle driving processes, especially the impact of slippery wheel/rail conditions that are difficult to quantify and evaluate on vehicle safety and autonomous operation, is converted into reward feedback obtained by vehicles from the environment. This method simplifies the modeling process of vehicle driving conditions and improves the efficiency of intelligent algorithm calculation. There are three major elements in RL, including state, reward, and strategy for selecting actions. When using RL method for creep control of the mining electric locomotive, the mining electric locomotive is considered as an intelligent agent, and the algorithm components are designed as follows.

3.1. State and Action

According to the starting or stopping of the mining electric locomotive, whether the driving section is a curve or a straight road, whether there are obstacles, whether the operating speed reaches the speed limit of the corresponding section, and whether the electric locomotive reaches the destination, the state of the RL creep control algorithm for the mine electric locomotive is set to 16 as shown in the Table 1. In addition, the torque applied on the axle has three forms: positive, negative, and zero, corresponding to the three actions of vehicle acceleration, deceleration, and uniform speed.

doi:10.20944/preprints202404.1512.v1

Table 1. State Design of the Mining Electric Locomotive Autonomous Control Based on RL

State	Meaning
begin	Electric locomotive start
to_the_end	Electric locomotive reaches the destination
obstacle_stop	Stop when encountering obstacles
c_within_obstacle	Close to the vehicle in front on curves
max_spd_c_no_obstacle	The maximum set speed is reached on curves
over_spd_c_no_obstacle	Overspeed on curves
below_spd_c_no_obstacle	The maximum set speed is not reached on curves
l_within_obstacle	Keep close to the obstacle in front when driving straight
near_to_the_end_brake	The speed is greater than 0 when approaching the terminal
near_to_the_end_drive	The speed is less than 0 when approaching the terminal
max_spd_l_to_c	The maximum speed allowed in the curve is reached when preparing to turn
over_spd_l_to_c	Overspeed when preparing to turn
below_spd_l_to_c	The maximum speed allowed in the curve is not reached when the vehicle is preparing to turn
max_spd_l_no_obstacle	The maximum set speed is reached on the straight track
over_spd_l_no_obstacle	Overspeed on the straight
below_spd_l_no_obstacle	The maximum set speed is not reached on the straight track

During the execution of the algorithm, the real-time position and speed of the mining electric locomotive are recorded every 0.01 seconds to determine the changes of electric locomotive state during the sampling interval. At the same time, the algorithm will also timely provide the actions that should be applied within the sampling time of 0.01s, ensuring the timeliness of control.

It is worth pointing out that JB/T 4091-2014 *'Technical Conditions for Explosion-proof Special Battery Electric Locomotives in Coal Mines'* stipulates that mining electric locomotives should pass through the curve radius at a 50% of hourly speed. For the CTY1.5/6 electric locomotive used in this case, the maximum speed of the vehicle is 2.83m/s (10.2km/h), and the hourly speed of the vehicle is 1.80m/s (6.5km/h). Therefore, when the mining electric locomotive passes a curve with a radius of curvature greater than 4m, the minimum speed of the vehicle is 0.90m/s.

3.2. Features of Reward Function

The appropriate definition of the reward function can directly determine whether the learning process of RL algorithms can efficiently and quickly achieve the desired training objectives. The main task of RL is to maximize rewards by selecting the best action at each sampling time. The ultimate goal of control algorithm for the electric locomotive is to achieve the optimal creep rate utilization in uncertain and harsh roadway environments, reduce the occurrence of slip and idling phenomena, and achieve the best operating efficiency under the premise of safe operation. Therefore, for this case, the reward function is directly related to the control effect of the mining electric locomotive's driving speed and creep rate under different states.

Whether the mining electric locomotive has encountered obstacles and reached the maximum speed of the corresponding road section is classified during the state setting. When setting the reward function, whether the vehicle is accelerating, decelerating, or running at the uniform speed in the current state should be considered. The vehicle changing at a faster speed to achieve the desired control purpose will present a better training effect. The following speed reward conditions is defined based on repeated experiments. The main way to measure the size of the reward value is to calculate the speed change that occurs during the unit sampling interval for the vehicle to achieve control objectives.

$$reward_{\text{speed}} = c_{\text{speed}} * (v_{\text{record}} - v_{\text{current}}) \tag{22}$$

where, c_{creep} represents the speed reward coefficient, which can be obtained through multiple experiments.

According to the control goal proposed in section 2.4 of the paper to control the creep rate within the range of ±10% to ±30%, the creep control reward conditions are set. We have comprehensively considered the creep rate of four wheels, and the evaluation criteria for the creep rate reward are set as follows.

When the absolute values of the creep rate of the left front wheel $|\xi_{\text{FL}}|$, right front wheel $|\xi_{\text{FR}}|$, left rear wheel $|\xi_{\text{BL}}|$, and right rear wheel $|\xi_{\text{BR}}|$ are all approximately equal to the optimal creep rate of 0.2 when taking two decimal places, the reward value can be set to:

$$reward_{\text{creep}} = c_{\text{creep\_max}} \tag{23}$$

In other cases, the reward value is set to:

$$reward_{\text{creep}} = \frac{c_{\text{creep\_max}}}{\sqrt{(|\xi_{\text{FL}}| - 0.2)^2 + (|\xi_{\text{FR}}| - 0.2)^2 + (|\xi_{\text{BL}}| - 0.2)^2 + (|\xi_{\text{BR}}| - 0.2)^2}} \tag{24}$$

where, $c_{\text{creep\_max}}$ and c_{creep} represent the reward coefficient for optimal creep rate and the reward coefficient for creep rate respectively, which can be obtained through multiple experiments.

We add the evaluation results of speed and creep with ratio of 1:1 to obtain the final reward function formula:

$$reward = reward_{\text{speed}} + reward_{\text{creep}} \tag{25}$$

3.3. Structure of the Creep Control Method

After setting up the intelligent agent, state, action, and reward function, a complete creep control framework for mining electric locomotives should be designed. As shown in the Figure 8, the mining electric locomotive serves as an intelligent agent to determine the current state S_t based on the conditions of the environment at the current moment t. Then the intelligent agent uses an improved ε-greedy method to select the actions to be executed, obtain the current action A_t, and apply it to the axle. Finally, the intelligent agent obtains the corresponding reward value R_{t+1} based on the current environment and state. This reward is used to update Q-table.

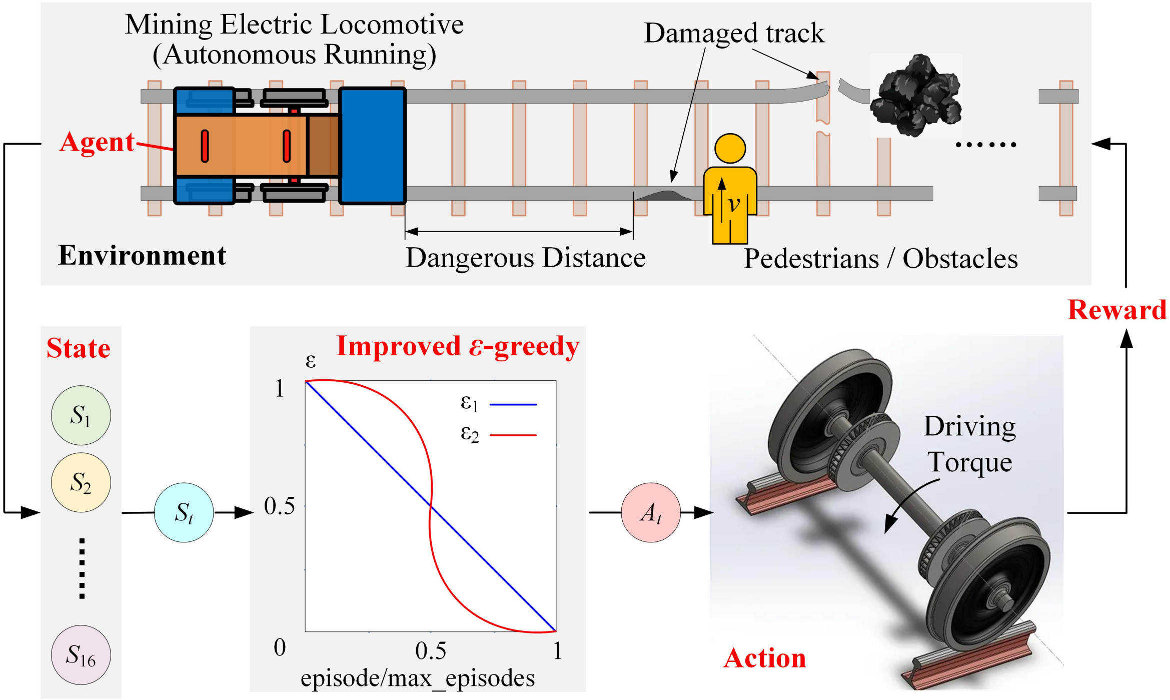


Figure 8. Details of the CTY1.5/6 creep control method.

4. Simulation Verification and Results

In this section, we varified the proposed RL based creep control method for autonomous operation of mining electric locomotives through simulation experiments. A dynamic model of the CTY1.5/6 mining electric locomotive is built through Simpack. And the algorithm framework based on the Python language environment is designed. MATLAB was used as the intermediate carrier for connecting the algorithm and model, mainly responsible for the collection of dynamic model data and the output of algorithm operation results. Therefore, in this paper, the validation of the algorithm will be carried out through a co-simulation platform built by Simpack, MATLAB, and Python.The simulation platform is used to varify the effectiveness of the creep control method for autonomous operation of mining electric locomotives, and compared it with traditional ε-greedy algorithms to highlight the advantages of the improved greedy algorithm in terms of learning efficiency.

4.1. Simulation Platform and Setup

The Windows 11 Professional 64 bit system is used as a simulation platform operating system, equipped with an Intel (R) Core (TM) i9-9900K CPU @ 3.60GHz CPU and NVIDIA GeForce RTX 2070 SUPER GPU. The composition software of the co-simulation platform is PyCharm Community Edition, MATLAB R2014b and Simpack 2018.1. The language environment of Python 2.7 and its equipped third-party libraries such as Matlabengine for Python R2014b and Openpyxl 2.6.4 provide the foundation for the operation of simulation platform algorithms.

We use the RL algorithm based on the improved ε-greedy mentioned above to train the creep control of the CTY1.5/6 mining electric locomotive. The number of iterations is 500, with a maximum runtime of 45s and a sampling interval of 0.01s. The learning rate of this algorithm is set to 0.2, and the Q-learning discount rate is 0.8. The initial and final values of ε are set to 0.01 and 1, respectively. The driving section of the mining electric locomotive is composed of a combination of 20m straight road-20m curved road-20m straight road. In general, the friction between the rail and the wheel is 0.2-0.4. Therefore, the friction coefficient on the track surface is set to 0.3.

4.2. Simulation Results

The reward values for creep control of the mining electric locomotive obtained through simulation using traditional ε-greedy and improved ε-greedy algorithms are shown in the Figure 9. It can be seen that under the condition of training 500 times, the agent obtains lager reward values in the first half of the training using traditional ε-greedy. The increasingly stable creep control reward value is obtained in the later stages of training during the use of improved ε-greedy algorithm for autonomous driving simulation in mining electric locomotives. This indicates that the improved ε-greedy algorithm we have designed can fully utilize the experience gained from exploration in the first half of the training, while the exploration of traditional ε-greedy in the first half of the training is clearly limited. And towards the end of the training stage, the reward value tends to stabilize to ensure the reliability of the training results, allowing the agent to repeatedly train and verify the optimal value.

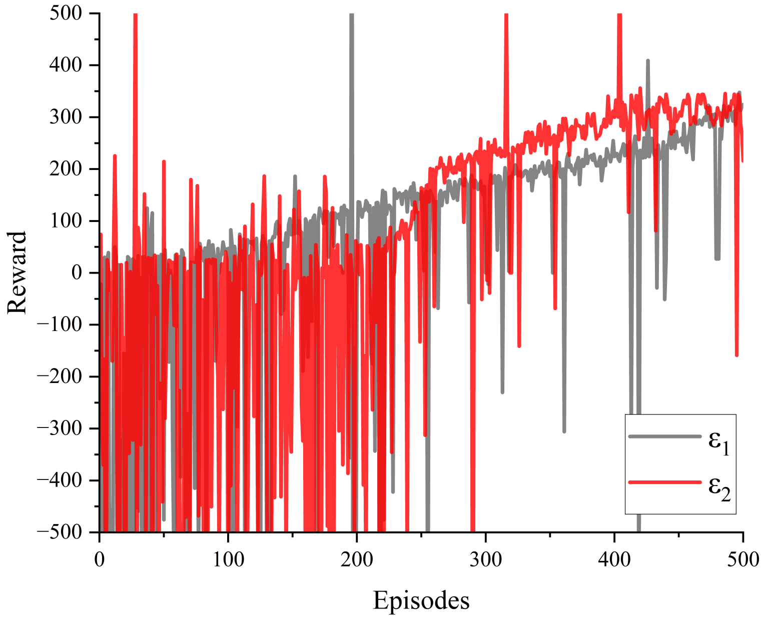


Figure 9. Results of reward with different values of ε.

We take the results every 125 iterations and plot the real-time position, speed, and creep rate of each wheel of the mining electric locomotive. From the 125th iteration to the 500th iteration (Figures 10 and 11), the mining electric locomotive can achieve their destination. However, the speed change is different. It can be easily observed that the mining electric locomotive cannot fully achieve the maximum speed limit for straight or curved roads during the initial training stage, and the vehicle cannot stop at its destination. As the training progresses, the mining electric locomotive acceleration increases and it can operate efficiently on speed limited sections, with accurate braking at the destination. On the first straight section of the road, the mining electric locomotive can maintain a safe distance when encountering obstacles.

In terms of controlling the creep rate, it can be seen that the creep rate fluctuates greatly during vehicle acceleration, which can easily lead to wheel spin. Meanwhile, when the vehicle is driving in a bend, there is also a fluctuation in creep rate, indicating the possibility of some wheels slipping at certain times during the turning process. After training the agent, it can be seen that as the number of iterations increases, the idle time during the vehicle acceleration process significantly decreases, almost eliminating the phenomenon of the vehicle slipping on curves. When the intelligent agent is trained 500 times, the creep rate of the vehicle during acceleration can be controlled within the range of 0.1 to 0.4. This indicates that the creep rate between the wheels and rails is well controlled during vehicle operation and is well utilized to drive wheel acceleration. This proves that the design of our learning algorithm is reasonable and effective, which can better utilize the creep rate and achieve the goal of safe and efficient creep control.

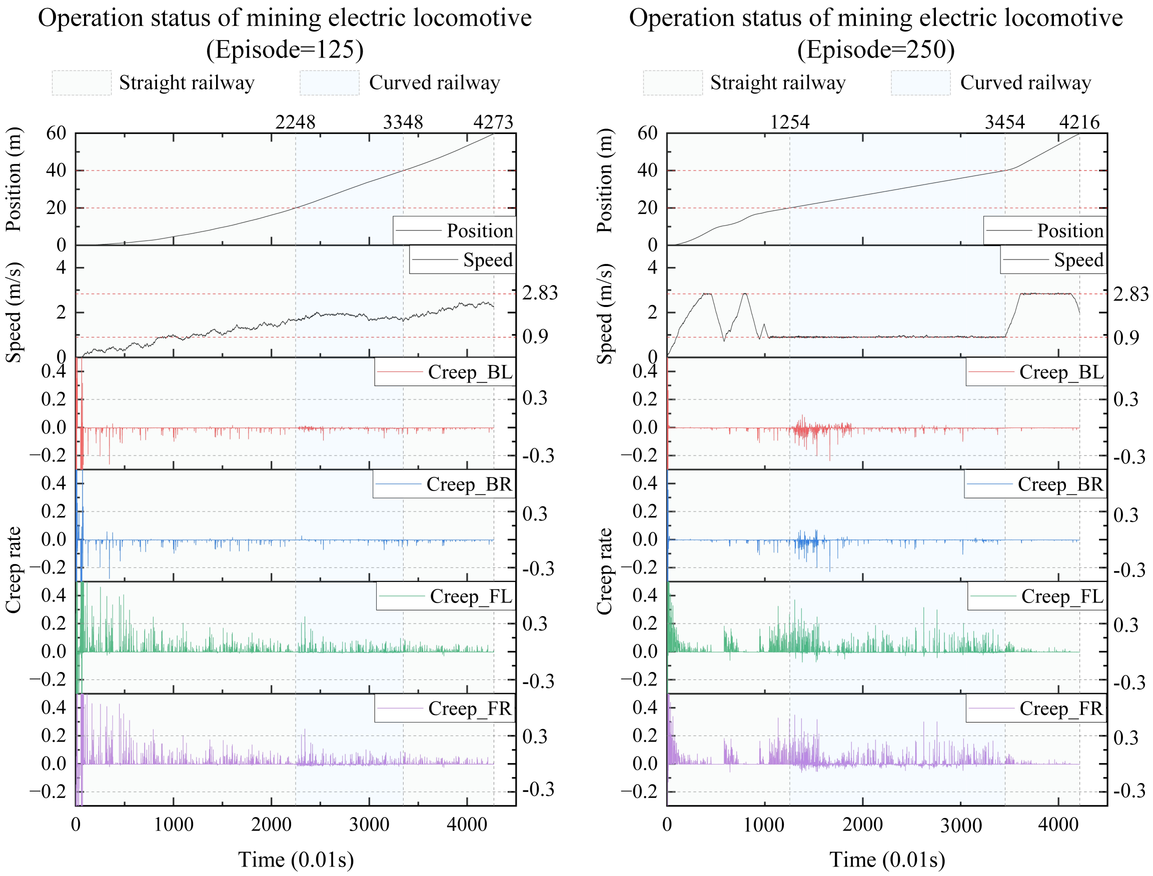


Figure 10. The operating status of mining electric locomotives with episode = 125 and episode = 250.

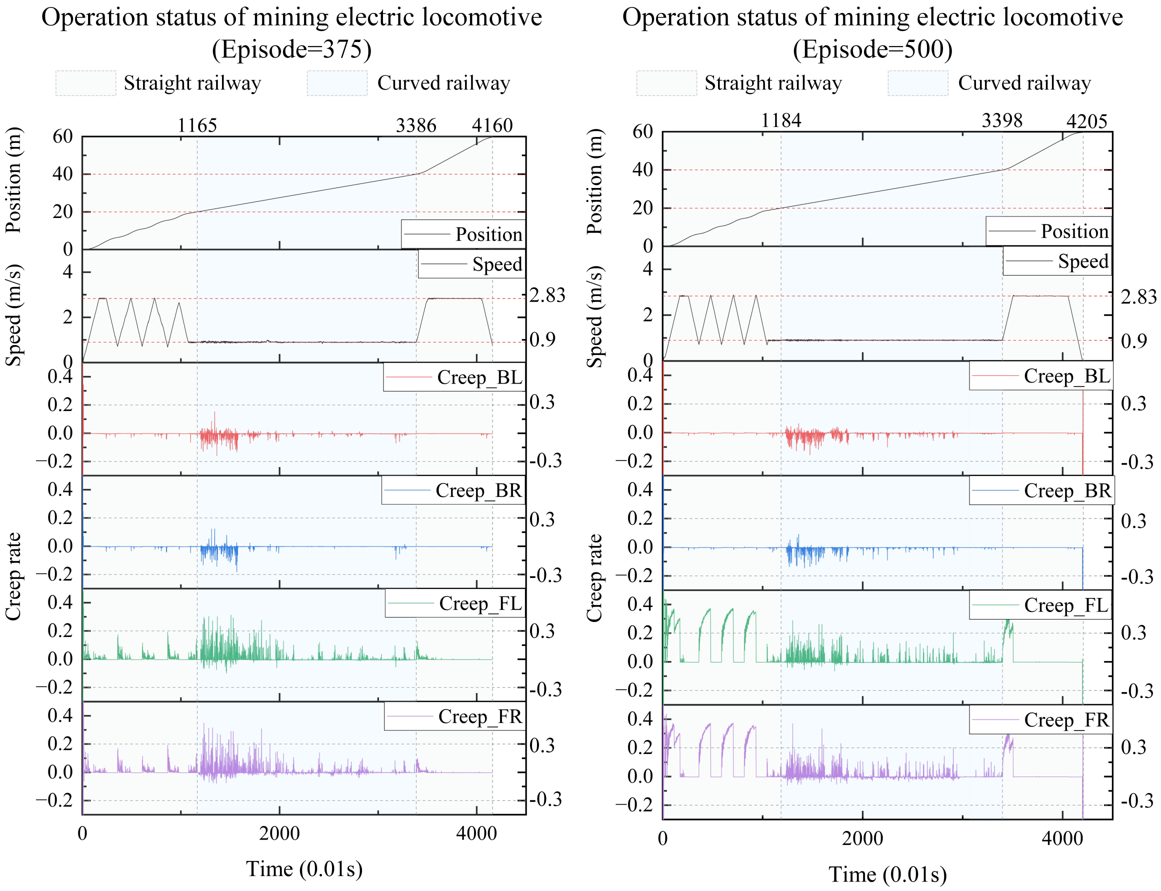


Figure 11. The operating status of mining electric locomotives with episode = 375 and episode = 500.

5. Conclusions

In order to reduce the occurrence of idling or slipping caused by wet track surface and improve vehicle safety when mining electric locomotive is driving, an improved ε-greedy of creep control strategy for mining electric locomotives based on RL algorithm is designed in this paper. By reproducing and proving the autonomous control method of mining electric locomotives based on improved ε-greedy, the analysis of the operating conditions of mining electric locomotives shows that creep rate is an important influencing factor on the operation of electric locomotives in wet and slippery tunnels. An RL based creep control method is proposed, which considers the efficiency of vehicle driving speed and the effectiveness of creep control in the algorithm design process. Finally, simulation verification is conducted based on a co-simulation platform that can simulate the dynamic process of mining electric locomotives underground. The results show that this method can achieve safe operation of mining electric locomotives in complex and ever-changing mine environments, including being able to drive with maximum efficiency on speed limited sections, follow at a safe distance, and reduce the occurrence of dangerous phenomena such as slipping and idling during vehicle driving.

Author Contributions: Conceptualization, Y.L.; methodology, Y.L.; software, Y.L.; validation, Y.L.; formal analysis, Y.L.; investigation, Y.L.; resources, Y.L., Z.Z and X.L.; data curation, Y.L.; writing—original draft preparation, Y.L.; writing—review and editing, Y.L. and X.L.; visualization, Y.L. and X.L.; supervision, Z.Z.; project administration, Z.Z.; funding acquisition, Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NATIONAL KEY R&D PROGRAM of funder grant number 2020YFB1314100, CHINA.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

doi:10.20944/preprints202404.1512.v1

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wang G.F.; Liu F.; Meng X.J.; et al. Research and practice on intelligent coal mine construction (primary stage). *Coal Science and Technology* **2019**, *47*, 1-36.
2. Ge S.R.. Present situation and development direction of coal mine robots. *China Coal* **2019**, *45*, 18-27.
3. Li Y.; Zhu Z.C.; Li X.Q.; Yang C.Y.; Lu H.. When mining electric locomotives meet reinforcement learning. *arXiv preprint.* **2023**, *arXiv:2311.08153*. Available online: https://doi.org/10.48550/arXiv.2311.08153
4. Fang C.C.; Jaafar S.A.; Zhou W.; Yan H.K.; Chen J.; Meng X.H.. Wheel-rail contact and friction models: A review of recent advances. *Proceedings of the Institution of Mechanical Engineers Part F - Journal of Rail and Rapid Transit* **2023**, *237*, 1245-1259.
5. Vollebregt E.; Six K.; Polach O.. Challenges and progress in the understanding and modelling of the wheel-rail creep forces. *Vehicle System Dynamics* **2021**, *59*, 1026-1068.
6. Zhao Y.H.; He X.; Zhou D.H.; Pecht M.G.. Detection and Isolation of Wheelset Intermittent Over-Creeps for Electric Multiple Units Based on a Weighted Moving Average Technique. *IEEE Transactions on Intelligent Transportation Systems* **2022**, *23*, 3392-3405.
7. Lu C.X.; Chen D.L.; Shi J.; Li Z.Q.. Research on wheel-rail dynamic interaction of high-speed railway under low adhesion condition. *Engineering Failure Analysis* **2024**, *157*, DOI 10.1016/j.engfailanal.2023.107935.
8. Zhao Y.H.; He X.; Zhou D.H.; Pecht M.G.. Detection and isolation of wheelset intermittent over-creeps for electric multiple units based on a weighted moving average technique. *IEEE Transactions on Intelligent Transportation Systems.* **2022**, *23*, 3392-3405.
9. Gao X.; Lu Y.. Study of locomotive adhesion control. *Railway Locomotive and Car* **2017**, *3*, 35-39.
10. Yamazaki O.; Ohashi S.; Fukasawa S.; Kondo K.. The proposal of re-adhesion control method with the advantage of individual control system. In Proceedings of International Conference on Electrical Systems for Aircraft, Railway, Ship Propulsion and Road Vehicles (ESARS), Aachen, Germany, 03-05 March 2015.
11. Yamashita M.; Soeda T.. Anti-slip re-adhesion control method for increasing the tractive force of locomotives through the early detection of wheel slip convergence. In Proceedings of 17th European Conference on Power Electronics and Applications (EPE ECCE-Europe), Geneva, Switzerland, 08-10 September 2015.
12. Wen X.K.; Huang J.C.; Zhang S.. Anti-slip re-adhesion control strategy of electric locomotive based on distributed MPC. In Proceedings of 2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), Zhangjiajie, China, 10-12 August 2019.
13. Wang S.; Wang X.G.; Huang J.C.; Sun P.F.; Wang Q.Y.. Adhesion Control of High Speed Train Based on Vehicle-control System. In Proceedings of 16th IEEE Conference on Industrial Electronics and Applications (ICIEA), Chengdu, China, 01-04 August 2021.
14. Çimen M.A.; Ararat Ö., Söylemez M.T.. A new adaptive slip-slide control system for railway vehicles. *Mechanical Systems and Signal Processing* **2018**, *111*, 265-284.
15. Wang S.; Wang X.G.; Huang J.C.; Sun P.F.; Wang Q.Y..Adhesion Control of High Speed Train Based on Vehicle-control System. In Proceedings of 2021 IEEE 16th Conference on Industrial Electronics and Applications (ICIEA), Chengdu, China, 01-04 August 2021.
16. Zhang S.; Huang Z.W.; Yang Y.Z.; Gao K.; Zhu C.. A Safe and Reliable Anti-Lock Wheel Control with Enhanced Forgotten Factor for Brake Operation of Heavy Train. In Proceedings of 2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC), Guangzhou, China, 12-15 December 2017.
17. Martin-Guerrero J.D.; Lamata L.. Reinforcement Learning and Physics. *Applied Sciences-Basel* **2021**, *11*, 8589.
18. Kulkarni S.R.; Lugosi G.. Finite-time lower bounds for the two-armed bandit problem. *IEEE Transaction on Automatic Control* **2000**, *45*, 711-714.
19. Tze L.L.; Herbert R.. Asymptotically efficient adaptive allocation rules. *Advances in Applied mathematics* **1985**, *6*, 4-22.
20. Auer P.; Cesa-Bianchi N.; Fischer P.. Finite-time analysis of the multiarmed bandit problem. *Machine Learning* **2002**, *47*, 235-256.

21. Kulkarni S.R.; Lugosi G.. Finite-time lower bounds for the two-armed bandit problem. *IEEE Transaction on Automatic Control* **2000**, *45*, 711-714.
22. Volodymyr K.; Doina P.. Algorithms for multi-armed bandit problems. *arXiv preprint.* **2014**, *arXiv:1042.6028*.
23. Malvezzi M.; Pugi L.; Papini S.; Rindi A.; Toni P.. Identification of a wheel-rail adhesion coefficient from experimental data during braking tests. *Proceedings of the Institution of Mechanical Engineers Part F-Journal of Rail and Rapid Transit* **2013**, *227*, 128-139.
24. Polach O.. Creep forces in simulations of traction vehicles running on adhesion limit. *Wear* **2005**, *258*, 992-1000.
25. Fu G.H.. An extension of Hertz's theory in contact mechanics. *Journal of Applied Mechanics-Transactions of the ASME* **2007**, *74*, 373-374.
26. Polach O.. A fast wheel-rail forces calculation computer code. *Vehicle System Dynamics* **1999**, *33*, 728-739.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.