information

Article

Cyclic Air Braking Strategy for Heavy Haul Trains on Long Downhill Sections Based on Q-Learning Algorithm

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

College of Electrical and Information Engineering, Hunan University of Technology, Zhuzhou 412000, 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 factor affecting the safe operation of trains on long downhill sections. However, a train's cycle braking strategy is constrained by multiple factors such as driving environment, speed, and air-refilling time. A Q-learning algorithm-based cyclic braking strategy for a heavy haul train on long downhill sections is proposed to address this challenge. First, the operating environment of a heavy haul train on long downhill sections is designed, considering various constraint parameters, such as the characteristics of special operating routes, allowable operating speeds, and train tube air-refilling time. Second, the operating status and braking operation of a heavy haul train on long downhill sections are discretized in order to establish a Q-table based on state–action pairs. The training of algorithm performance is achieved by continuously updating Q-tables. Finally, taking the heavy haul train formation as the study object, actual line data from the Shuozhou–Huanghua Railway are used for experimental simulation, and different hyperparameters and entry speed conditions are considered. The results show that the safe and stable cyclic braking of a heavy haul train on long downhill sections is achieved. The effectiveness of the Q-learning control strategy is verified.

Keywords: heavy haul train; long steep downhill; cyclic braking; Q-learning; intelligent control

Citation: Zhang, C.; Zhou, S.; He, J.; Jia, L. Cyclic Air Braking Strategy for Heavy Haul Trains on Long Downhill Sections Based on Q-Learning Algorithm. *Information* **2024**, *15*, 271. https://doi.org/10.3390/info15050271

Academic Editor: Xiao-Fang Liu

Received: 25 March 2024
Revised: 2 May 2024
Accepted: 9 May 2024
Published: 11 May 2024

1. Introduction

Heavy haul trains have a large transportation capacity, high efficiency, and low transportation costs; thus, they have received widespread attention from countries worldwide. To control speed when heavy haul trains operate on long downhill sections, the braking system must increase cyclic air braking [1]. The existing strategy used for air braking mainly relies on the conductor's experience, which is insufficient for meeting the safety and efficiency requirements for heavy haul train operation [2]. Therefore, an intelligent control strategy must be developed to improve the air braking performance of heavy haul trains on long downhill sections [3].

With the rapid development of heavy haul trains, scholars have recently conducted extensive research on the cyclic braking of heavy haul trains on long downhill sections. Related methods can be summarized as mechanistic model-, machine learning-, and reinforcement learning-based methods.

In terms of mechanistic model-based methods, a neural network-based air braking model was proposed to accurately predict pressure changes in the key components of train air braking systems [4]. In [5], a new hybrid model of long short-term memory (LSTM) was developed to describe the changes in control force. In [6], a long short-term memory model with delayed information was constructed to solve the problem of deep learning models being unable to explain the impact of model inputs on system outputs. A real-time slope estimation model based on Kalman filtering was constructed for the electric and air braking system of heavy haul trains [7]. Traditional physical-driven models usually fail to reflect the "true" dynamics of heavy haul trains because of the strong nonlinearity and uncertainty

in the mechanistic model due to air resistance, frequently switching working conditions, and variations in external influencing factors such as weather and temperature. During heavy haul train operations, a large amount of data are accumulated, providing support for the research on data-driven circulating air braking strategies for heavy haul trains.

In terms of machine learning-based research, an intelligent driving strategy for heavy haul trains based on expert knowledge and machine learning is proposed, in order to determine feasible air pressure reduction and the exact time to apply and release air brakes [8]. In [9], an optimization model for the operation of heavy-duty trains was established, achieving optimal control while maximizing operating distance and minimizing air braking time. In [10], to address the issue of the severe imbalance in the proportion of operating data for heavy haul trains under different working conditions, a random forest algorithm was used to extract data and establish a model for automatic air brakes. In [11], based on a train dynamics model, the model parameters—including energy consumption, running time, and distance of pneumatic braking—were optimized, and the artificial bee colony (ABC) algorithm was introduced to find reasonable switching points for different states.

Reinforcement learning can be used to handle large-scale state spaces and dynamically changing environments, and is characterized by a strong real-time decision-making ability. Reinforcement learning has received widespread attention in studies on braking strategies for heavy haul trains. In [12], a long downhill section operation optimization method suitable for long-formation heavy haul trains was developed to improve the braking performance of 20,000-ton heavy haul trains. In [13], a deep reinforcement learning method with a reference system was constructed, which satisfies the constraints on speed, time, and position during train operation and reduces the tracking errors of reinforcement learning. In [14], a double-switch Q-network (DSQ network) architecture was designed to solve the problem of the optimal control of multiple electric locomotives in heavy haul trains. However, fully using the massive amounts of data generated by trains during operation is a key issue in reinforcement learning methods that needs to be addressed.

The Q-learning algorithm is a widely recognized and extensively used reinforcement learning method. It not only boasts a solid theoretical foundation, but also features a relatively simple application process. Additionally, it has demonstrated excellent performance in numerous practical scenarios, providing strong practical support for its utilization in the field of heavy haul train braking. Significantly, the Q-learning algorithm has a unique advantage in handling discrete action spaces, making it well-suited to address the challenges faced by heavy haul trains operating on long downhill sections. Based on these comprehensive considerations, we developed a cyclic air braking strategy for heavy haul trains on long downhill sections based on the Q-learning algorithm. The main contributions of this study are as follows:

(1) A heavy haul train model with operational constraints was constructed, considering the vehicle's characteristics on long and steep slopes of railway lines, as well as heavy haul trains equipped with traditional pneumatic braking systems. In addition, with the optimization objectives of safe train operation and operational efficiency, a Q-learning algorithm-based cycle braking strategy for heavy haul trains on long downhill sections was developed under constraints such as interval speed limits and air-refilling time.
(2) Simulations and experiments were conducted under actual heavy haul train operating conditions, and the experimental results were compared under different parameters and ramp speeds. The experimental results showed that the proposed intelligent control strategy performs well in various scenarios, demonstrating its effectiveness and practicality in train braking.

The rest of this article is organized as follows: A model for heavy haul trains operating on long downhill sections is described in Section 1, introducing constraints on train operation and the performance indicators of train operation. Section 2 introduces the method of circulating air braking for heavy haul trains based on the Q-learning algorithm. The effectiveness and robustness of the proposed method were verified through simulation

experiments, as described in Section 3. Finally, Section 4 provides a summary of the study and outlines prospects for future research.

2. Heavy Haul Train Model

2.1. Dynamics Model

During the operation of heavy haul trains, various factors such as track gradient, train formation, and on-board mass exert diverse forces on each train. However, in this study, the forces of the interactions between carriages were not considered in the calculation of additional resistance. As a result, the forces acting on the train during its operation primarily comprised locomotive traction, braking force (including electric and pneumatic braking), fundamental running resistance, and additional resistance. According to the principles of Newtonian dynamics, the mathematical expression of each train model can be formulated as follows:

$$M\dot{v} = F - B_1 - B_2 - F_R. \tag{1}$$

Generally, the running resistance F_R encountered by a heavy haul train during braking on a long and steep downhill slope 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 [15].

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

According to previous research [16], the formula for calculating the basic resistance of a heavy haul train is as follows:

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

The additional resistance is determined by the slope force g_R, curvature resistance c_R, and the tunnel resistance t_R [16], as shown in Equation (4). The specific calculations [17] of these factors are given in Equation (5).

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

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

To facilitate the understanding, the main symbols are introduced in Table 1.

Table 1. Description of symbols used in the train model.

Symbol	Description	Symbol	Description
M	Sum of the masses of all carriages	$\dot{v}$	Acceleration of the heavy haul train
R	Curve radius	L_s	Tunnel length
B_1	Output electric brake force	B_2	Output air brake force
u^d_{max}	Maximum electric brake force	u_a	Air brake force
b^a	Binary variable of air braking	b^d	Relative output ratio of the electric brake force
F_R	Resistance of train	g	Gravity acceleration
$\varphi_1, \varphi_2, \varphi_3$	Running resistance constant	L_a	Air brake distance
v	Running speed of train	V^r_{min}	Minimum release speed of air brake
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
i	Gradient of the track on which the train is running	V_{max}	Upper limit of train running speed

2.2. Running Constraints

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

When a train operates on a long and steep downhill section, cyclic braking is adopted for speed control. To ensure sufficient braking force in the next braking cycle, sufficient time is required to refill the air pipe to full pressure [18]. That is, the duration of the release phase shall not be less than the minimum air filling time T_a specified by the operating procedures.

$$t_{j+1}^{b} - t_{j}^{r} \geq T_a, \tag{6}$$

where t_{j+1}^{b} represents the time point where the air brake is engaged in the (j + 1)th cycle, and t_{j}^{r} indicates the time point where the air brake is released in the j-th cycle. T_a is closely related to the formation of the train and the pressure drop in the train air pipe. For fixed train parameters, the air-filled time under a certain pressure drop needs to be determined. A longer train and a larger pressure reduction generally require longer air-filled time.

To ensure safety, the speed of a heavy haul train cannot exceed the speed limit V_{max} at any point on long and steep downhill line sections. This value often depends on the infrastructure of the railway line or the temporary setup during operation. Additionally, train speed must be greater than the minimum air brake release speed V_{min}^{r}. The specified limit is designated as 40 km/h for a 10,000-ton heavy haul train formation [19]. Therefore, the speed should meet the following requirement:

$$\mathrm{V}_{\mathrm{min}}^{\mathrm{r}} \leq v \leq V_{\mathrm{max}}. \tag{7}$$

Regarding the optimization objectives of this study, the operation of a heavy haul train on long downhill sections is also constrained by the relative output ratio of the braking force [20]. The two main types of braking devices used for heavy haul trains include variable resistance and pneumatic braking systems. Variable resistance braking, also known as regenerative braking, can feed energy back to other locomotives to provide power. Pneumatic braking systems produce braking force by reducing the air pressure in the train's air brake pipe [21].

The output electric brake force B_1 of a heavy haul train depends on the maximum electric brake force $u_{max}^{d}(v)$ and the relative output ratio b^d. The output pneumatic braking force depends on whether air braking is applied. Therefore, a train's braking force can be expressed as

$$\begin{cases} B_1 = b^d \mathrm{u}_{\mathrm{max}}^{\mathrm{d}}, 0 \leq b^d \leq 1 \\ B_2 = b^a u^a, b^a = 0 \; or \; b^a = 1' \end{cases} \tag{8}$$

where the maximum electric brake force u_{max}^{d} is a piecewise function related to the operating speed [22], and the air braking force u_a is a function of the air pressure drop [23].

2.3. Performance Indicators

This study mainly focuses on the safety and maintenance cost of the heavy haul train operation process. The maintenance cost is expressed as the air-braking distance. Hence, two indicators were used to evaluate the control of the heavy haul train.

- Safety: Safety is a prerequisite for train operation. The running speed of a heavy haul train must be kept under the upper limit but cannot be lower than V_{min}^{r}. Here, K is defined in order to indicate whether the train's speed remains within the speed limit.

$$K = \begin{cases} 1, \; \mathrm{V}_{\mathrm{min}}^{r} \leq v \leq \mathrm{V}_{\mathrm{max}} \\ 0, \; \text{otherwise} \end{cases}. \tag{9}$$

- Air-braking distance: As excessive wear is caused by the friction between the wheels and brake shoes when the air brake is engaged for a long distance, the replacement of air brake equipment increases maintenance costs. 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

$$L_a = \int_0^T b^a(t) * v(t)dt. \tag{10}$$

3. Algorithm Design

Reinforcement learning is a machine learning approach tailored for goal-oriented tasks [24]. Unlike traditional methods, reinforcement learning does not instruct the agent on how to act, but rather guides the agent through interactions with the environment to learn the correct strategies [25]. In this section, we first define the train operation process as a Markov decision process (MDP). Second, we describe a control algorithm based on Q-learning that learns the cyclic braking strategy for long downhill sections.

3.1. Markov Decision Process

Before applying the Q-learning algorithm, the process of controlling train operation on a long and steep downhill slope needed to be defined as a Markov decision process (MDP), that is, the formalization of sequential decision-making [24,26]. A schematic diagram of the MDP interaction of a heavy haul train running on a long and steep downhill slope is shown in Figure 1.

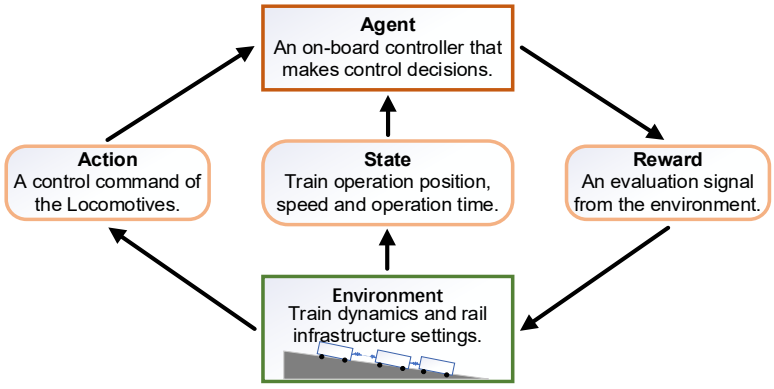

Figure 1. Schematic diagram of MDP interaction during operation of heavy haul trains.

As shown in Figure 1, the MDP consists of five elements: the agent, environment, action, state, and reward. A heavy haul locomotive is defined as an agent that makes control decisions. The heavy haul train dynamics and railway infrastructure settings are defined as the environment. During the interaction process, the agent performs actions based on the environment, and the environment responds to the agent, with new heavy haul train states and reward signals based on operational constraints. Therefore, location, speed, and operating time are defined as the states of the heavy haul train.

$$\mathrm{s}_k = [P_k, V_k, T_k], k = 0, 1, 2, \cdots, n, \tag{11}$$

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 train speed, and T_k is the train's running time.

The control action is defined as the setting of the relative electric brake force and air brake notch.

$$a_k = [b_k^a, b_k^d], k = 0, 1, \cdots, n, \tag{12}$$

where b_k^a is a binary variable representing the air brake control command, and b_k^a is the relative output ratio of the electric brake force output by the train locomotive, which is limited by the constraint condition in Equation (8).

The control output of a heavy haul train in each period is determined only by the speed, position, and operating time of the train. Thus, the process of controlling a heavy haul train can be exactly defined using reinforcement learning as a Markov decision-making process [26], which is expressed as follows:

$$s_0 \xrightarrow{a_0} s_1, r_1 \xrightarrow{a_1} \cdots s_k, r_k \cdots \xrightarrow{a_{n-2}} s_{n-1}, r_{n-1} \xrightarrow{a_{n-1}} s_n. \tag{13}$$

3.2. Q-Learning Algorithm

In this section, the Q-learning algorithm-based intelligent control method is described for heavy haul trains operating on steep downhill slopes. The Q-learning algorithm is a reinforcement learning algorithm that learns in an environment without prior knowledge. Based on the principle of temporal difference control, the agent continuously updates the Q-value function through interactions with the environment. Using the Q-value as the evaluation criterion, the algorithm iteratively seeks the optimal action to maximize the expected total reward obtained during the interaction with the environment. The iteration process of the Q-learning algorithm involves learning the optimal actions from the Markov decision process (MDP). In a single simulation process, Q-learning updates the Q-values in real time to form new strategies for the next simulation, as shown in the control process diagram in Figure 2.

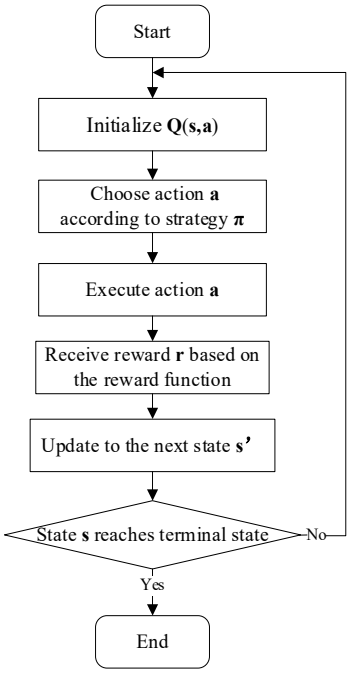

Figure 2. Flowchart of controller for Q-learning algorithm.

(1) Randomly initialize Q (s, a), $\forall s \in S, a \in A(s)$.
(2) According to the ε-greedy policy π and the current state s, action a is selected from the Q-table. Execute action a as determined by the decision-making process; then, obtain the reward value r by interacting with the environment and proceed to the next state. Update the Q-Table, i.e., s→s′; continue until the termination state is reached.
(3) By following this procedure, after multiple iterations, the optimal policy and the optimal state–action value function can both be obtained.

When the number of algorithm iterations reaches a certain quantity, the termination condition is met. The generation of the optimal policy is no longer determined by the greedy

policy, but is based on selecting actions according to the optimal Q-values corresponding to each state at each time, forming the optimal policy.

3.2.1. Policy Design

To ensure that the algorithm balances exploration and exploitation capabilities, an ε-greedy policy is adopted, defining the agent's behavior at a given time step. Formally, the policy is a function that outputs the probability of selecting each possible action relative to the Q-function. It can be represented as the following:

$$\pi(a|s_k) = \begin{cases} (1-\varepsilon) + \frac{\varepsilon}{|\mathrm{A}(s_k)|}, a = a^* \\ \frac{\varepsilon}{|A(s_k)|}, a \neq a^* \end{cases}, \tag{14}$$

where $|A(s_k)|$ is the number of actions in the action set when the state is s_k, $a^* = \mathrm{argmax}_a Q(s,a)$, $\varepsilon \in (0,1)$.

Specifically, using the ε-greedy policy to select control actions during the train operation process involves randomly choosing actions with a probability of ε, and adopting the action with the highest estimated Q-value with a probability of $1-\varepsilon$. This approach enhances the algorithm's global search capability.

3.2.2. Reward Function Design

The optimization goal of the reinforcement learning problem is reflected by the reward function. For the train control process in question, to ensure safe operation, the operating speed cannot exceed the upper limit. Therefore, the constraint in Equation (10) must be satisfied. If the speed is higher than the upper limit V_{max} or lower than the minimum remission speed V^r_{min}, a negative reward R_c is given to the agent. If the air brake is engaged by a heavy haul train at step k, a zero reward is given. Otherwise, a positive reward R_d is given to encourage the release of the air brake. Therefore, the award is defined as follows:

$$r_{k+1} = \begin{cases} \mathrm{R}_c, V_{k+1} < V^r_{\min} \; or \; V_{k+1} > V_{\max} \\ 0, h^a_k = 1 \; and \; V^r_{\min} \leq V_{k+1} \leq V_{\max} \\ R_d, \; h^a_k = 0 \; and \; V^r_{\min} \leq V_{k+1} \leq V_{\max} \end{cases}. \tag{15}$$

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

Algorithm 1: The Q-learning-based control strategy for cyclic air braking of the heavy haul train.

///Initialization///
1: Initialize Q function Q (s, a) randomly.
///Training process//
2: for episode = 1, . . . M do
3: Initialize the state s_0 of the train.
4: for k = 0, 1, . . ., N − 1 do
5: Select action **a** according to ε-greedy policy $\boldsymbol{\pi}$.
6: Perform action **a**; receive rewards **r** and the next state of the train s'
7: Update the Q-Table through the equation, that is,
$\mathrm{Q}(s,a) \leftarrow \mathrm{Q}(s,a) + \alpha[r + \gamma \mathrm{max}_{a'} \mathrm{Q}(s\prime, a) - \mathrm{Q}(s,a)]$
8: Update the next state of train, s←s′
9: end for
10: end for
11: Output the well-trained Q-Table
///Online control process///
12: Initialize the state s_0 of the train
13: for k = 0, 1, . . ., N − 1 do
14: According to $a = \mathrm{argmax}_a Q(s)$ select action.
15: Perform action **a** and obtain the next state s′
16: end for

4. Algorithm Simulation and Analysis

This section describes the simulation experiments that were conducted using real data from the Shuozhou–Huanghua Railway in China. First, the setup of the experimental parameters and the data are introduced. Second, the experimental results are presented and analyzed in three main parts: the model training process, effectiveness testing in practical applications, and robustness testing of the algorithm.

4.1. Experimental Parameter Settings

To validate the effectiveness of the proposed intelligent control algorithm, simulation experiments were conducted using "1 Locomotive + 100 Wagons" in combination with the route data from the Shuohuang Railway in China. Our aim was to obtain a speed tracking curve for a heavy haul train on long downhill sections. The train consisted of HXD1 electric locomotives and C80 freight cars, with the specific train parameters shown in Table 2. The total length of the train route was S = 20,000 m, and the slope of the route mostly ranged from 10‰ to 12‰, which complied with the requirements of the Technical Management Regulations for Chinese Railways for long downhill sections. Additionally, the speed limit on this route was 80 km/h, and specific route data are provided in Table 3. The hyperparameters for the Q-learning algorithm were set as shown in Table 4.

Table 2. 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

Table 3. 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	19,130–20,000	10.9
6750–12,430	11.3		

Table 4. Algorithm hyperparameters.

Parameter	**Value**	**Parameter**	**Value**
Maximum training episode M	100,000	Minimum air-refilling time T_{AI}	50
Discount rate γ	0.95	Learning rate λ	0.001
Initial value of ε	0.98	Final value of ε	0.1
Positive reward R_d	5	Negative reward R_c	-50
Minimum braking speed $V^r_{\min}$	30 km/h	Maximum braking speed $V^r_{\max}$	80 km/h

4.2. Simulation Experiment Verification

4.2.1. Model Training Process

Using the parameter settings described above, the proposed Q-learning algorithm was validated through simulation experiments. During this study, the learning rate λ of the Q-learning algorithm was defined, and the sensitivity of this parameter was analyzed. Three groups of experiments were set up, with λ values set to 0.0001, 0.001, and 0.01, separately. The iterative Q-learning process in each group of experiments was observed. The initial

speed V0 of the heavy haul train entering the long downhill section was set to 40 km/h. Other hyperparameters were set according to Table 4. The more interactions between the reinforcement learning agent and the environment, the richer the experience, and the more accurate the strategies. During training, the agent and the environment interacted 1 million times, including 100,000 episodes. For each episode, the total reward value corresponding to the solution generated based on Q-values was recorded. The cumulative reward change curve of the optimized algorithm is shown in Figure 3.

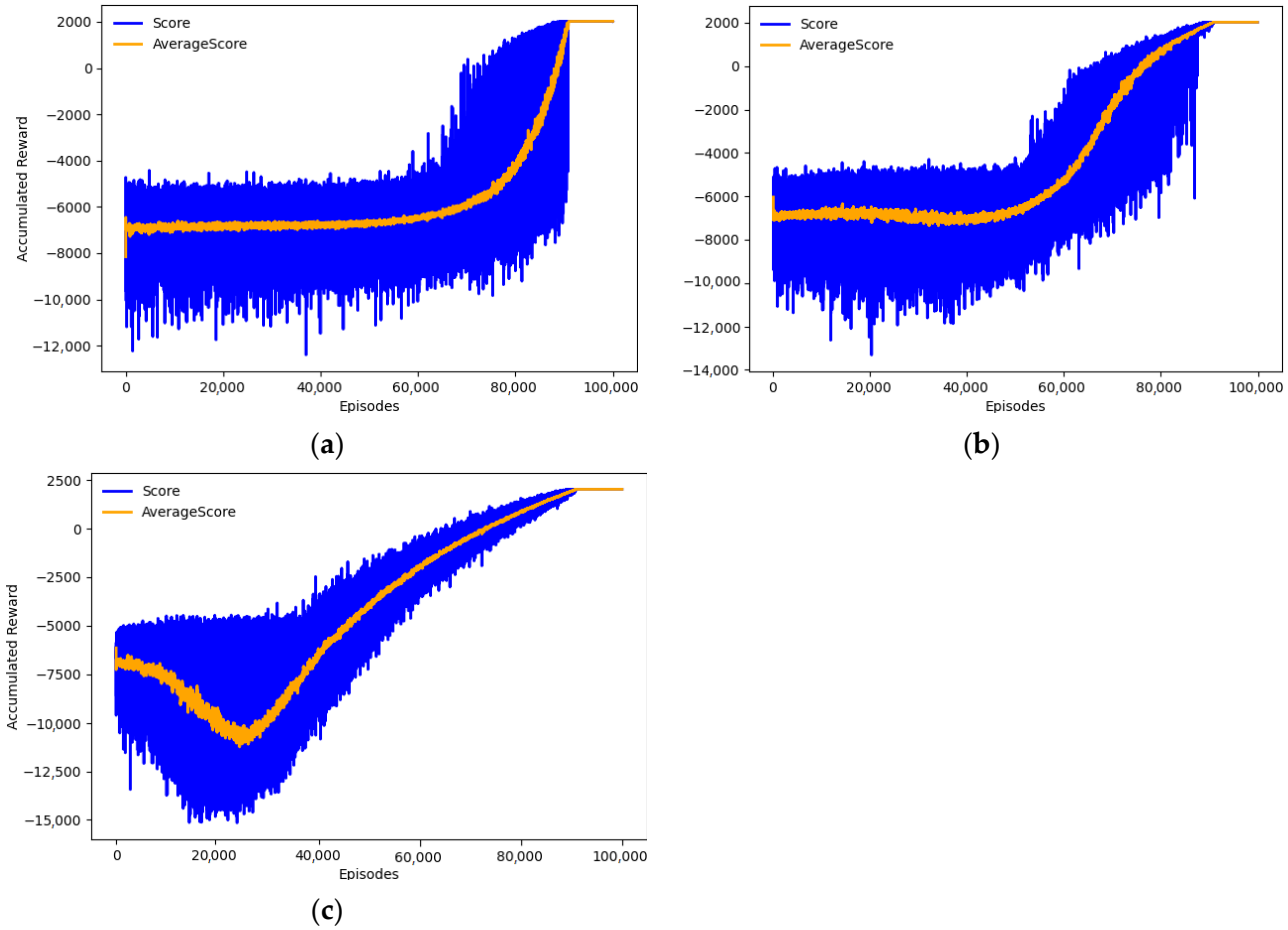

Figure 3. The cumulative reward change curves of the algorithm for different learning rates: (**a**) the learning rate λ of Q−learning algorithm is 0.0001; (**b**) the learning rate λ of Q−learning algorithm is 0.001; (**c**) the learning rate λ of Q−learning algorithm is 0.01.

The best training performance was achieved for the experiment depicted in Figure 3b. Compared with the other two groups of experiments, when λ = 0.001, the cumulative reward change curve of the Q-learning algorithm exhibited a faster and more stable convergence rate, as well as a higher convergence value. Therefore, we recommend setting the learning rate λ to 0.001 during training. Owing to the presence of the ε-greedy policy in the Q-learning algorithm, the agent initially randomly explores during training, and the action selection during decision-making is random. Consequently, the optimization space for Q-values is large, resulting in relatively small reward values and optimization effects. As exploration proceeds, the agent gradually learns the correct braking strategy, and the cumulative reward value continuously increases. As training progresses, the control policy optimization of the Q-learning algorithm tends to stabilize and approach the optimal state. The agent tends to adopt the optimal action with the maximum Q-value, leading to a stable cumulative reward curve, which indicates convergence of the Q-learning algorithm's iterations.

4.2.2. Effectiveness Testing of Practical Application

After training the Q-learning algorithm, the effectiveness of the control algorithm in periodically braking the train was verified, considering different entry states of the train and the state transition process. The Q-values corresponding to different entry states and state transitions were read directly from the Q-table. The effectiveness of the control algorithm for periodic braking of the train was validated. The speed tracking curves of the train under different entry speeds are shown in Figure 4. When the entry speed is 30 km/h, the train adopts a three-cycle braking optimization strategy through reinforcement learning training. Similarly, when the entry speed is 40 km/h, the train also adopts a three-cycle braking optimization strategy. However, when the entry speed is 50 km/h, due to the increase in entry speed, the train adopts a four-cycle braking optimization strategy. Additionally, according to the simulation results, when the train runs on a long downhill section, air braking tends to be applied at the maximum speed limit during the cyclic braking process, and the braking is released appropriately at the right time. Despite the train entering the downhill section at the three different entry speeds mentioned above, the running speed of the train increases. This is because, on long downhill sections, trains tend to initially maintain a coasting state to save energy and ensure a higher running speed, and then apply braking at the appropriate time.

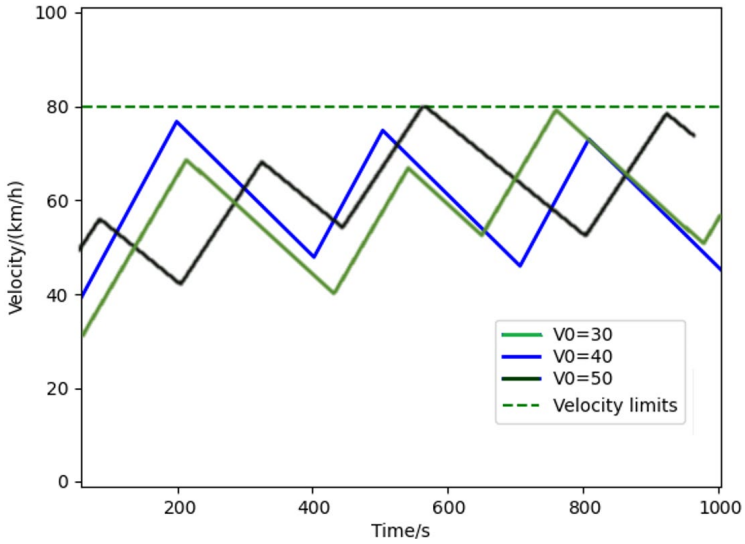

Figure 4. Periodic braking strategies of heavy haul trains at different entry speeds.

Figure 4 shows that after training, for the three entry speeds mentioned above, the reinforcement learning agent can control train speed by applying air braking before reaching the maximum speed limit. This ensures that the train remains within a safe operating speed range until exiting the section. This indicates that the Q-learning algorithm can effectively train agents to develop good control strategies, keeping the train speed within the speed limits and maintaining a relatively high average speed. This validates the effectiveness of the algorithm.

4.2.3. Performance Comparison Experiment

To verify the robustness of the Q-learning algorithm in controlling heavy haul trains through cyclic braking on long downhill sections, the optimized results for different entry speeds were compared. The key parameters of the Q-learning algorithm were set as follows: the learning rate λ was 0.001; the maximum number of iterations M was 100,000; the discount factor γ was 0.95; the exploration rate ε was 0.1; and the state transition time interval Δt was 50.

Table 5 shows that under different conditions, the Q-learning algorithm with the same hyperparameter settings—aiming to optimize air braking distance, running time, and running efficiency—shows robustness in braking distance and braking efficiency when heavy haul trains perform cyclic air braking on long downhill sections. Additionally, the Safety Indicator K documented in Table 5 reveals a significant degree of stability in braking performance for algorithms utilizing fixed hyperparameters, regardless of the specific environmental conditions encountered. This dependability is paramount in guaranteeing the safety of heavy-duty trains during downhill braking, effectively mitigating the risks associated with speed loss or ineffective braking methods, thereby minimizing the potential for accidents. Ultimately, the empirical data strongly corroborate the efficacy of the proposed Q-learning algorithm in securing safe train operations.

Table 5. Comparison of simulation results.

V_0/(km/h)	Target				
	Safety Indicator K	Air Braking Distance/m	Planned Running Time/s	Actual Running Time/s	Average Speed/(km/h)
30	1	9843.6	1000	1074.6	67
40	1	10,181.3	1000	1014	71
50	1	10,547.4	1000	993.2	72

5. Conclusions

In this study, a cyclic air braking strategy for heavy haul trains on long downhill sections based on the Q-learning algorithm is proposed. Aiming to minimize air brake distance and maximize operating efficiency, various constraints on actual train operation were simultaneously considered, including factors such as the air-filled time of the reservoir, the operating speed, and the operation–action switch. Our main conclusions are as follows:

(1) For heavy haul trains running on long and steep downhill sections, a multi-objective optimization model under multiple constraint conditions was constructed. A Q-learning algorithm with a finite Q-Table was introduced. The train states were discretized, and HXD1 locomotives and C80 freight trains were compared as the study objects for simulation verification. The proposed method enabled the train to adapt to complex train operating environments and route conditions. By adjusting the Q-learning algorithm hyperparameters, the convergence speed of the algorithm was improved while ensuring safe train operation.

(2) To validate the performance of the proposed Q-learning algorithm, comparative experiments were conducted under different parameter conditions. The experimental results demonstrated that the proposed Q-learning algorithm exhibits a stable optimization performance and effectively generates train speed profiles that satisfy constraints, providing a valuable reference for the intelligent assisted driving of heavy haul trains on long downhill sections.

In future work, the impacts of other environmental factors on the cyclic air braking method for heavy haul trains should be further explored—for example, considering different weather conditions, track conditions, or train load conditions, and how to optimize control in these complex environments. Additionally, introducing neural networks or combining other algorithms with Q-learning may further improve the performance of the cyclic air braking method for heavy haul trains.

Author Contributions: Funding acquisition, C.Z. and J.H.; resources, C.Z. and L.J.; validation, C.Z. and L.J.; writing—original draft, S.Z.; writing—review and editing, C.Z. and S.Z. 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 (grant nos. 62173137, 52172403, 62303178) and the Project of Hunan Provincial Department of Education, China (grant 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.

References

1. Jiang, B.; Tian, C.; Deng, J.; Zhu, Z. China's railway train speed, density and weight in developing. *Railw. Sci.* **2022**, *1*, 131–147. [CrossRef]
2. Liu, W.; Su, S.; Tang, T.; Cao, Y. Study on longitudinal dynamics of heavy haul trains running on long and steep downhills. *Veh. Syst. Dyn.* **2022**, *60*, 4079–4097. [CrossRef]
3. Hu, Y.; Ling, L.; Wang, K. Method for analyzing the operational performance of high-speed railway long ramp EMUs. *J. Southwest Jiaotong Univ.* **2022**, *57*, 277–285.
4. Zhou, X.; Cheng, S.; Yu, T.; Zhou, W.; Lin, L.; Wang, J. Research on the air brake system of heavy haul trains based on neural network. *Veh. Syst. Dyn.* **2023**, 1–19. [CrossRef]
5. Xu, K.; Liu, Q. Long and Short-Term Memory Mechanism Hybrid Model for Speed Trajectory Prediction of Heavy Haul Trains. In Proceedings of the 2023 IEEE Symposium Series on Computational Intelligence (SSCI), Mexico City, Mexico, 5–8 December 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 476–481.
6. Yin, J.; Ning, C.; Tang, T. Data-driven models for train control dynamics in high-speed railways: LAG-LSTM for train trajectory prediction. *Inf. Sci.* **2022**, *600*, 377–400. [CrossRef]
7. Tian, C.; Wu, M.; Zhu, L.; Qian, J. An intelligent method for controlling the ECP braking system of a heavy-haul train. *Transp. Saf. Environ.* **2020**, *2*, 133–147. [CrossRef]
8. Wang, X.; Li, S.; Tang, T.; Wang, H.; Xun, J. Intelligent operation of heavy haul train with data imbalance: A machine learning method. *Knowl.-Based Syst.* **2019**, *163*, 36–50. [CrossRef]
9. Su, S.; Huang, Y.; Liu, W.; Tang, T.; Cao, Y.; Liu, H. Optimization of the speed curve for heavy-haul trains considering cyclic air braking: An MILP approach. *Eng. Optim.* **2023**, *55*, 876–890. [CrossRef]
10. Wei, S.; Zhu, L.; Chen, L.; Lin, Q. An adaboost-based intelligent driving algorithm for heavy-haul trains. *Actuators* **2021**, *10*, 188. [CrossRef]
11. 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: Piscataway, NJ, USA, 2020; pp. 1–6.
12. Wang, J.; Wang, C.; Zeng, Z. Research on operation optimization of heavy-haul combined trains in long and steep downhill sections based on reinforcement learning. *Electr. Drive Locomot.* **2023**, *6*, 139–146. [CrossRef]
13. Shang, M.; Zhou, Y.; Fujita, H. Deep reinforcement learning with reference system to handle constraints for energy-efficient train control. *Inf. Sci.* **2021**, *570*, 708–721. [CrossRef]
14. 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. [CrossRef]
15. Wang, X.; Su, S.; Cao, Y.; Qin, L.; Liu, W. Robust cruise control for the heavy haul train subject to disturbance and actuator saturation. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 8003–8013. [CrossRef]
16. Wu, Q.; Yang, X.; Jin, X. Hill-starting a heavy haul train with a 24-axle locomotive. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* **2022**, *236*, 201–211. [CrossRef]
17. Zhu, Q.; Su, S.; Tang, T.; Xiao, X. Energy-efficient train control method based on soft actor-critic algorithm. In Proceedings of the 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), Indianapolis, IN, USA, 19–22 September 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 2423–2428.
18. Ge, X.; Ling, L.; Chen, S.; Wang, C.; Zhou, Y.; Xu, B.; Wang, K. Countermeasures for preventing coupler jack-knifing of slave control locomotives in 20,000-tonne heavy-haul trains during cycle braking. *Veh. Syst. Dyn.* **2022**, *60*, 3269–3290. [CrossRef]
19. 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]
20. Zhang, Q.; Zhang, Y.; Huang, K.; Tasiu, I.A.; Lu, B.; Meng, X.; Liu, Z.; Sun, W. Modeling of regenerative braking energy for electric multiple units passing long downhill section. *IEEE Trans. Transp. Electrif.* **2022**, *8*, 3742–3758. [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. Ma, Q.; Yao, Y.; Meng, F.; Wei, W. The design of electronically controlled pneumatic brake signal propagation mode for electronic control braking of a 30,000-ton heavy-haul train. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* **2024**, *238*, 268–279. [CrossRef]
23. He, J.; Qiao, D.; Zhang, C. On-time and energy-saving train operation strategy based on improved AGA multi-objective optimization. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit.* **2024**, *238*, 511–519. [CrossRef]
24. Palm, H.; Arndt, L. Reinforcement Learning-Based Hybrid Multi-Objective Optimization Algorithm Design. *Information* **2023**, *14*, 299. [CrossRef]

25. Wang, X.; Wu, C.; Xue, J.; Chen, Z. A method of personalized driving decision for smart car based on deep reinforcement learning. *Information* **2020**, *11*, 295. [CrossRef]
26. Wang, R.; Zhuang, Z.; Tao, H.; Paszke, W.; Stojanovic, V. Q-learning based fault estimation and fault tolerant iterative learning control for MIMO systems. *ISA Trans.* **2023**, *142*, 123–135. [CrossRef] [PubMed]