algorithms

MDPI

Article

Deep Q-Network Algorithm-Based Cyclic Air Braking Strategy for Heavy-Haul Trains

Changfan Zhang, Shuo Zhou, Jing He and Lin Jia *

College of Electrical and Information Engineering, Hunan University of Technology, Zhuzhou 412007, China; zcf@hut.edu.cn (C.Z.); 14381@hut.edu.cn (S.Z.); hejing@hut.edu.cn (J.H.)
* Correspondence: jialin@hnu.edu.cn

Abstract: Cyclic air braking is a key element for ensuring safe train operation when running on a long and steep downhill railway section. In reality, the cyclic braking performance of a train is affected by its operating environment, speed and air-refilling time. Existing optimization algorithms have the problem of low learning efficiency. To solve this problem, an intelligent control method based on the deep Q-network (DQN) algorithm for heavy-haul trains running on long and steep downhill railway sections is proposed. Firstly, the environment of heavy-haul train operation is designed by considering the line characteristics, speed limits and constraints of the train pipe's air-refilling time. Secondly, the control process of heavy-haul trains running on long and steep downhill sections is described as the reinforcement learning (RL) of a Markov decision process. By designing the critical elements of RL, a cyclic braking strategy for heavy-haul trains is established based on the reinforcement learning algorithm. Thirdly, the deep neural network and Q-learning are combined to design a neural network for approximating the action value function so that the algorithm can achieve the optimal action value function faster. Finally, simulation experiments are conducted on the actual track data pertaining to the Shuozhou–Huanghua line in China to compare the performance of the Q-learning algorithm against the DQN algorithm. Our findings revealed that the DQN-based intelligent control strategy decreased the air braking distance by 2.1% and enhanced the overall average speed by more than 7%. These experiments unequivocally demonstrate the efficacy and superiority of the DQN-based intelligent control strategy.

Keywords: heavy-haul train; long and steep downhill; cyclic air braking; DQN; intelligent control

Citation: Zhang, C.; Zhou, S.; He, J.; Jia, L. Deep Q-Network Algorithm-Based Cyclic Air Braking Strategy for Heavy-Haul Trains. *Algorithms* **2024**, *17*, 190. https://doi.org/10.3390/a17050190

Academic Editors: Marko Đurasević and Bruno Gašperov

Received: 4 April 2024
Revised: 23 April 2024
Accepted: 29 April 2024
Published: 30 April 2024

1. Introduction

Heavy-haul train transportation has been widely valued around the world due to its advantages such as large capacity, high efficiency and low transportation cost. When heavy-haul trains run on long and steep downhill railway sections, cyclic air braking is needed for controlling the train speed [1]. At present, the utilization of air braking mainly relies on the driving experience of the driver. However, existing braking methods based on drivers' experience cannot meet the safety and efficiency requirements of heavy-haul train operation [2]. Therefore, it is of great importance to develop an intelligent control strategy for the cyclic braking of heavy-haul trains running on long and steep downhill sections [3].

To date, air braking methods for heavy-haul trains running on long and steep downhill railway sections have been studied in-depth by many researchers. The main solutions can be classified into imitation learning methods based on expert data, numerical solution methods based on an optimal train control models, and reinforcement learning based on a Markov decision process.

In terms of imitation learning based on expert data, it is necessary to provide data on experts' driving courses during the training stage to simulate an expert's driving behavior in a supervised manner. For example, Ref. [4] combined expert data with the concept of generative adversarial learning and proposed a representative generative adversarial

imitation learning algorithm. Ref. [5] integrated expert data with reinforcement learning to train a new intelligent driving control strategy. The reinforcement learning algorithm supervised by the expert model was able to make the operation more efficient and stable. However, this type of imitation learning method did not directly establish safety assessments and constraints, and the expert data it used failed to fully cover special emergency traffic situations.

In terms of research on numerical solutions based on train optimal control models, Pontryagin's Maximum Principle (PMP) was employed in ref. [6] to ascertain the most effective driving approach for controlling trains through the use of generalized motion equations. By considering line conditions, such as different slopes and different speed limits, a version of the key equation for train traction energy consumption was proposed, and it was used to calculate the speed of train cyclic braking. Ref. [7] established an optimized model based on a train dynamic model, and integrated the artificial bee colony algorithm to find the appropriate switching points of different states and develop an optimal operation strategy for heavy-haul trains running on long and steep downhill railway sections. Ref. [8] combined an approximate model of data-learning systems with model predictive control to solve planning problems with safety constraints. Ref. [9] considered the energy consumption and comfort of a heavy-haul train when determining the optimal strategy for cyclic braking. Two methods were implemented, including a pseudo-spectral method and a mixed-integer linear programming method, to develop an optimal driving strategy for the train. However, this type of method requires comprehensive data to build the model. Moreover, the accuracy of the model has a strong influence. Therefore, the problem of model bias under complex uncertainty scenarios is prominent.

Reinforcement learning algorithms based on a Markov decision process have been increasingly applied to the intelligent control of trains in recent years. Ref. [10] proposed an optimal model of operation energy consumption leveraging a Q-learning algorithm. Then, a cyclic braking strategy was developed based on the train status value function to solve the problems of train punctuality and energy-saving operation optimization. Ref. [11] designed an intelligent train control method using a reinforcement learning algorithm based on policy gradient. The performance of the agent was continuously optimized to realize the self-learning process of the controller. An improved Q-learning algorithm was proposed in Reference [12]. The target reward for energy consumption and time are updated in different ways. In Reference [13], a Q-SARSA algorithm was proposed by combining Q-learning and the SARSA update rules, which considerably improved the efficiency of subway operations by combining deep fully connected neural networks with it. These methods can achieve cyclic braking, energy savings and emission reduction in a heavy-haul train without the need for a pre-designed reference speed curve. However, the state space in this algorithm is discretized, which results in slow convergence of the optimization algorithm during the learning process and the curse of dimensionality problem. Thereafter, ref. [14] proposed a reinforcement learning algorithm based on two-stage action sampling to solve the combinatorial optimization problem in heavy-haul railway scheduling. This approach not only alleviates the curse of dimensionality but also naturally satisfies the optimization objective and complex constraints. Ref. [15] proposes a reinforcement learning method for multi-objective speed trajectory optimization to simultaneously achieve energy efficiency, punctuality and accurate parking. Ref. [16] employed a double-switch Q-network (DSQ network) architecture to achieve fast approximation of the action value function and enhance the parameter sharing of states and actions. However, the methods used in refs. [14–16] still cannot make full use of the large amount of unstructured data generated during train operation, that is, the data utilization rate is reduced. Furthermore, in the design of a train model, the environmental characteristics of heavy-haul trains running on long and steep downhill railway sections are not considered. Therefore, it is difficult to apply such models to the optimal control of heavy-haul trains running on long and steep downhill sections.

Based on the above works, an intelligent algorithm utilizing DQN for the cyclic braking of heavy-haul trains is proposed in this paper to solve the problem of the cyclic air braking of heavy-haul trains traversing lengthy and steep downhill railway sections. The main contributions of this work are as follows:

(1) A model with operation constraints is constructed for heavy-haul trains equipped with a conventional pneumatic braking system and traversing lengthy and steep downhill railway sections. In addition, the performance indexes of a train running on a long and steep downhill section are introduced to evaluate the control performance of the heavy-haul train.

(2) The action value functions are approximated based on a neural network. In accordance with the Q-learning algorithm, the neural network is combined with reinforcement learning to avoid the occurrence of the curse of dimensionality inherent in the Q-learning algorithm. Therefore, this method is suitable for solving the train control problem characterized by continuous state space.

(3) A prioritized experience replay mechanism is proposed. The samples are prioritized and selected according to their importance so that important and high-reward samples are selected and trained more frequently. Therefore, the learning efficiency of the algorithm is improved for important samples, which accelerates the convergence speed of the algorithm. As a result, the performance of the algorithm can be improved.

The rest of this paper is organized as follows: In Section 1, the design of the control model for heavy-haul trains is presented. The constraints of the train operation and the performance indexes of the train control are introduced into the model. The train control problems of a heavy-haul train running on a long and steep downhill section are described in detail. In Section 2, a cyclic air braking method based on the DQN algorithm is established for heavy-haul trains. In Section 3, the validity and robustness of the proposed method are verified via simulation. Finally, the conclusions of this study are summarized in Section 4.

2. Model Construction for Heavy-Haul Trains

2.1. Dynamic Model

During the operation of a heavy-haul train, factors such as track gradient, train composition and on-board mass cause the train to be subjected to diverse forces. In this study, the interaction forces between the cars are not taken into account when calculating the additional resistance. Therefore, the forces imposed on the train during operation mainly include locomotive traction force, braking force (including electric braking and pneumatic braking), basic running resistance and additional resistance. Essentially, a heavy-haul train is a distributed power network system consisting of multiple locomotives and freight cars. According to the Newtonian principle of dynamics, the mathematical expression of each train model can be defined as follows [17]:

$$M\dot{v} = F - U_1 - U_2 - F_R, \tag{1}$$

Usually, the running resistance F_R encountered by a heavy-haul train during braking on a long and steep downhill section is mainly composed of basic resistance M_R and additional resistance L_R. These resistances depend on the operating speed of the heavy-haul train as well as its physical characteristics [18]. The running resistance is calculated as follows:

$$F_R = M_R + L_R, \tag{2}$$

According to previous research, the calculation formula of the basic resistance for a heavy-haul train is as follows [19]:

$$M_R = M(\varphi_1 + \varphi_2 v + \varphi_3 v^2), \tag{3}$$

The additional resistance is determined by the section force g_R, the curvature resistance c_R and the tunnel resistance t_R [19], as shown in Equation (4). The specific calculations of these factors [20] are presented in Equation (5):

$$L_R = g_R + c_R + t_R, \tag{4}$$

$$\begin{cases} g_R = Mg\sin(\arctan\frac{i}{1000}) \\ c_R = 600/R \\ t_R = 0.00013L_s \end{cases}, \tag{5}$$

For heavy-haul trains, the control inputs for the locomotives include the traction force and braking force, whereas the input for freight trains includes only the braking force. There are mainly two types of brake equipment for heavy-haul trains: one is rheostatic brake equipment and the other is pneumatic braking equipment. Rheostatic brake equipment, also known as the regenerative brake, can feed back energy to other locomotives to provide power. A pneumatic braking system achieves braking force by reducing the air pressure in the train's air braking pipes [21].

The traction force F of a heavy-haul train depends on the relative output ratio h^{tr} of the maximum traction force $u^{tr}_{max}(v)$, while the electric braking force U_1 is determined by the maximum electric braking force $u^{d}_{max}(v)$ and its relative output ratio h^d. Therefore, the traction force and electric braking force of a train can be calculated according to Equation (6):

$$\begin{cases} F = F^{tr}(u^{tr}_{max}(v), h^{tr}) \\ U_1 = F^{d}(u^{d}_{max}(v), h^{d}) \end{cases}, \tag{6}$$

where $u^{tr}_{max}(v)$ and $u^{d}_{max}(v)$ are the piecewise functions of the train's running speed [22]. In addition, it is impossible for each locomotive to output electric braking force and traction force at the same time [23]; thus, the relative output ratio of h^d and h^{tr} is $h^d \times h^{tr} = 0$.

The air braking system of a heavy-haul train is its main braking force and the key to ensuring the safety of train operation. Figure 1 is a diagram showing the structural composition of the air braking system of a heavy-haul train [24]. According to Reference [25], U_2 in Equation (1) can be calculated based on Equation (7):

$$U_2 = \theta_b \times \varphi_b \times \beta_s \times 1000, \tag{7}$$

where θ_b and φ_b are intimately associated with the train's physical characteristics, such as the brake lever and transmission efficiency of locomotives and freight cars, while the service braking coefficient β_s is assigned a value based on the rated pressure p of the air pipes and the corresponding air pressure drop Δp. Therefore, when air braking is applied by heavy-haul trains, the output of the air braking force depends on the air pressure drop Δp [25]. Equation (7) can be rewritten as a function of Δp as follows:

$$U_2 = h^a \times F(\Delta p), \tag{8}$$

where h^a is a binary variable that determines whether air braking is engaged ($h^a = 1$) or released ($h^a = 1$).

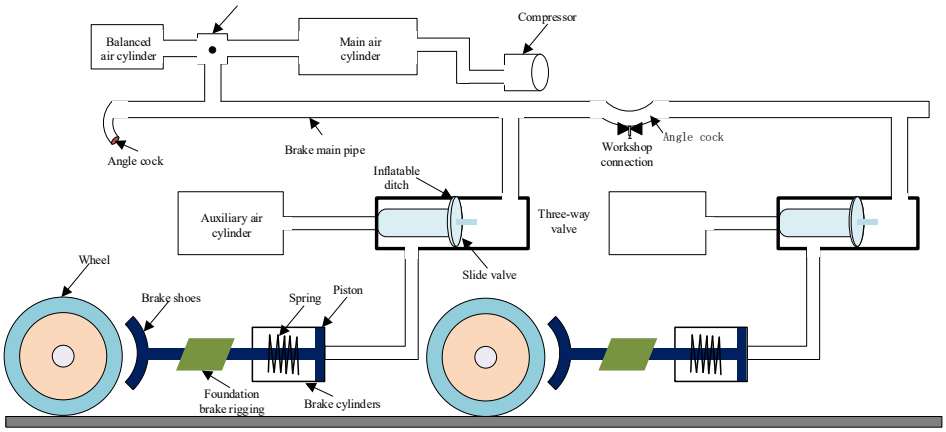

Figure 1. The compositional structure of the air braking system for heavy-haul trains.

2.2. Running Constraints

The aim of this research on the circulating air braking of a heavy-haul train on a long and steep downhill section is essentially to solve a multi-constraint and multi-objective optimization problem. Considering the actual requirements of train driving control and model design, the running constraints set in this research are as follows:

(1) Air-refill time: When a train operates on a long and steep downhill section, it necessitates the adoption of cyclic braking for speed control. To guarantee sufficient braking force in the next braking cycle, sufficient time should be ensured for refilling the air pipe to its maximum pressure [26]. In other words, the release phase must not be shorter than the minimum air-filling time T_a, as stated in the operating guidelines.

$$t_{j+1}^{b} - t_{j}^{r} \geq T_a, \tag{9}$$

In the above formula, T_a is closely related to the formation of the train and the pressure drop within the train air pipe. For predetermined train parameters, the air-filling time needs to be ascertained.

(2) Speed limit: To ensure safety, the speed of the heavy-haul train cannot exceed the speed limit $\overline{V}$ at any point on the lengthy and steep downhill line stretch. This value often depends on the underlying framework of the railway line or the provisional setup. Additionally, the speed of the train should be greater than the minimum air brake release speed V_{min}^{r}. The specified limit is designated as 40 km/h for a 20,000-ton heavy-haul train formation [19]. Therefore, the speed should meet the following requirement:

$$V_{min}^{r} \leq v \leq \overline{V}, \tag{10}$$

(3) Electric braking force: Heavy-haul trains are equipped with both an electric braking system and an air braking system. When the train engages electric braking, the braking current is regulated by adjusting the series excitation resistance of the electric braking system, so as to generate a continuous electric braking force to slow down the running speed of the heavily loaded train. In actual operation, the constraint is the maximum electric braking force [27], so the relative output ratio of the electric braking force should satisfy

$$0 \leq h^{d} \leq 1, \tag{11}$$

2.3. Performance Indicators

This study primarily focuses on the safety and servicing cost of the heavy-haul train operation process. The maintenance cost is expressed by air braking distance. Hence, two indicators are introduced to evaluate the control performance of the heavy-haul train.

(1) Safety: Safety serves as the fundamental prerequisite for train operation. The speed of the heavy-haul train should be kept under the upper limit. Yet, it cannot be lower than

V^r_{min}. Here, the parameter Y is defined to indicate whether the train speed remains within the speed range.

$$Y = \begin{cases} 1, & V^r_{min} \le v \le \overline{V} \\ 0, & otherwise \end{cases}, \tag{12}$$

(2) Air braking distance: As excessive wear will be caused by the friction between wheels and brake shoes when the air brake is engaged for a long distance, a maintenance cost will be generated by the replacement of air brake equipment. By reducing the air brake distance during operation, the maintenance cost can be reduced. Therefore, the air brake distance L_a of a heavy-haul train is defined as Equation (13).

$$L_a = \int_0^T b^a(t) \times v(t) dt, \tag{13}$$

3. DQN Control Algorithm

Reinforcement learning is a machine learning method for goal-oriented tasks. It does not tell the agent how to act, but instead, guides the agent to learn the correct strategy through interaction with the environment. In this section, the train operation process will first be defined as a Markov decision process. Then, a control algorithm based on DQN is proposed to learn the cyclic braking strategy of train operation when running on a long and steep downhill railway section.

3.1. Markov Decision Process

Prior to implementing the DQN algorithm, it is essential to define the control process for train operation when running on a long and steep downhill section as a Markov decision process (MDP), which involves formalizing sequential decision-making. An illustration of the MDP interaction during the operation of heavy-haul trains running on long and steep downhill sections is shown in Figure 2.

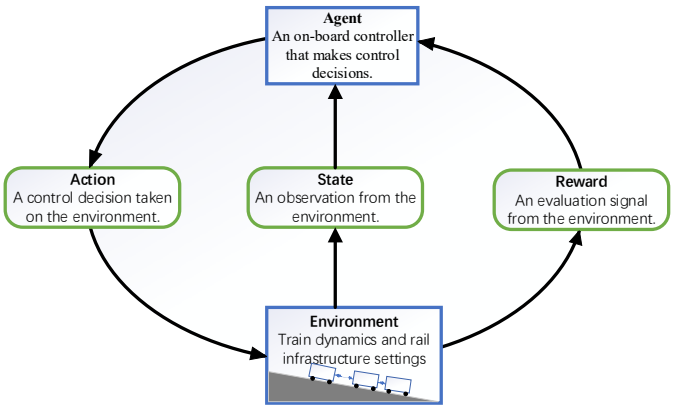

Figure 2. Illustration of MDP interaction during operation of heavy-haul trains.

The locomotive of a heavy-haul train is described as the agent responsible for making control decisions, while the heavy-haul train's dynamics and the railway's underlying structure arrangements are defined as the environment. Assuming that the time interval between two consecutive states is Δt, the control process is evenly divided into N steps with respect to the planned operating time T, and the position, operating speed and operating time of the train are taken as the states of the heavy-haul train.

$$s_k = [P_k, V_k, T_k], k = 0, 1, 2, \cdots, n, \tag{14}$$

where s_k is the status of the heavy-haul train at step k, P_k is the position of the train, V_k is the train speed and T_k is the train running time. s_0 and s_n represent, respectively, the initial

state and final state of the heavy-haul train running on a long and steep downhill section, and their calculation is as follows:

$$s_0 = [0, V_0, 0], \tag{15}$$

$$s_n = [P, V_n, T], \tag{16}$$

Starting from the initial state s_0 of the train, succeeding states in the train control process are determined through actions until the final state s_n is reached. Specifically, at a given train state s_k, the agent chooses the optimal action a_k from all feasible actions. After executing the action a_k, the agent can obtain the next state s_{k+1} and receive the corresponding reward r_{k+1}. The next state is s_{k+1}, which is solely influenced by the current state s_k and action a_k. Its calculation is as follows:

$$s_{k+1} = \Phi(s_k, a_k), k = 0, 1, \cdots, n-1, \tag{17}$$

where $\Phi()$ represents the functional dynamics of the system associated with the heavy-haul train dynamic model, while (s_k, a_k) is denoted as the state–action pair. Notably, the agent lacks prior knowledge of train dynamics. Such knowledge is solely utilized for generating driving experience throughout the interaction and to supply the agent with training data. Herein, the control action refers to the setting of electric braking and the air braking notch.

$$a_k = [h_k^a, h_k^d], k = 0, 1, \cdots, n, \tag{18}$$

where h_k^a is a binary variable, and h_k^d denotes the relative output ratio of the electric braking force generated by the train locomotive, which is subjected to the constraint condition outlined in Equation (11).

The control output of a heavy-haul train in each cycle is determined solely by the speed and time of the current train. Thus, the control process of a heavy-haul train can be defined by reinforcement learning as a Markov decision process, which is expressed as follows:

$$s_0 \xrightarrow{a_0} s_1, r_1 \xrightarrow{a_1} \cdots s_k, r_k \cdots \xrightarrow{a_{n-2}} s_{n-1}, r_{n-1} \xrightarrow{a_{n-1}} s_n, \tag{19}$$

3.2. DQN Algorithm Model

In this section, an intelligent control method based on the DQN algorithm is designed for a heavy-haul train running on a long and steep downhill section. The DQN algorithm combines reinforcement learning with deep neural networks to design a neural network that approximates the action value function. By applying the neural network, the curse of dimensionality caused by the discretization of the state space in the Q-learning algorithm is avoided. To solve instability when using a neural network to approximate the action value function, an independent target Q-network updated at regular intervals and a prioritized experience replay mechanism are incorporated into the DQN. Hence, the performance of the DQN algorithm in the cyclic braking control process of the heavy-haul train is further improved. The overall structure of the algorithm is shown in Figure 3.

The intelligent cyclic air braking system for heavy-haul trains designed in this research consists of a sensing module, a strategy-making module, an action execution module and a reward feedback module. Among them, the sensing module is responsible for collecting the train state information, the strategy-making module devises an optimal braking strategy by using the DQN algorithm, the action execution module converts the strategy into specific braking instructions, and the reward feedback module evaluates the effect of each braking action. The key modules are described in more detail below.

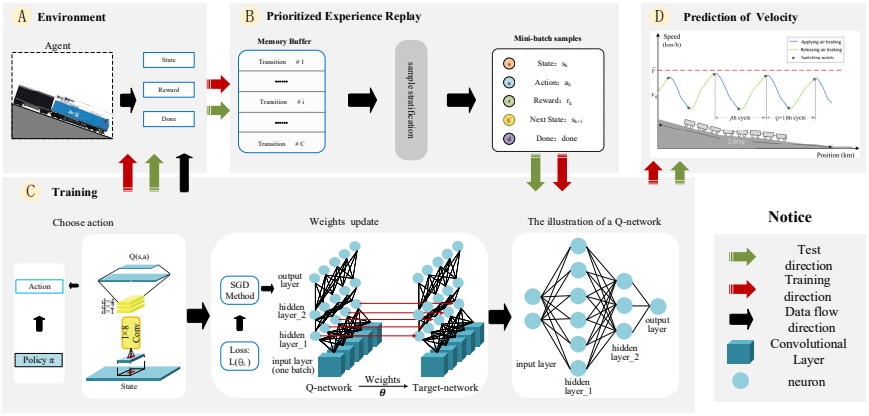

Figure 3. Overall framework of DQN-based intelligent control algorithm for heavy-haul trains.

3.2.1. Policy Design

Policy decides how an agent behaves at a specific time step. In this scenario, the train controller is treated as an agent and the state of the heavy-haul train can be described using Equations (14)–(17). Formally, policy is a function that calculates the probability of selecting each potential action relative to the Q function, and its expression is as follows:

$$\pi(a|s) = \Gamma[s, Q(s,a)], \tag{20}$$

where $\Gamma[]$ is the mapping policy. The action selection policy consists of two sub-policies, i.e., $\pi = \{\pi_1, \pi_2\}$. Concretely speaking, to assure that the duration of the release air brake release stage satisfies the constraint of minimum air-refilling time in Equation (9), when the air brake release action is selected at the kth step, that is, $a_k^+ = \left[0, h_k^d\right]$, then this action ought to be sustained for the subsequent $n_t - 1$ steps, with $a_{k+1} = a_{k+2} = \cdots = a_{k+n-1} = a_k^+$ and $n_t = T_{AI}/\Delta t$. Additionally, the ε-greedy strategy (Equation (21)) is employed to determine the control actions for subsequent stages of the train operation process, wherein an action is chosen randomly with a probability of ε, whereas with the probability of $1 - \varepsilon$, the action $a*$ with the largest estimated Q value will be taken.

$$\pi(a|s_k) = \begin{cases} (1-\varepsilon) + \frac{\varepsilon}{|A(s_k)|}, a = a* \\ \frac{\varepsilon}{|A(s_k)|}, a \neq a* \end{cases}, \tag{21}$$

where $|A(s_k)|$ is the number of actions in the event of s_k.

3.2.2. Reward Design

The optimization goal of the reinforcement learning problem is reflected by the reward function. For the train control process in question, the operating speed cannot exceed its upper limit to ensure a safe operation. Therefore, the constraint of Equation (10) must be fulfilled. On the condition that the speed exceeds the upper limit $\overline{V}$ or is below the minimum remission speed V_{min}^r, a negative reward R_c will be given to the agent. If the heavy-haul train engages in air braking at step k, it will receive a reward of zero. Conversely, if air braking is not engaged, a positive reward R_d will be granted to incentivize the release of air braking. Hence, the specified reward is delineated as listed below:

$$r_{k+1} = \begin{cases} R_c, V_{k+1} < V_{min}^r \text{ or} V_{k+1} > V_{max}^r \\ 0, h_k^a = 1 \text{ and } V_{min}^r \leq V_{k+1} \leq V_{max}^r \\ R_d, h_k^a = 0 \text{ and} V_{min}^r \leq V_{k+1} \leq V_{max}^r \end{cases}, \tag{22}$$

3.2.3. Prioritized Experience Replay Design

The experience of the heavy-haul train at step k is recorded as a transition $e_k = (s_k, a_k, r_{k+1}, s_{k+1}, done)$, and then, all experiences obtained over many training episodes are stored in the buffer D. The term *done* in an array is a sign signal. If *done* is True, it indicates that the training state s_{k+1} is the final state or a negative reward R_c will be imparted. During training, the DQN algorithm updated by means of sampling prioritization is applied to small-batch transitions $(s_k, a_k, r_{k+1}, s_{k+1}, done)$ according to the priority from the memory buffer storing experience. This mechanism enables greater data efficiency, as experiences of higher priority will be more frequently selected for weight updates. Samples are selected probabilistically to ensure that experiences with zero TD error can be sampled. The sampling formula $P(i)$ of the priority experience replay is presented in Equation (23):

$$P(i) = \frac{p_i^\alpha}{\sum_k p_k^\alpha}, \tag{23}$$

where p_i represents the priority of the ith experience sample; α determines the utilization priority; and p_i is defined according to Equation (24),

$$p_i = |\delta_i| + \eta, \tag{24}$$

where δ_i is the temporal difference error (TD error), representing the difference between the reward received by the agent after performing an action and the expected reward. η is a very small value, taken as 0.001 in this paper. It is used to prevent an experience sample from not being played back into the experience pool after its TD-error value reaches 0.

The smaller the TD error, the smaller the actual difference between the Q value estimated from the sampled data and the target Q value. A small TD error indicates that the agent is able to manage the corresponding situation well and there is no need to sample frequently or to train this experience. When there is significant deviation between the estimated Q value and the target Q value, this indicates that the experience is relative to the strategy currently learned by the agent, that is, the experience value is high, and frequent training should be prioritized to cope with similar changes that could occur. In addition, sampling by priority breaks the strong correlation between successive samples that adversely affects the reinforcement learning algorithm. As a result, it reduces the variance in the learning updates.

3.2.4. Action Value Function Design

Defined as $Q(s, a; \theta)$, the action value function demonstrates the quality of the action taken by the agent in a given state. In the form of a formula, the value of an action can be represented as the sum of all rewards that the agent is able to accumulate in subsequent steps, commencing from the current state and action. During the learning process, the agent is coached to choose an action in a manner that optimizes the overall future reward, referred to as the expected return, as shown in the following formula:

$$R_k = \sum_{k\prime=k}^{N} \gamma^{k\prime-k} r_{k\prime}, \tag{25}$$

The optimal Q function $Q^*(s,\ a)$ is characterized as the highest cumulative reward attainable by adhering to the policy π upon entering the state s and executing the action a. In the form of $Q(s, a)$, it is a function consisting of two parts: the state value function and the advantage function. The action value function expresses the expected return obtained by executing the action in the specified states, that is, indicating that taking this action is good or bad. The formula for calculating the optimal Q function is presented in Equation (26):

$$Q*(s,a) = \max_{\pi} E[R_k | s_k = s, a_k = a, \pi], \tag{26}$$

Theoretically, the optimal Q function obeys the Bellman equation, which is shown in Equation (27):

$$Q*(s,a) = E_{s\prime}[r(s,a) + \gamma max_{a\prime} Q*(s\prime, a\prime)], \tag{27}$$

where s' and a' denote the state and the potential action at the subsequent time step, respectively.

In the DQN framework, the neural network is employed to approximate the optimal function. The Q function is estimated by the neural network, with the weight parameters denoted as θ (henceforth called the Q-network). Throughout the training process, the Q-network undergoes training through weight adjustments aimed at reducing the mean square error to a minimum, as defined by Equation (27). In this context, the best possible target value, $r + \gamma max_{a\prime} Q^*(s\prime,\ a)$, is approximated with the target value $y = r + \gamma max_{a\prime} \hat{Q}(s\prime,\ a\prime;\ \theta^-)$. Thus, the loss function of updating the weight for the network is as follows:

$$L_i(\theta_i) = E_{s,a\sim p(),s\prime\sim P}[r + \gamma \underset{a\prime}{max} Q(s\prime, a\prime, \theta^-) - Q(s,a,\theta)]^2, \tag{28}$$

where L_i is the loss of the ith iteration; p is the joint probability distribution of states and actions; θ is the weight parameter of the Q-network in the iteration; and θ^- is the weight parameter of the target network. The target network has an identical structure to the Q-network. To revise the Q-network, it is necessary to initially compute the derivative of the loss function concerning θ, and the gradient $\nabla_\theta L$ will be obtained. Then, the stochastic gradient descent (SGD) method is used to update the weight θ as follows:

$$\theta + \lambda \bullet \nabla_\theta L \rightarrow \theta, \tag{29}$$

where λ is the learning rate. The weight of the target network θ^- is updated periodically. When the parameters of the Q-network undergo J iterations of updating, a weight θ^- will be assigned to the target network to generate a new target network. Then, $\hat{Q}$ will be adopted to generate the target value y of the Q-learning to perform subsequent parameter updates J times for the Q-network. Through this approach, the overestimation or training instability caused by the correlation between the Q-network and the target network $\hat{Q}$ can be avoided; thus, the algorithm can calculate more accurately and stably, and the convergence speed can be increased.

Algorithm 1 summarizes the control method for heavy-haul trains based on the DQN algorithm.

Algorithm 1 DQN-Based Intelligent Control Strategy for Circulating Air Brake of the Heavy-Haul Train

///Initialization///
1: Use weight θ to randomly initialize the Q-network.
2: Use weight $\theta^- = \theta$ to initialize target network $\hat{Q}$.
3: initialize the experience pool D with size C, greedy probability ε, small-batch sample size n_e, discount rate γ, learning rate λ and the parameter update episode J of the target network
///Process of training///
4: for episode = 1,. . .M do
5: initialize the state s_0 of train through Equation (13)
6: for k = 0,1,. . .,$N-1$ do
7: choose action a_k based on strategy π
8: execute action a_k; receive rewards r_{k+1} and the next state of the train s_{k+1} according to Equation (20) and Equation (15), respectively. Determine whether the train reaching the target point is done, forming a quadruple (s_k,a_k,r_k,s_{k+1}, done)

9: calculate the priority p_i, of the quadruple, then calculate the sampling probability $P(i)$, and finally, calculate the importance sampling weight ω_i and store the quadruple (s_k, a_k, r_k, s_{k+1}, done) in the experience pool D with the probability $P(i)$
10: draw n_e quadruples (s_k,a_k,r_k,s_{k+1}, done) from the experience pool D according to sampling probability
11: If done = True, set $y_k = r_k$; otherwise, $y_k = r_k + \gamma max_{a\prime} Q\left(s_{k+1}, a\prime; \theta^-\right)$
12: calculate the $\nabla_\theta L$ of gradient $(y_k - Q(s_k, a_k; \theta))^2$ with respect to weight θ
13: update the weight θ by SGD method
14: update the weight θ^- of the target network $\hat{Q}$ according to $\theta^- = \theta$ every J perods
15: end for
16: end for

4. Algorithm Simulation and Analysis

In this study, simulations were carried out with real data obtained for a section of the Shuohuang railway line in China to confirm the efficacy of the proposed algorithm. Firstly, the experimental parameter data setting was introduced. Then, the simulation results were presented and analyzed. The simulations were divided into three parts: the model training process, a practical application performance test, and performance comparison using different algorithms.

4.1. Experimental Parameter Settings

To prove the effectiveness of the intelligent control method, a heavy-haul train weighing 20,000 tons was taken as the object of the simulations. The train formation consisted of 1 HXD1 electric locomotive + 108 freight cars + 1 HXD1 electric locomotive + 108 freight cars. The HXD1 electric locomotive had the abilities of traction and regenerative braking. The whole train was equipped with air braking. For specific train parameters, see Table 1.

Table 1. Train parameters.

Locomotive Parameters		Freight Car Parameters	
Parameter Name	**Value**	**Parameter Name**	**Value**
Model	HXD1	Model	C80
Mass	200 t	Mass	100 t
Length	35.2 m	Length	13.2 m

Based on the data for the section spanning from Longgong Station to Beidaniu Station of the Shuozhou–Huanghua Line, simulations were carried out to obtain a speed curve of the heavy-haul train running on a long and steep downhill section. The operation section had a total length S = 23,800 m, with a gradient of 10–12‰ for the extended and precipitous downhill stretch. The speed limit on this line was 80 km/h. The specific data are presented in Table 2.

Table 2. Route information.

Distance (m)	Gradient (-‰)	Distance (m)	Gradient (-‰)
0–1000	1.5	12,430–14,080	10.5
1000–1400	7.5	14,080–16,330	11.4
1400–6200	10.9	16,330–19,130	10.6
6200–6750	9	19,130–22,260	10.9
6750–12,430	11.3	22,260–23,800	3.3

The parameter settings of Algorithm 1 used in this paper are listed in Table 3:

Table 3. Algorithm hyperparameters.

Parameter	Value	Parameter	Value
Maximum training episode M	100,000	Minimum air-refilling time T_{AI}	100
Update period of target network	300	Batch size of sampling n_e	32
Discount rate γ	0.94	Learning rate λ	0.001
Initial value of ε	0.98	Final value of ε	0.1
Capacity of memory buffer D	5000	Planned operation time T	1500 s
Positive reward R_d	5	Negative reward R_c	−20
Minimum braking speed V^r_{min}	40 km/h	Maximum braking speed V^r_{max}	80 km/h

4.2. Simulation Experiment Verification

4.2.1. Model Training Process

In this study, simulations were carried out under different parameters and training cycles with the proposed DQN algorithm for verification purposes. Under the specified circumstances, the time interval Δt was 100 s. In addition, the initial speed V_0 of the heavy-haul train running on a long and steep downhill section was 65 km/h. The change curve of the cumulative reward obtained using the proposed algorithm is shown in Figure 4. As shown in Figure 4a–c, the learning rates were determined to be 0.03, 0.003 and 0.0003, respectively, for the cumulative reward curve that changes depending on the number of episodes in different experiments. It is evident that during the short-term training period, the total cumulative reward gradually increases and converges in a good direction after training, as shown in Figure 4a–c. It should be noted that the best training performance is in Figure 4b, where the learning rate was determined to be 0.003. The average total reward's evolution curve exhibits a swifter and more consistent convergence rate, along with a higher convergence value. Consequently, a learning rate of 0.003 was selected for the other experiments.

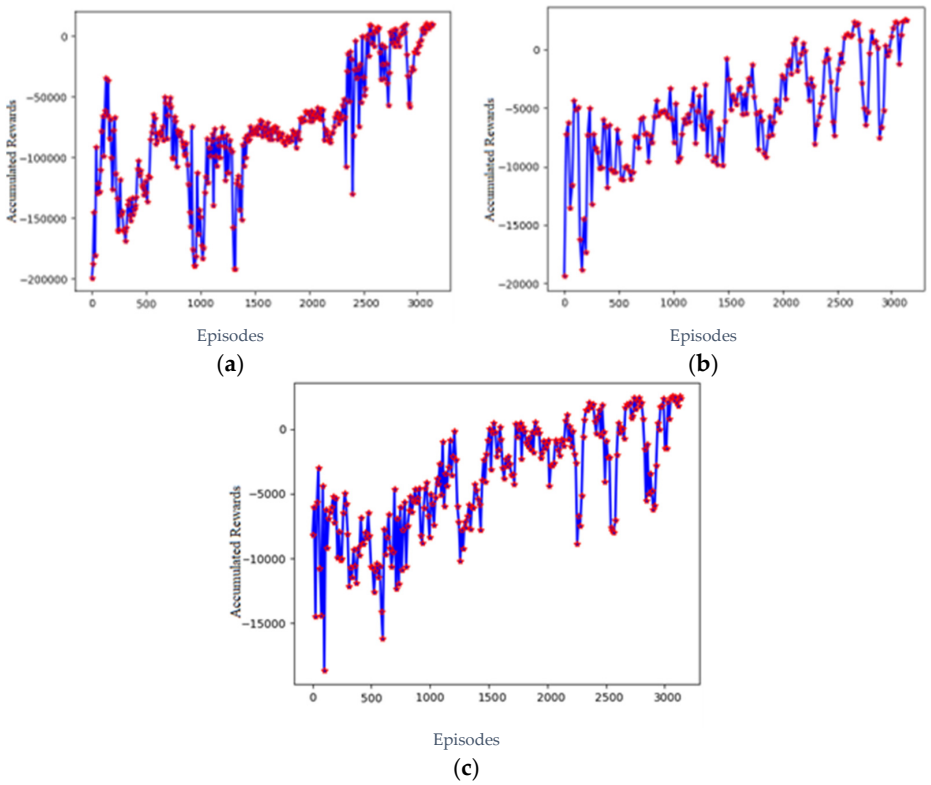

Figure 4. Change curves of cumulative reward at different learning rates: (**a**) the learning rate λ of DQN algorithm is 0.03; (**b**) the learning rate λ of DQN algorithm is 0.003; (**c**) the learning rate λ of DQN algorithm is 0.0003.

Figure 5 shows the change curve of the total cumulative reward with the number of training episodes at a learning rate of 0.003 and a training period of 10,000 episodes. It also illustrates the change curve of the cumulative reward after the algorithm has converged. The other parameters of the DQN algorithm are the same as those shown in Table 3.

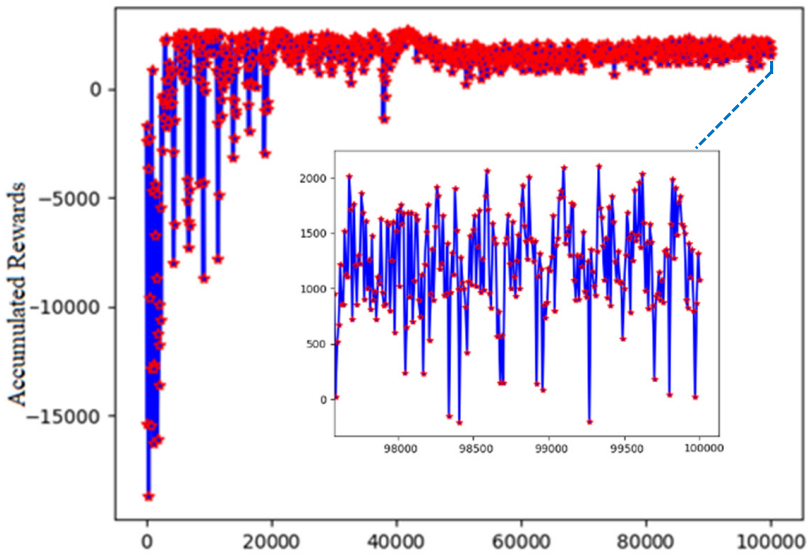

Figure 5. Change curve of total cumulative reward of the DQN algorithm with λ = 0.003.

It can be seen from Figure 5 that in the 10,000-episode training experiment, the fluctuation range of the cumulative reward gradually decreases as the number of training episodes increases. From the change curve of the cumulative reward, it seems that the cumulative reward was maintained at a significant level within a designated range upon practice when executed about 5000 times, and the average cumulative reward converged to about 1400. After training was executed 10,000 times, the average cumulative reward remained at 1450, indicating that the proposed algorithm is convergent.

4.2.2. Practical Application Performance Test

Figure 6a,b exhibit the outputs of the air braking force and electric braking force during the control process of the DQN algorithm, respectively. The curve in Figure 7 shows the train speed controlled by the DQN algorithm. With the above parameter, i.e., the index Done was 1 with a learning rate of 0.003, the speed of the train was kept within the limitation range during the operation. This means that the agent had learned a safe and feasible control strategy during the test simulations to ensure the safe operation of the heavy-haul train. Moreover, it is evident from Figures 6 and 7 that during the process of train control, cyclic air braking was engaged when the train was running on the long and steep downhill section. Specifically, air braking was not applied during the initial 80 s. Since the overall braking force was not sufficient to decelerate the heavy-haul train, the train speed increased during this period. In the following 100 s, air braking and electric braking were utilized concurrently ($h^a = 1$ and $h^d = 0.8$), resulting in the train's velocity dropping from 76 km/h to 52 km/h. Subsequently, air braking was discontinued, while electric braking was maintained constantly for the next 100 s ($h^a = 0$ and $h^d = 0.8$). At 210 s, the train speed reached 78 km/h. At this moment, air braking was reactivated to reduce the train's velocity ($h^a = 1$ and $h^d = 0.8$).

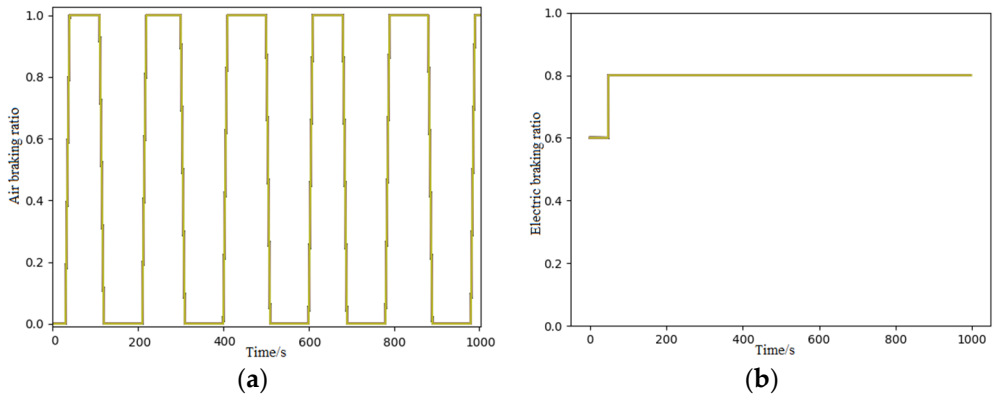

Figure 6. Output of brake command during braking: (**a**) output of air braking; (**b**) output of electric braking.

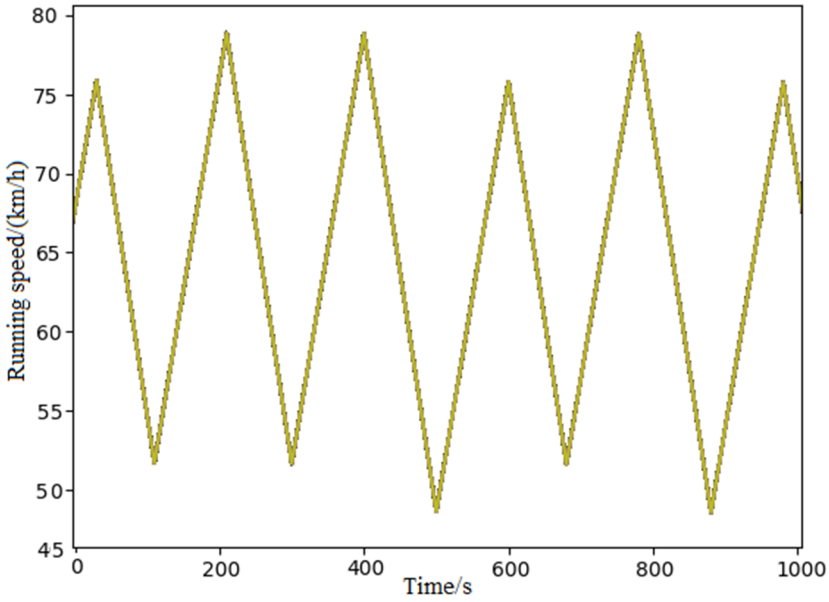

Figure 7. Control Strategy for Running Speed of The Heavy-Haul Train.

Herein, an intelligent air braking strategy employing the DQN algorithm was used to implement five braking cycles for a heavy-haul train running on a long and steep downhill railway section. This approach ensured the safety of train operation. Furthermore, Figure 7 unequivocally illustrates that the train's safe operation indicator Y maintains a value of 1, which signifies that the train velocity consistently remains within the predefined safety limits during its entire operation. Figure 7 shows that the DQN agent in the experiment learned a safe and feasible control strategy to ensure the safe operation of heavy-haul trains. It can be seen that after training, the algorithm is able to control the train speed by applying the air brakes before the train reaches the maximum speed of 80 km/h under the premise of a slope speed of 68 km/h, and by releasing the air brakes before the speed drops below the minimum release speed of 40 km/h after braking. This is done to ensure the train can maintain safe operation until it exits the long downhill section.

4.2.3. Comparison of Algorithm Performance

To prove the superiority of the DQN algorithm in dealing with the cyclic braking of a heavy-haul train running on a long and steep downhill section, the optimization results with the above settings were compared with those of the Q-learning algorithm [10]. The key parameters of the Q-learning algorithm were as follows: the learning rate λ was 0.05, the

maximum number of iterations M was 100,000, the discount rate γ was 0.95, the acquisition probability ε was 0.1, and the state transition interval Δt was 10.

It can be seen from Table 4 that compared with the results of the Q-learning algorithm from [10], the proposed algorithm based on DQN is superior in terms of braking distance, braking efficiency and operation efficiency during the cyclic braking of a heavy-haul train running on a long and steep downhill section. The safe operation of the train is, thus, effectively ensured.

Table 4. Comparison of simulation results.

Algorithm	Target				
	Safety Indicator Y	Air Braking Distance/m	Planned Running Time/s	Actual Running Time/s	Average Speed/(km/h)
Q learning	1	7576.9	1500	1360	63
DQN	1	7417.6	1500	1260	68

From Table 4, it can be observed that under the same conditions and with the same hyperparameter settings, the DQN algorithm, which is focused on air braking distance, actual running time and the train's average speed, shows superiority in braking distance and running efficiency when heavy-haul trains perform cyclic air braking on long downhill sections. Our findings reveal that the DQN-based intelligent control strategy decreased the air braking distance by 2.1% and enhanced the overall average speed by over 7%. These results unequivocally demonstrate the efficacy and superiority of the DQN-based intelligent control strategy. Our experimental results indicate that the DQN algorithm proposed in this paper effectively ensures the safety of train operation.

5. Conclusions

In this article, we explore the optimal working condition (braking and braking release) transition during cyclic braking when a heavy-haul train is running on a long and steep downhill section. Aiming at obtaining the shortest air brake distance and the highest operating efficiency, various constraints of the actual operation are considered at the same time, including factors such as the air-refilling time of the auxiliary reservoir, operating speed and operation action switch. The main conclusions are as follows:

(1) To achieve the optimization of multiple objects, a mathematical model is established for a heavy-haul train running on a long and steep downhill section. An intelligent cyclic braking system design based on the DQN algorithm is introduced for the train to adapt to a variety of complex operation environments and line conditions. To improve the convergence speed of the algorithm, the priority experience replay mechanism is used instead of ordinary experience replay. By prioritizing experiences, the agent can choose experiences with a higher priority for learning. In this way, the agent can more quickly learn important experiences that are more conducive to the rapid convergence of the algorithm. As a result, it improves the performance of the control algorithm.

(2) To verify the performance of the proposed DQN algorithm, comparative simulations were carried out and tested with different parameters. The simulation results show that the DQN algorithm proposed in this article exhibits better optimization performance and can effectively generate train driving speed curves that fulfill the specified constraints. This provides a valuable reference for the application of cyclic braking in heavy-haul trains running on long and steep downhill sections.

This study primarily focuses on the intelligent control of a cyclic air braking strategy for heavy-haul trains. However, during the research phase, we failed to comprehensively consider all environmental factors that could affect braking effectiveness. In particular, weather conditions (such as temperature, humidity and wind speed), track conditions (such as track flatness and friction coefficient) and variations in train load were not within the

scope of our research. In future, we plan to conduct in-depth research on these environmental factors to evaluate the braking performance of heavy-haul trains more comprehensively under different conditions. In addition, the proposed DQN algorithm could be further improved and a more efficient network structure could be developed. This, in turn, could improve the performance of cyclic air braking with respect to heavy-haul trains.

Author Contributions: Conceptualization, C.Z.; Formal Analysis, J.H.; Funding Acquisition, C.Z.; Investigation, L.J.; Methodology, S.Z.; Resources, L.J.; Software, S.Z.; Validation, J.H.; Writing—Original Draft, S.Z.; Writing—Review and Editing, L.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Nos. 62173137, 52172403, 62303178) and the Project of the Hunan Provincial Department of Education, China (Nos. 23A0426, 22B0577).

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

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

Nomenclature

M	Sum of the masses of all carriages	$\dot{v}$	Acceleration of heavy-haul train
F	Locomotive traction force	v	Running speed of train
U_1	Output electric braking force	U_2	Output air brake force
u^d_{max}	Maximum electric brake force	u^{tr}_{max}	Maximum traction force
F_R	Resistance of train	g	Gravity acceleration
$\overline{V}$	Upper limit of train running speed	V^r_{min}	Minimum release speed of air braking
h^{tr}	Relative output ratio of traction force	h^d	Relative output ratio of the electric braking force
θ_b	Equivalent emergency brake ratio of air braking	φ_b	Equivalent friction coefficient of air braking
R	Curve radius	L_s	Tunnel length
β_s	Service brake coefficient of air braking	i	Gradient of the line section
$\varphi_1, \varphi_2, \varphi_3$	Running resistance constant	L_a	Air brake distance
t^b_{j+1}	Time point of engaging air brake in the $(j+1)$th cycle	t^r_j	Time point of releasing air brake in the jth cycle

References

1. Zhang, Z. *Optimization Analysis of Smooth Operation for Ten-Thousand Ton Trains of Shuohuang Railway*; Southwest Publishing House: Chengdu, China, 2017.
2. Lu, Q.; He, B.; Wu, M.; Zhang, Z.; Luo, J.; Zhang, Y.; He, R.; Wang, K. Establishment and analysis of energy consumption model of heavy-haul train on large long slope. *Energies* **2018**, *11*, 965. [CrossRef]
3. Dong, S.; Yang, S.; Yang, C.; Ni, W. Analysis of Braking Methods for Express Freight Train on Long Ramp. *Mod. Mach.* **2022**, *2*, 28–34.
4. Kuefler, A.; Morton, J.; Wheeler, T.; Kochenderfer, M. Imitating Driver Behavior with Generative Adversarial Networks. In Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (IV), Los Angeles, CA, USA, 11–14 June 2017; IEEE: New York, NY, USA, 2017.
5. Yang, H.; Wang, Y.; Li, Z.; Fu, Y.; Tan, C. Expert supervised SAC reinforcement learning for optimizing the operation of heavy-duty trains. *Control Theory Appl.* **2022**, *39*, 799–808.
6. Howlett, P. An optimal strategy for the control of a train. *Anziam J.* **1990**, *31*, 454–471. [CrossRef]
7. Huang, Y.; Su, S.; Liu, W. optimization on the Driving Curve of Heavy Haul Trains Based on Artificial Bee Colony Algorithm. In Proceedings of the 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), Rhodes, Greece, 20–23 September 2020; IEEE: New York, NY, USA, 2020.
8. Du, L. Model Based Security Reinforcement Learning. Master's Thesis, Harbin Institute of Technology, Harbin, China, 2021.
9. Wang, Y.; De Schutter, B.; van den Boom, T.J.; Ning, B. Optimal trajectory planning for trains—A pseudo spectral method and a mixed integer linear programming approach. *Transp. Res. Part C Emerg. Technol.* **2013**, *29*, 97–114. [CrossRef]

10. Zhang, M.; Zhang, Q.; Zhang, Z. A Study on Energy-Saving Optimization for High Speed Railways Based on Q-learning Algorithm. *Railw. Transp. Econ.* **2019**, *41*, 111–117.
11. Zhang, M.; Zhang, Q.; Liu, W.; Zhou, B. A Policy-Based Reinforcement Learning Algorithm for Intelligent Train Control. *J. China Railw. Soc.* **2020**, *42*, 69–75.
12. Zhang, W.; Sun, X.; Liu, Z.; Yang, L. Research on Energy-Saving Speed Curve of Heavy Haul Train Based on Reinforcement Learning. In Proceedings of the 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC), Bilbao, Bizkaia, 24–28 September 2023; IEEE: New York, NY, USA, 2023; pp. 2523–2528.
13. Sandidzadeh, M.A.; Havaei, P. A comprehensive study on reinforcement learning application for train speed profile optimization. *Multimed. Tools Appl.* **2023**, *82*, 37351–37386. [CrossRef]
14. Wu, T.; Dong, W.; Ye, H.; Sun, X.; Ji, Y. A Deep Reinforcement Learning Approach for Optimal Scheduling of Heavy-haul Railway. *IFAC-PapersOnLine* **2023**, *56*, 3491–3497. [CrossRef]
15. Lin, X.; Liang, Z.; Shen, L.; Zhao, F.; Liu, X.; Sun, P.; Cao, T. Reinforcement learning method for the multi-objective speed trajectory optimization of a freight train. *Control Eng. Pract.* **2023**, *138*, 105605. [CrossRef]
16. Tang, H.; Wang, Y.; Liu, X.; Feng, X. Reinforcement learning approach for optimal control of multiple electric locomotives in a heavy-haul freight train: A Double-Switch-Q-network architecture. *Knowl.-Based Syst.* **2020**, *190*, 105173.
17. Wang, M.Y.; Kou, B.Q.; Zhao, X.K. Analysis of Energy Consumption Characteristics Based on Simulation and Traction Calculation Model for the CRH Electric Motor Train Units. In Proceedings of the 2018 21st International Conference on Electrical Machines and Systems (ICEMS), Jeju, Republic of Korea, 7–10 October 2018; IEEE: New York, NY, USA, 2018; pp. 2738–2743.
18. Yu, H.; Huang, Y.; Wang, M. Research on operating strategy based on particle swarm optimization for heavy haul train on long down-slope. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; IEEE: New York, NY, USA, 2018; pp. 2735–2740.
19. Huang, Y.; Bai, S.; Meng, X.; Yu, H.; Wang, M. Research on the driving strategy of heavy-haul train based on improved genetic algorithm. *Adv. Mech. Eng.* **2018**, *10*, 1687814018791016. [CrossRef]
20. Su, S.; Tang, T.; Li, X. Driving strategy optimization for trains in subway systems. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* **2018**, *232*, 369–383. [CrossRef]
21. Niu, H.; Hou, T.; Chen, Y. Research on Energy-saving Operation of High-speed Trains Based on Improved Genetic Algorithm. *J. Appl. Sci. Eng.* **2022**, *26*, 663–673.
22. Zhang, Z. *Train Traction Calculation*; China Railway Press: Beijing, China, 2013.
23. Su, S.; Wang, X.; Cao, Y.; Yin, J. An energy-efficient train operation approach by integrating the metro timetabling and eco-driving. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 4252–4268. [CrossRef]
24. Ma, H. Research on Optimization Control of Long Downhill Braking Process for Heavy Haul Trains. Master's Thesis, East China Jiaotong University, Nanchang, China, 2019.
25. Wu, J. *Train Traction Calculation*; Southwest Jiaotong University Press: Chengdu, China, 2013; pp. 50–141.
26. Wei, W.; Jiang, Y.; Zhang, Y.; Zhao, X.; Zhang, J. Study on a Segmented Electro-Pneumatic Braking System for Heavy-Haul Trains. *Transp. Saf. Environ.* **2020**, *2*, 216–225. [CrossRef]
27. Su, S.; Tang, T.; Xun, J.; Cao, F.; Wang, Y. Design of Running Grades for Energy-Efficient Train Regulation: A Case Study for Beijing Yizhuang Line. *IEEE Intell. Transp. Syst. Mag.* **2019**, *13*, 189–200. [CrossRef]