Hindawi
Advances in Civil Engineering
Volume 2021, Article ID 7619214, 18 pages
https://doi.org/10.1155/2021/7619214

Research Article

Research on Vibration Reduction Control Based on Reinforcement Learning

Rongyao Yuan,[1] **Yang Yang**,[2,3] **Chao Su**,[1] **Shaopei Hu**,[1] **Heng Zhang**,[1] **and Enhua Cao**[1]

[1]*College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210098, China*
[2]*Power China Kunming Engineering Corporation Limited, Kunming 650051, China*
[3]*Department of Hydraulic Engineering, Tsinghua University, Beijing 100084, China*

Correspondence should be addressed to Yang Yang; yangyhhu@foxmail.com and Chao Su; csu_hhu@126.com

Received 10 June 2021; Accepted 22 June 2021; Published 2 July 2021

Academic Editor: Loke Foong

Magnetorheological (MR) dampers, as an intelligent vibration damping device, can quickly change the damping size of the material in milliseconds. The traditional semiactive control strategy cannot give full play to the ability of the MR dampers to consume energy and reduce vibration under different currents, and it is difficult to control the MR dampers accurately. In this paper, a semiactive control strategy based on reinforcement learning (RL) is proposed, which is based on "exploring" to learn the optimal value of the MR dampers at each step of the operation, the applied current value. During damping control, the learned optimal action value for each step is input into the MR dampers so that they provide the optimal damping force to the structure. Applying this strategy to a two-layer frame structure was found to provide more accurate control of the MR dampers, significantly improving the damping effect of the MR dampers.

1. Introduction

MR damper is currently the fastest research and development field [1]. It is a new type of intelligent damping device that uses MR effects. It can be well combined with the control system and it is reliable. At present, it has become a new generation of civil engineering structural vibration damping devices and has been initially used in civil engineering structural vibration control. Ji et al. [2] proposed a scheme of using MR damping technology to control pipeline vibration. The test results show that all three pipeline vibration control methods based on MR dampers can effectively control pipeline vibration. The pipeline amplitude and acceleration reduced to 22.31 dB and 16.34 dB, respectively, while the amplitude attenuation rate and acceleration attenuation rate reached 92.34% and 84.77%, respectively. Ghasemikaram et al. [3] verified that the performance of the MR damper is suitable for controlling extreme cyclic oscillations of wings and external storage devices under flutter conditions. Based on on-site monitoring of wind and rain excitation events, Ni et al. [4] used two MR dampers to control the vibration of the Dongting Lake cable-stayed bridge. Abdeddaim et al. [5] used a MR damper to link two adjacent buildings and effectively reduced the vibration response of the structure.

Because the parameterized model of MR damper is not as complicated as the nonparametric model in terms of structure, its research and application are extensive. At present, the commonly used dynamic model of MR damper is Bingham model [6], Bouc–Wen model [7], and modified Bouc–Wen model [8]. The research [9] shows that, compared with the Bingham model, the Bouc–Wen model can accurately reflect the nonlinear performance of the MR damper at low speeds and the hysteresis characteristics of the simulated MR damper and has strong versatility; compared with the modified Bouc–Wen model, it has fewer parameters and is easy to digitally model. It has been widely used in the modeling of the dynamic characteristics of MR damper. Therefore, this article will use the Bouc–Wen model to calculate the damping force.

MR damper controls the damping force by adjusting the magnitude of current or voltage. MR damper vibration control methods mainly include active control, passive control, and semiactive control. Research [5, 10–13] shows that active control and semiactive control have better damping effects than passive control. However, compared to active control, semiactive control methods can change the stiffness and damping of the structure with a small amount of energy to reduce the vibration response of the structure [14, 15]. Since semiactive control has both the excellent control effect of active control and the simple advantages of passive control, it also overcomes the shortcomings of active control that requires a lot of energy and the narrow tuning range of passive control. Therefore, semiactive control has great prospects for research and application development. In the semiactive control of the MR damper, in order to apply the optimal control force to the control structure, the control current or voltage of the MR damper needs to be calculated by the control system through a semiactive control strategy, which attracts a large number of scholars. Bathaei et al. [16] used two different fuzzy controllers to study the seismic vibration of the adaptive MR damper, which can further reduce the maximum displacement, acceleration, and foundation shear force of the structure. Hazaveh et al. [17] used discrete wavelet transform (DWT), linear quadratic regulator (LQR), and limiting optimal control algorithm to determine the optimal control force. The semiactive control performance is evaluated by comparing the maximum displacement, total base shear force, and control energy. Kim [18] proposed two semiactive control methods for seismic protection of structures using MR dampers. The first method is a simple adaptive control method, and the second method is a fuzzy control method based on genetic algorithms. The results show that it can control the displacement and acceleration response of the structure effectively. The control strategies proposed above can improve the effect of vibration reduction to varying degrees. However, due to the nonlinearity of the MR damper, the above algorithm cannot control the MR damper accurately.

RL plays an important role in solving the optimal control of complex linear systems with unknown models and it is combined with control theory to form adaptive dynamic programming theory which is a data-driven intelligent control algorithm with learning and optimization capabilities. What is more, it has rich theoretical research results in the fields of robust control, optimal control, and adaptive control. The value-based RL algorithm obtains the optimal value function, selects the action corresponding to the maximum value function, and implicitly constructs the optimal strategy. Representative algorithms include Q-learning [19] and SARSA. Q-learning has a process of selecting the maximum value, which is more suitable for optimal control than SARSA. In 2014, Brodley and Health [20] proposed a deterministic strategy gradient algorithm. In 2015, Littman made a review of RL in "Nature" [21]. Littman and Michael et al. [22] used Q-learning algorithm to realize autonomous movement control of robots. Hara et al. [23] used machine learning control algorithms to control robots and improve learning efficiency.

In this paper, we propose a semiactive control strategy based on RL and apply this strategy to the two-layer framework. The results show that it has a significant improvement in vibration reduction effect compared to the semiactive control strategy based on simple Bang-Bang.

The remainder of this paper is structured as follows. Section 2 describes the principle and model of MR damper. Section 3 introduces the principles of RL. In Sections 4 and 5, the semiactive control strategy based on RL is proposed and applied it to a two-layer frame structure. Compared with simple Bang-Bang, it is obvious that this strategy is better. Based on the above results, we conclude in Section 6.

2. MR Damper Model

2.1. Mechanical Model of MR Damper. MR damper is made of MR fluid, which has the features of simple device, low energy consumption, fast response, large damping, and wide dynamic range. Its structure includes electrical control line, piston rod, piston, orifice, and buffer accumulator. So far, the mechanical model of MR damper can be roughly divided into two types: parametric model and nonparametric model. Since the nonparametric model has a very complex structure, scholars at home and abroad have fully considered the characteristics of different stages on the process of MR fluid yielding and the structural characteristics of the MR damper. Parametric models mainly include Bingham viscoplastic model, modified Bingham viscoplastic model, Bouc–Wen model, modified Bouc–Wen model, and phenomenological model. But the Bouc–Wen model is a model with a smooth transition curve, which can fit the test results well. This model is easy to carry out numerical calculation, has strong versatility, can reflect various hysteresis lines, and has been widely used in hysteretic system modeling. In this paper, the RD-8041-1MRD MR damper produced by the American Lord company is used for vibration reduction control research, and its Bouc–Wen model is

$$f = c_0\dot{x} + 1.31 \times (x + 20.75) + \alpha z. \tag{1}$$

The expression of variable z is

$$\dot{z} = -47.76|\dot{x}|z|z|^{0.24} - 36.37\dot{x}|z|^{1.24} + 53.17\dot{x}. \tag{2}$$

And its schematic diagram is shown in Figure 1.

x represents the displacement of the piston rod of the MR damper and the damping coefficient; c_0 is a constant. A is a coefficient determined by the control system and magnetic field MR fluid. By adjusting the parameters γ, β, and A, the linear behavior of the force-velocity curve during unloading and the smoothness of the transition from before to after yielding can be controlled. k_0 is linear spring stiffness. x_0 is the initial deformation of the spring. The results of the dynamic characteristics of the MR damper show that the model can describe the force-displacement relationship of the MR damper well, and the force-velocity relationship curve is closer to the test results.

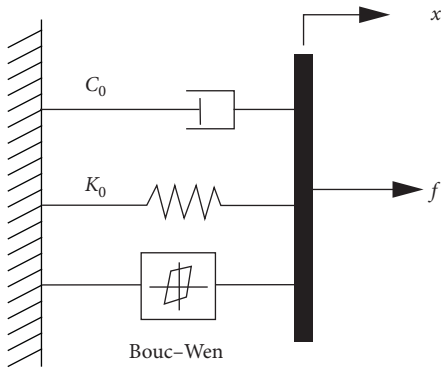

Figure 1: Schematic diagram of the Bouc–Wen model.

2.2. Semiactive Control Strategy of MR Damper. Under the control device equipped with MR dampers, the motion equation of a controlled system with n degrees of freedom subjected to vibration is

$$M_s\ddot{x}_g + C_s\dot{x} + K_s x = Ef - M_s I\ddot{x}_g, \tag{3}$$

where M_s, C_s, and K_s, respectively, represent the mass, damping, and stiffness matrix of the structure; x is the displacement vector of the structure relative to the ground; $\ddot{x}_g$ represents the one-dimensional ground acceleration, f is the control force vector generated by n MR dampers, I is unit column vector, and E is the position matrix of MR damping.

The corresponding state equation is

$$\dot{z} = Az + Bf + W\ddot{x}_g, \tag{4}$$

where z is the state vector. A, B, and W are state matrix, control device position indication matrix, and earthquake action matrix, respectively.

The semiactive control strives to achieve the optimal control force, so a semiactive control algorithm is needed to control the magnetorheological damper to apply the optimal control force F. The specific steps include the following:

(1) Obtain the displacement and speed of each incremental step of the control point through the URDFIL subroutine, and store the data in the global variable COMMON block at the same time.

(2) According to the displacement and speed, the semiactive control algorithm is used to calculate the optimal control force F of the magnetorheological damper based on the Bouc–Wen model, and the data is stored in the global variable COMMON block.

(3) Pass the control force F into the subroutine DLOAD through the global variable COMMON block, thereby applying the control force to the corresponding control area.

(4) Repeat the above process for each incremental step until the end of the program.

The current semiactive control algorithms mainly include simple Bang-Bang control algorithm, optimal Bang-Bang control algorithm, and limit Hrovat optimal control algorithm. This paper combines the commonly used simple Bang-Bang control algorithm to realize the semiactive control of the MR damper. The simple Bang-Bang control algorithm can be expressed as

$$c_d(t) = \begin{cases} c_{d\max}, & x\dot{x} > 0, \\ c_{d\min}, & x\dot{x} \le 0, \end{cases} \tag{5}$$

where x and $\dot{x}$ are the displacement and velocity of the damper piston rod relative to the cylinder; $c_{d\max}$is the maximum damping coefficient; $c_{d\min}$ is the minimum damper coefficient.

According to formula (5), the main operation of the simple Bang-Bang algorithm is when the structure deviates from the equilibrium position and vibrates, the magnetorheological damper applies the maximum damping coefficient to the structure; that is, the maximum current is used. When the structure vibrates to the equilibrium position, the magnetorheological damper applies a minimum damping coefficient to the structure; that is, the current is zero. Therefore, the simple Bang-Bang is equivalent to Passive-off and Passive-on control, and the actual damping of the control algorithm is between these two.

3. Principles of Reinforcement Learning

3.1. Basic Concepts of Reinforcement Learning. RL [24] is a branch of machine learning, which is used to solve continuous decision-making problems. RL can learn how to achieve the goals set by the task in a complex and uncertain environment [25].

Figure 2 shows the basic framework of RL. When the agent completes a task, it first interacts with the surrounding environment through actions. Under the action of action and environment, the agent will generate a new state. The environment will give an immediate return.

In this cycle, the agent continuously interacts with the environment and generates a lot of data. RL uses the generated data to modify its own action strategy and interacts with the environment to generate new data and then uses the new data to further improve its own behavior. After many iterations of learning, the agent finally learns how to complete the response, the optimal action for the task.

3.2. Markov Decision Processes. Markov decision process [26] is a framework that can solve most of the RL problems and has been widely used in various RL fields.

3.2.1. Markovian. Markovian means that, in a stochastic process, the next state of the system is only related to the current state and has nothing to do with the historical state and its actions. Setting $\{S_t, t \in T\}$ as a stochastic process, if for any moment $t_1 < t_2 < \cdots t_n \in T$, any state $S_1, S_2, \ldots, S_t \in E$. The conditional distribution function of the random variable S_t under the known variable $S_1, S_2, \ldots, S_t$ is only related to the current state S_t and has nothing to do with the previous state $S_1, S_2, \ldots, S_{t-1}$; that is, the conditional distribution function satisfies

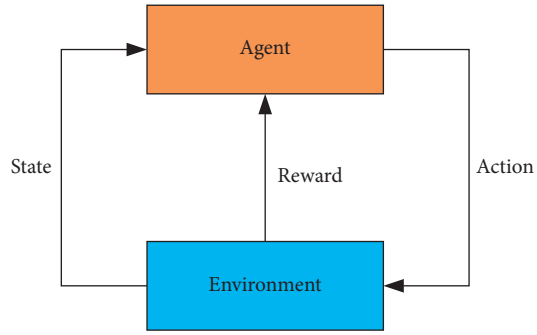

Figure 2: The basic framework of the reinforcement learning.

$$F(S_{t+1}|S_t) = F(\{S_{t+1}|S_1, S_2, \ldots, S_t\}), \tag{6}$$

namely,

$$P[S_{t+1}|S_t] = P[S_{t+1}|S_1, \ldots, S_t]. \tag{7}$$

Therefore, when the state of the system at time t satisfies $P[S_{t+1}|S_t] = P[S_{t+1}|S_1, \cdots, S_t]$, the state satisfies the Markovian. Markovian is also called memorylessness; that is, the current state S_t contains all relevant historical information $S_1, S_2, \ldots, S_t$. Once the current state S_t is known, the historical information will be discarded.

3.2.2. Markov Decision Processes. Markovian describes the nature of each state S_t, If each state in the stochastic process $\{S_t, t \in T\}$ satisfies the Markovian, the stochastic process is called Markov Stochastic Process. The Markov process is a two-tuple (S, P). In the formula, S is the set of finite states and P is the state transition probability. The state transition probability matrix is defined as

$$P = \begin{bmatrix} P_{11} & \cdots & P_{1n} \\ \vdots & \ddots & \vdots \\ P_{n1} & \cdots & P_{nm} \end{bmatrix}. \tag{8}$$

The Markov Stochastic Process mainly describes the transfer relationship between states. In the transfer process, assigning different reward values to each process is the Markov Decision Process. The Markov Decision Process can be represented by a tuple (S, A, P, R, γ) in which S is a finite set of states, A is a limited set of actions, P is the probability of state transition, R is the reward function, and γ is the attenuation coefficient; the specific Markov Decision Process is described in Figure 3.

The agent whose initial state is S_0 selects an action a_0 from the action set A. After executing the action a_0, the agent transfers to the next state S_1 according to the state transition probability P. Then, when the next action a_1 is executed, it moves to S_2, and then the next action a_2 is executed until the last action is completed.

The goal of RL is to find the optimal strategy π to obtain the largest reward (Reward) given by a Markov Decision Process, so as to estimate the pros and cons of a strategy. It can be described as

$$\pi(a|s) = P[A_t = a|S_t = s]. \tag{9}$$

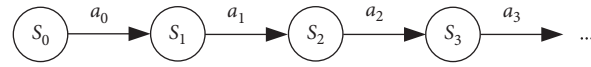

Figure 3: The dynamic process of Markov.

Strategy π specifies an action probability in each state s; the strategy π can specify a certain action in each state s. In the case of a certain strategy π, with a given strategy π, the cumulative return can be calculated, which can be defined as

$$G_t = R_{t+1} + R_{t+2} + \cdots = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}. \tag{10}$$

In order to evaluate the value of each state S_t in RL, it is necessary to use a definite quantity to describe the value of state S_t, but the cumulative return G_t is a random variable and cannot be used as a definite value to describe the value of the state. While its expected value is a certain value, therefore the expected value of cumulative returns G_t is used in RL to quantify the value of each state S_t.

3.2.3. State-Value Function. When the agent adopts strategy π, the cumulative return obeys a distribution, and the expected value of the cumulative return at state s is defined as a state-value function:

$$v_\pi(s) = E_\pi\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1}|S_t = s\right]. \tag{11}$$

The corresponding state-action value function can be expressed as

$$q_\pi(s, a) = \left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1}|S_t = s,\ A_t = a\right]. \tag{12}$$

The purpose of calculating the state-value function is to build a RL algorithm to get the optimal strategy from the data, each strategy corresponds to a state-value function, and the optimal strategy corresponds to the optimal state-value function, which is

$$v^*(s) = \max_\pi q_\pi(s). \tag{13}$$

The optimal state-action value function A is the largest state-action value function of all strategies, which can be expressed as

$$q^*(s, a) = \max_\pi q_\pi(s, a). \tag{14}$$

If the optimal state-action value function is known, the optimal strategy can be directly expressed by maximizing $q^*(s, a)$, namely,

$$\pi^*(s) = \arg\max_{a \in A} q^*(s, a). \tag{15}$$

Therefore, RL is looking for a strategy $\pi^*(s)$ that can maximize the value function under any initial conditions s.

3.3. Semiactive Control RL Framework of MR Damper. The basic framework of reinforcement learning includes agent, environment, action, state, and reward. Based on the basic framework of reinforcement learning, this paper establishes a RL framework for semiactive control of MR. As shown in Figure 4, in this framework, the MR damper is used as the main body of learning—the intelligent body; the structure that needs vibration reduction control is used as the learning environment; the action corresponds to the MR damper applying damping force to the structure perform vibration reduction control; reward is an evaluation of the control effect of the structure through the evaluation function; the state describes the situation between the agent and the environment, and it is related to the environment and the agent. In the damping control of MR damper, the state of RL corresponds to the response of the structure.

The agent exists in the environment and takes actions on the environment. These actions will make the agent get corresponding rewards. The purpose of RL is to obtain a strategy through learning. Under this strategy, the agent can make appropriate actions at the right time and obtains the greatest reward. Corresponding to the RL framework of MR damper semiactive control, the MR damper is installed on the structure that needs vibration damping control, and the damping force is applied by the damper to control the vibration of the structure. For every action applied by the MR damper to the structure, the structure will produce a corresponding response. The response is used to determine whether the action of the MR damper reduces the structural response, and rewards are given according to the evaluation function. The reward may be positive or negative. Through repeated damping control of the structure, MR damping continuously explores and learns in the damping control process and finally learns an optimal control strategy.

3.4. RL Algorithm: Q-Learning Algorithm. RL algorithms mainly include model learning algorithms and model-free learning algorithms. Model algorithms are built on the condition that the various elements of the entire Markov process model are known, but when MR dampers are used to reduce structural vibration in the control process, it is difficult to know the elements in the Markov decision process corresponding to the task in advance. Therefore, when the model is unknown, that is, when the transition probability and reward function of the Markov decision process are unknown, RL adopts a model-free algorithm for learning. Model-free RL algorithms mainly include two types: one is Monte Carlo RL algorithm, and the other is time difference algorithm. The model-free algorithm is to obtain the optimal strategy through learning when the Markov Decision Process is unknown. The algorithm does not rely on the transition probability and reward model of previous experience [27].

Figure 5 shows the update method of the value function of different RL algorithms. From Figure 5, it can be seen that, in the Monte Carlo algorithm [28], a complete trajectory is needed to calculate a certain state-value function and update it, resulting in low algorithm efficiency. The Q-learning algorithm and the SARSA algorithm update the value function after each step of the strategy, so the efficiency is higher. In addition, the Q-learning algorithm [29] is superior to other algorithms in convergence and stability among the three algorithms, so this paper uses the Q-learning algorithm to study the semiactive control of the MR damper.

Q-learning is an off-policy temporal-difference method; it was first proposed by Watkins in his doctoral thesis in 1989. Q-learning uses the state-action reward value and $Q(s, a)$ as the estimation function. During the interaction, the value of the correct action keeps increasing; otherwise, the value will decrease. By comparing the Q value, the agent tends to choose the best action. The update formula is as follows:

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha\left[R_{t+1} + \gamma \max_a Q(S_{t+1}, a) - Q(S_t, A_t)\right], \tag{16}$$

where γ is the attenuation coefficient, which refers to the influence of future strategies on the present under the current strategy. The value range is $[0, 1]$; α is the learning rate, which refers to the ratio of the Q current learning value covering the old Q value; $\alpha = 0$ means that the value table is not updated, and $\alpha = 1$ means that the Q value table is completely updated. A is the action set and S is the state set.

The calculation process of Q-learning algorithm is shown in Algorithm 1. The main steps of Q-learning are as follows: (1) use the greedy strategy (ε – greedy) to select an action; (2) the agent takes action to obtain the reward and new state; and (3) update the value table using formula (16).

4. Study on the Effect of Different Vibration Reduction Control Methods

This section adopts the passive control method: Passive-off, Passive-on ($I = 0.4$ A), Passive-on ($I = 0.6$ A), and semiactive control method (simple Bang-Bang) to control the vibration of the structure. Figures 6–8 show the vibration reduction control effect of 4 control methods installed on the upper part of the first layer of pillars with 4 dampers. It can be seen from the figure that the semiactive control method has the best damping effect, and Passive-off has the worst damping effect, but Passive-on in the passive control algorithm also shows a better control effect. For Passive-on, by increasing the input current, the vibration reduction effect can be improved and the structural response can be reduced. In addition, it can be seen from Figure 8 that although the acceleration response of the simple Bang-Bang semiactive control is the smallest, the acceleration has a sudden change in local time, which is due to the simple Bang-Bang assumption that when the structure vibrates away from the equilibrium position, the MR damper adopts the maximum damping coefficient, that is, the maximum current; when the structure vibrates toward the equilibrium position, the MR damper adopts the minimum damping coefficient. That is, the current is zero. It is equivalent to Passive-off and Passive-on control, and the actual damping force of the simple Bang-Bang control algorithm jumps between these two. Therefore, the change of excessive damping force causes a sudden change in acceleration.

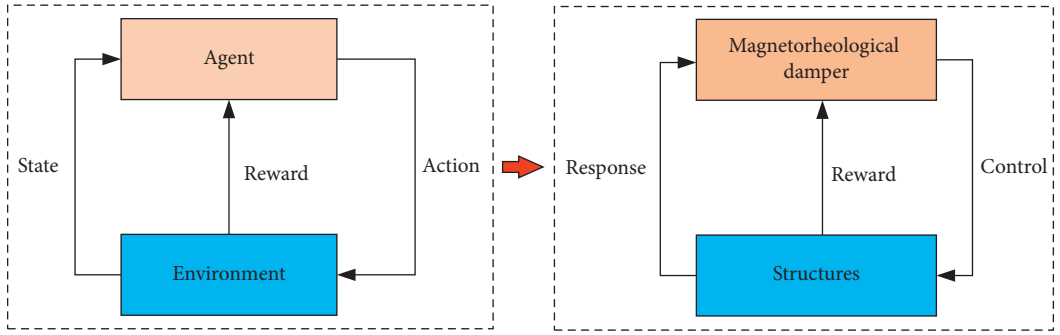

Figure 4: The framework of reinforcement learning for the semiactive control of MR damper.

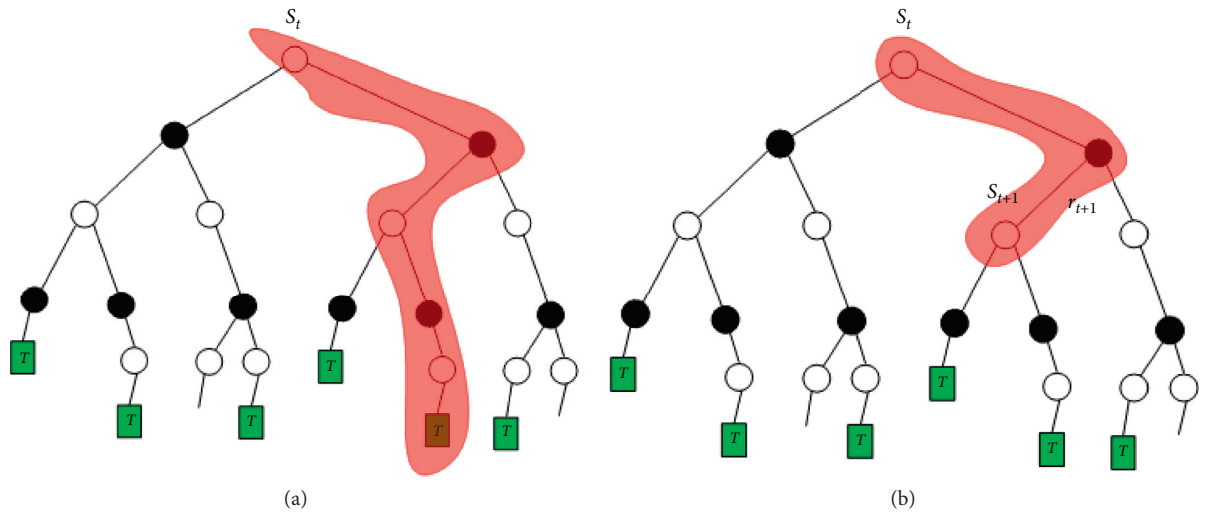

Figure 5: The updating method of value function for the different RL algorithms. (a) Monte Carlo RL. (b) Temporal-difference RL.

Input: environment E; action space A; initial state x_0; discount coefficient γ; learning rate α;
(1) $Q(s,a) = 0; \pi(s,a) = 1/|A(s)|;$
(2) $x = x_0$
(3) For $t = 1, 2, 3 \ldots$
(4) $r, s' =$ reward and transfer status generated by performing actions in environment E;
(5) $a' = \pi(s')$
(6) $Q(s,a) = Q(s,a) + \alpha(r + \gamma Q(s',a') - Q(s,a));$
(7) $\pi(s) = \arg\max Q(s,a'');$
(8) $s = s', a = a'^{a};$
(9) End
Output: strategy π

Algorithm 1: The updating method of value function for the different reinforcement learning algorithms.

5. Semiactive Control Strategy Based on RL Algorithm

5.1. Mission Details. In the semiactive control of the MR damper, the goal of the semiactive control algorithm is to calculate the optimal damping force to be applied to the structure, so as to achieve the optimal control effect [30]. Therefore, the goal of RL is to obtain the optimal damping force of each step of the MR damper through learning. According to the RL framework of semiactive control and Q-learning algorithm, this paper proposes a semiactive control strategy based on RL Q-learning algorithm. The specific principle is shown in Figure 9. The method includes two modules: the learning module and the semiactive control module. In the learning module, the system applies current to the MR damper through the Q-learning algorithm to control the damper to control the vibration of the structure. The system is based on the structure's response (speed,

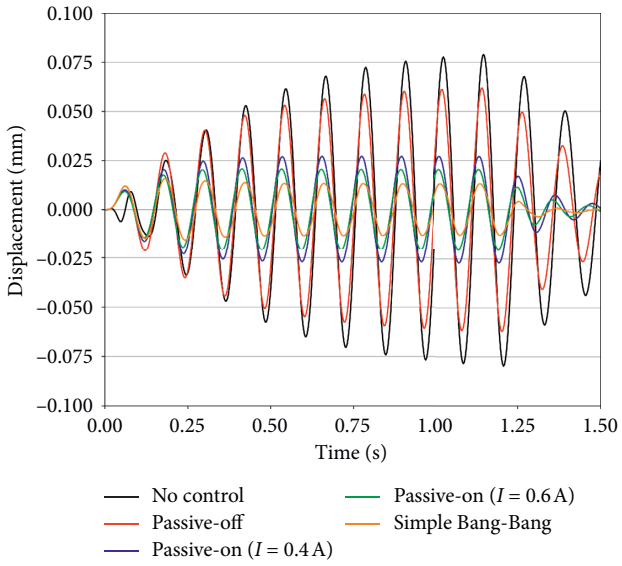

Figure 6: The time history of structural displacement response in the center point of the second floor for different control methods.

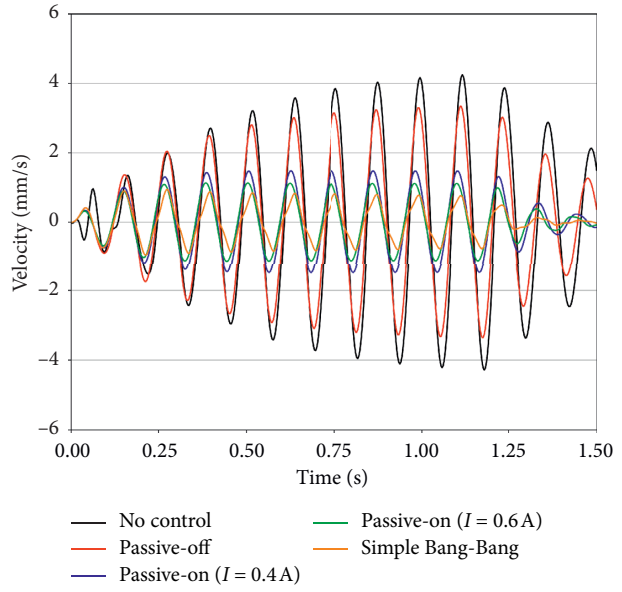

Figure 7: The time history of structural velocity response in the center point of the second floor for different control methods.

displacement, and acceleration) and the corresponding reward evaluation function; calculate the reward value for each action. Finally, in each state, the action with the largest value is selected as the optimal action in that state. After a certain period of learning, a Q value table is formed. The Q value table is a mapping pair of semiactive control strategies. Through the Q value table, the optimal control strategy, that is, the optimal control current, can be obtained. In the semiactive control module, the semiactive control of the structure by the MR damper can be realized by calling the control strategy that has been learned, that is, the current value learned at each step.

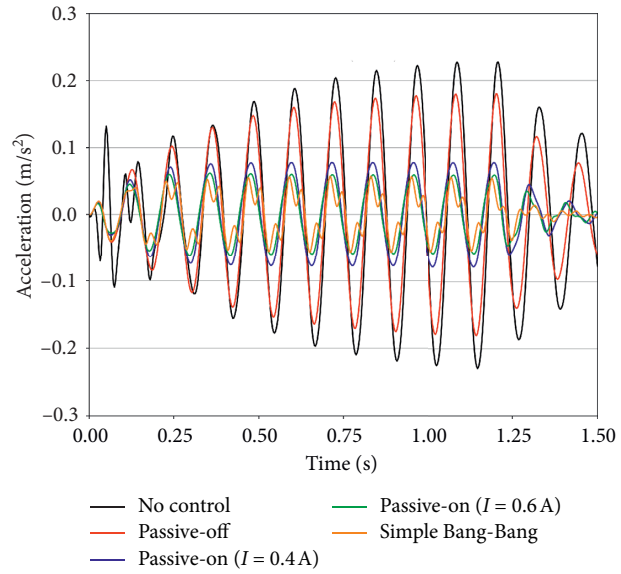

Figure 8: The time history of structural acceleration response in the center point of the second floor for different control methods.

5.2. Establishment of Reward Evaluation Function. The reward evaluation function is the feedback of the environment to the agent's decision in RL, reflecting the effect of the agent's actions on the environment. The MR damper selected in this paper is to adjust the damping force by controlling the applied current to change the size of the magnetic field. Therefore, for this type of MR damper, the application of current to the MR damper is used as the action of the agent in the RL. In the finite element analysis, the ratio of the response value U_{coni} of each incremental step to the response value U_{unconi} of each incremental step when the structure is controlled by the MR damper is used to reflect the reduction of each action. Therefore, the reward evaluation function $R(U_i)$ can be expressed as

$$R(U_i) = \begin{cases} -C\dfrac{U_{\text{coni}}}{U_{\text{unconi}}}, & \dfrac{U_{\text{coni}}}{U_{\text{unconi}}} > 1, \\ 0, & \dfrac{U_{\text{coni}}}{U_{\text{unconi}}} = 1, \\ C\dfrac{U_{\text{unconi}}}{U_{\text{coni}}}, & \dfrac{U_{\text{coni}}}{U_{\text{unconi}}} < 1, \end{cases} \tag{17}$$

where C is the reward magnification factor.

It can be seen from (17) that when $U_{\text{coni}}/U_{\text{unconi}} = 1$, it indicates that the response of the structure under the action does not change, and the action does not have the effect of damping vibration. At this time, the reward of the action is 0; when $U_{\text{coni}}/U_{\text{unconi}} > 1$, it indicates that the response of the structure under the action increases. Not only does the action fail to reduce the vibration, but it increases the response of the structure. Therefore, the reward (penalty) of the action is $-C(U_{\text{coni}}/U_{\text{unconi}})$. And

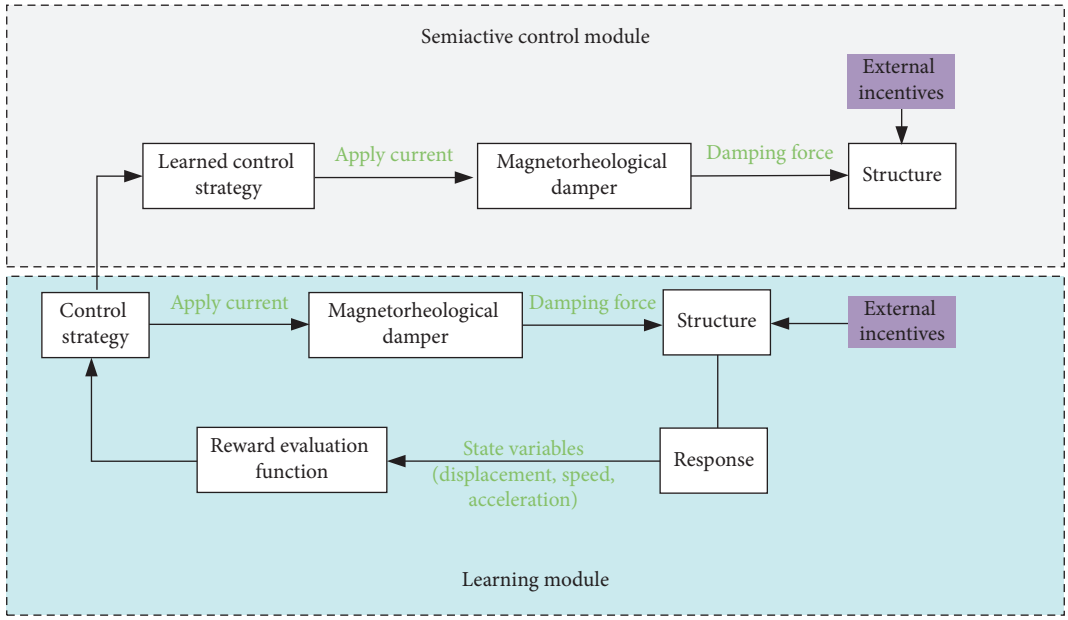

Figure 9: The schematic diagram of semiactive control based on the reinforcement learning algorithm.

when U_{coni}/U_{unconi}, it shows that the response of the structure under the action is reduced, and the action has a damping effect, so the reward is $C(U_{coni}/U_{unconi})$. Therefore, if an action reduces the structural response more, the reward for that action is higher. Conversely, if an action increases the structural response more, the penalty for that action is also higher. Through the reward evaluation function, the RL algorithm can give its reward score according to each action. Finally, the optimal action of each step is determined, thereby forming the optimal control strategy and realizing the optimal control of the MR damper.

5.3. Greedy Strategy. In the RL Q-learning algorithm, the algorithm selects an appropriate action based on the current state and value table. When choosing an action, two methods are generally used to select the appropriate action. The first is based on past "experience"; that is, the action with the highest score is selected every time you learn. The second method is to use the "exploratory" method, that is, to randomly choose an action each time you learn. In the two methods, if all "experience" is used and the action with the highest score is selected each time, it is likely to be confined to the existing experience and it is difficult to find more valuable actions. However, if only the "exploratory" method is used, and random selection of actions is used every time, most of the actions selected may be too low or worthless, resulting in slower convergence of the calculated value table.

Therefore, in order to balance the "experience" and "exploration" in the algorithm, scholars have proposed a method to effectively balance the two ε greedy. The mathematical expression of ε greedy strategy is

$$\pi(a|s) = \begin{cases} 1 - \varepsilon + \dfrac{\varepsilon}{|A(s)|}, & a = \arg\max_a Q(s,a), \\ \dfrac{\varepsilon}{|A(s)|}, & a \neq \arg\max_a Q(s,a). \end{cases} \tag{18}$$

For strategy $a = \arg\max_a Q(s,a)$ that adopts the maximization value function, the probability of its optimal action being selected is $1 - \varepsilon + (\varepsilon/|A(s)|)$, and the probability of each nonoptimal action being selected is $\varepsilon/|A(s)|$. Therefore, when the ε greedy strategy is adopted, each action may be selected, and different learning paths will be generated through multiple learning. In RL, a small value of ε is generally set first, and the agent has a probability of ε according to the above formula to randomly select actions to explore the experience. The agent has a $1 - \varepsilon$ probability to take action based on the learned Q value.

5.4. Algorithm Implementation. This section implements the semiactive control strategy of MR damper based on the RL Q-learning algorithm through the secondary development of Abaqus. The specific algorithm flowchart is shown in Figure 10. According to the Bouc–Wen model and the MR damper model selected in this paper, the MR damper only needs to control the current to adjust the damping to control the structure. In this section, applying current to the MR damper is taken as action A_i in RL, and the response of each incremental step in Abaqus is taken as state S_i. Because the MR damper selected in this article has a maximum operating current of 1.0 A, therefore in this section, there are 11 optional actions for each state S_i, namely, applying 11 different intensities of current, I = 0 A, I = 0.1 A, I = 0.2 A, I = 0.3 A, I = 0.4 A, I = 0.5 A, I = 0.6 A, I = 0.7 A, I = 0.8 A, I = 0.9 A, and I = 1.0 A.

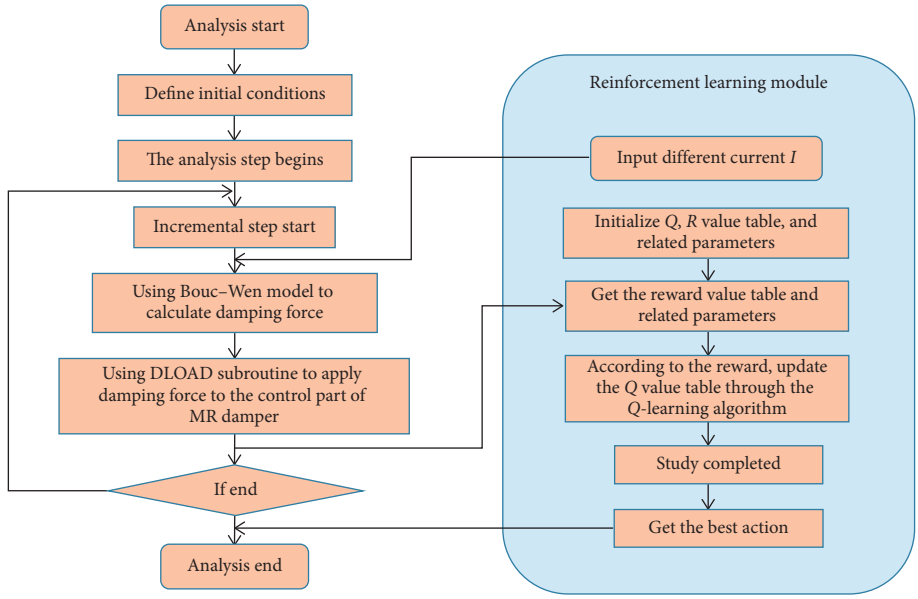

Figure 10: The flowchart of Q-learning reinforcement learning algorithm for MR damper based on Abaqus.

By inputting different currents (actions), the MR damper is used to control the vibration of the structure, and the response value of each incremental step is output. The reward value of each action is calculated through the reward evaluation function to form a reward value table—R value table. The reward value $r_{m\times n}$ calculated by the reward evaluation function is stored in the R value table, which is a $m \times n$ matrix

$$R = \text{State} \begin{matrix} \\ \begin{matrix} \text{Action} \end{matrix} \end{matrix} \begin{bmatrix} r_{11} & \cdots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{m1} & \cdots & r_{mn} \end{bmatrix}. \tag{19}$$

In the formula, $r_{m\times n}$ represents the reward value of each action, where m is the number of states and n is the number of actions. After getting the reward R value table, the RL Q-learning algorithm can learn according to the R value table. In reinforcement learning, the algorithm records the reward value obtained by each learning in the Q value table; that is, the Q value table is the learned experience value, and the Q value table is a $m \times n$ matrix, which is of the same order as the R value table.

$$R = \text{State} \begin{matrix} \\ \begin{matrix} \text{Action} \end{matrix} \end{matrix} \begin{bmatrix} q_{11} & \cdots & rq_{1n} \\ \vdots & \ddots & \vdots \\ q_{m1} & \cdots & q_{mn} \end{bmatrix}. \tag{20}$$

In the formula, q_{mn} represents the experience value learned for each action in this state, where m is the number of states and n is the number of actions. The Q value table is updated according to the Q-learning algorithm, and the formula is as follows:

$$Q(S_t, A_t) = Q(S_t, A_t) + \alpha\left[R_{t+1} + \gamma \max_a Q(S_{t+1}, a) - Q(S_t, A_t)\right], \tag{21}$$

where γ is attenuation factor (discounting factor), α is learning rate (learning rate), A is action set, and S is state set.

Finally, through the above steps, the final reward value of each action (current) in each state, that is, each incremental step, can be obtained. Therefore, the optimal action (current) of each step can be selected to form the optimal control strategy. In the next damping control of the MR damper, this strategy can be used to perform the optimal half of the MR damper.

6. Simulation Results and Analysis

6.1. Calculation Model and Boundary Conditions. This section takes a two-layer frame structure (including plate, beam, and column structure) as an example and uses the secondary developed Abaqus program to study the semi-active control strategy based on the RL Q-learning algorithm. The calculation model is shown in Figure 11. The structural parameters and material parameters are shown in Tables 1 and 2. The element type is a spatial second-order tetrahedral ten-node element (C3D10). A tangential dynamic load of 2000 N is applied to the center of the first floor of the structure at a frequency of 8.333 Hz, and the time history of the external force couple is given by $F(t) = f(t) \times F_D$, where $f(t)$ is shown in Figure 12 and the bottom of the structure is set as a fixed end constraint. In the study of MR damper vibration reduction, in order to apply the model to large-scale structures, the model parameters are generally enlarged so that it can be used in large-scale civil engineering [31]. Therefore, in order to obtain a good damping effect, this section enlarges the performance parameters of the RD-8041-1MRD MR damper by 5 times and then conducts vibration damping research on the structure.

6.2. Selection of Reward Evaluation Function. In order to construct the reward evaluation function in RL, it is necessary to determine the evaluation index of the state. This section uses three state variables: displacement, velocity, and

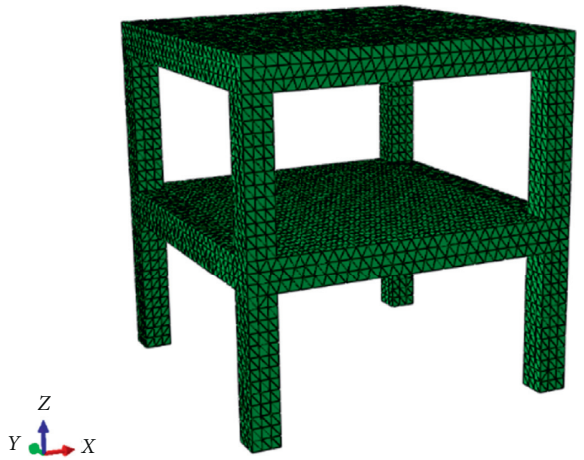

Figure 11: The mesh of two-story frame structure.

Table 1: The parameters of two-story frame.

Parameters	Floor (m)	Beam (m)	Column (m)	Span (m)	Floor height (m)
Parameter value	0.2	0.6 × 0.6	0.5 × 0.5	6.0	3.0

Table 2: Material parameters.

Material parameters	Concrete strength	Gravity density (kN/m^3)	Static elastic modulus (GPa)	Poisson's ratio	Rayleigh damping	
					α	β
Parameter value	C30	24.0	30.0	0.167	2.67	0.0009

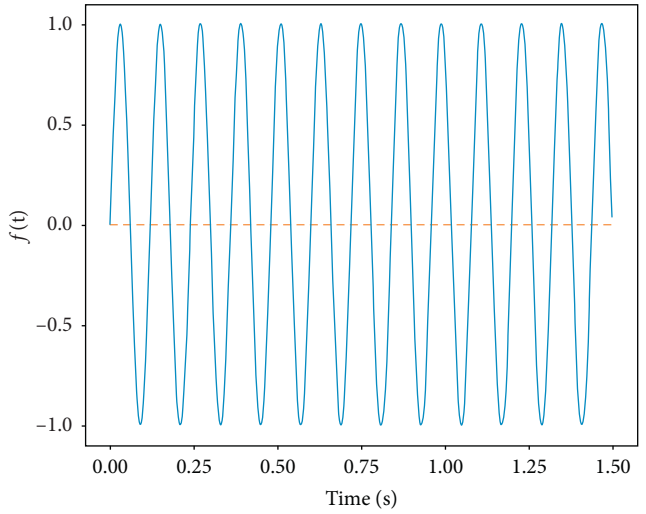

Figure 12: Time variation factor f_t of force couple.

acceleration to establish reward evaluation functions. Determine which index is used as the reward evaluation function for the best learning effect. The parameters in the RL are shown in Table 3, where the learning rate of this RL $\alpha = 0.8$; the attenuation coefficient $\gamma = 0.4$; the greedy strategy $\varepsilon = 0.1$; the number of learning times is 1000.

Three different reward evaluation functions have been learned 1000 times, and the corresponding optimal actions are obtained. The results are shown in Figure 13, Figure 13(a) is the average reward value of different reward evaluation functions; Figure 13(b) is the action value after the 1000th learning with displacement as the reward evaluation function; Figure 13(c) is the action value after the 1000th learning with speed as the reward evaluation function; Figure 13(d) is the action value after the 1000th learning with acceleration as the reward evaluation function.

It can be seen from Figure 13(a) that the reward value tends to converge after learning about 600 times for the 3

Table 3: The parameters of the reinforcement learning.

Parameter name	Learning rate α	Attenuation coefficient γ	Greedy strategy ε value	Number of learning times
Parameter value	0.8	0.4	0.1	1000

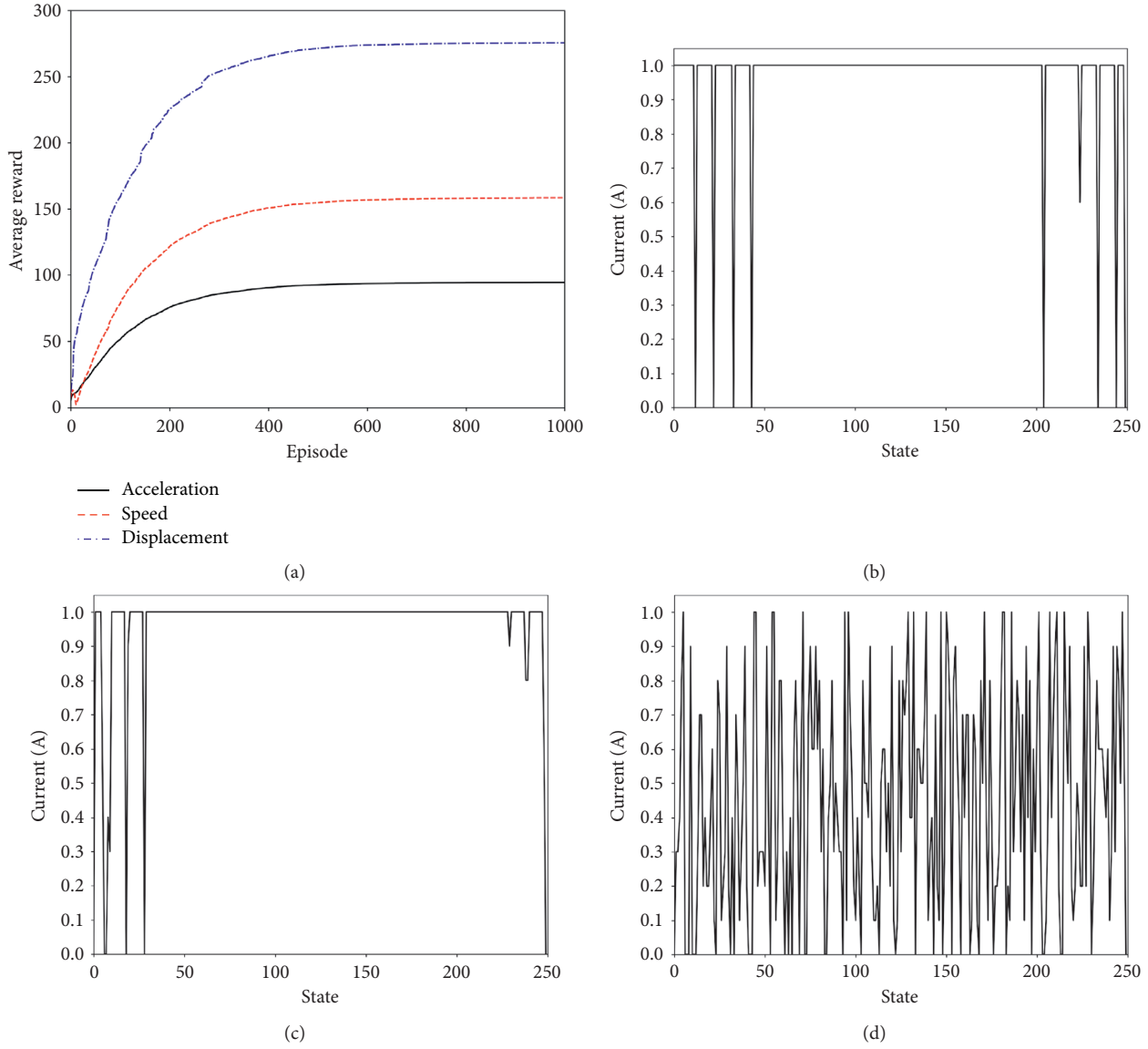

Figure 13: The result of reinforcement learning with different reward evaluation functions. (a) Average reward value of different reward evaluation functions. (b) Action value obtained by displacement reward evaluation function. (c) Action value obtained by speed reward evaluation function. (d) Action value obtained by acceleration reward evaluation function.

different reward evaluation functions. Figures 13(b)–13(d) are the learning results with displacement, speed, and acceleration as the reward evaluation function, that is, the corresponding action value in each state—the applied current value. Table 4 and Figures 14–16 show the vibration reduction effect of MR damper after learning with different reward evaluation functions. From the results in the graph, it can be seen that the vibration reduction effect after learning with the speed reward evaluation function is the best, and the learning effect with the acceleration reward evaluation function is the worst. Among them, the maximum displacement response using the speed reward evaluation function is reduced by 45.63%, the maximum speed response is reduced by 47.73%, and the maximum acceleration response is reduced by 48.17%. The learning effect of the displacement reward evaluation function is closer to that of the speed reward evaluation function.

6.3. Selection of Reinforcement Learning Parameters. The main parameters in the RL Q-learning include learning rate α, attenuation coefficient γ, and greedy strategy ϵ

Table 4: The result of MR damper vibration reduction under different reward evaluation functions.

Observation index	Reward evaluation function		
	Displacement	Speed	Acceleration
Maximum displacement response (mm)	0.044 (45.13%)	0.044 (45.63)	0.056 (29.50%)
Maximum speed response (mm/s)	2.308 (47.21%)	2.286 (47.73%)	2.876 (34.24%)
Maximum acceleration response(m/s^2)	0.123 (47.62%)	0.122 (48.17%)	0.179 (24.00%)
Maximum damping force (N)	7064	7063	9903

Note. The value in parentheses is the response reduction rate, that is (no control situation−control situation)/no control situation.

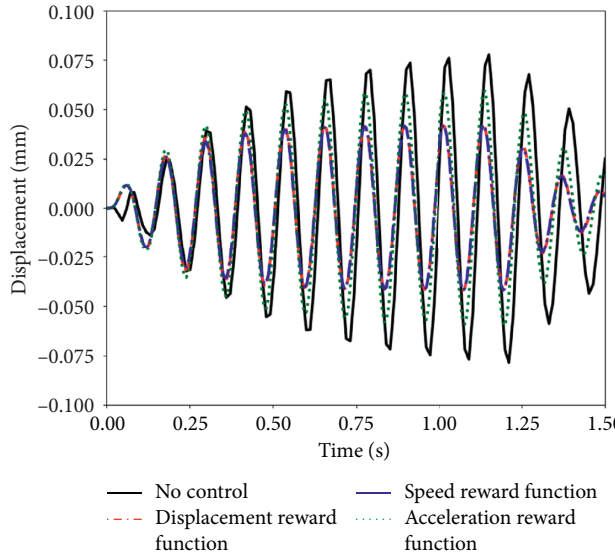

Figure 14: The time history of displacement response of the center point of the second floor for the different reward evaluation functions.

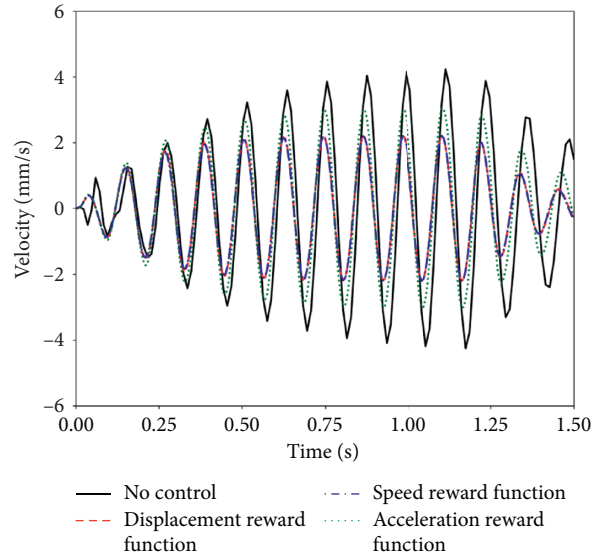

Figure 15: The time history of velocity response of the center point of the second floor for the different reward evaluation functions.

value. Different parameters have a greater impact on the accuracy and success rate of learning, so reasonable selection of RL parameters can obtain the best learning effect. Among them, the correct rate of learning represents the probability of the learned action being the action with the largest reward value; the success rate represents the completion ratio of the learning reward value table.

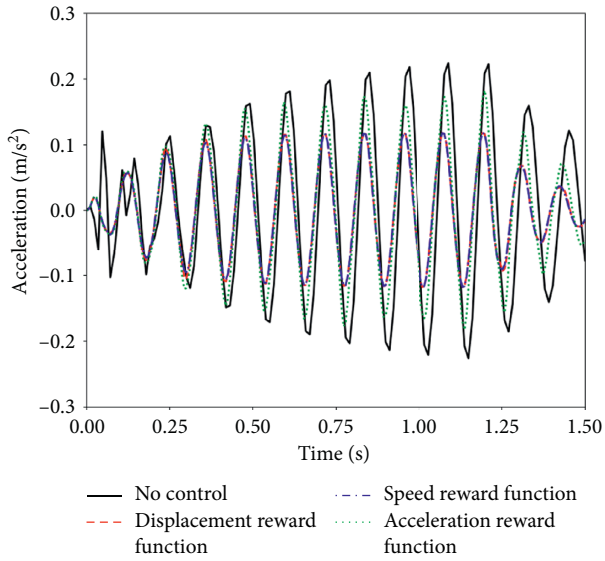

Figure 16: The time history of acceleration response of the center point of the second floor for the different reward evaluation functions.

6.3.1. Selection of Learning Rate α. The function of the learning rate is to learn the reward value brought by the current strategy during the update of the Q value table. In order to study the influence of the learning rate α on the learning effect, this section sets up 5 different working conditions, among which $\alpha = 0.2, 0.4, 0.6, 0.8$, and 1.0 are five learning rates for RL, the attenuation coefficient γ is 0.4, and the greedy strategy value is 0.1. The number of learning times is 1000.

Table 5 shows the effect of learning rate α on the learning effect. The learning rate of working condition 1 is 0.2, the correctness rate of the action at this time is 40.73%, the success rate is 80.83%, and the learning effect is not good at this time. As the learning rate α increases, the learning effect gradually improves. When the learning rate α is increased to 0.8 and 1.0, the correct rate of the action reaches 100.00%, and the success rate is 100.00%, and the calculation converges. Therefore, as the learning rate increases, the correct rate of RL also increases. However, in the RL Q-learning, when the learning rate $\alpha = 1.0$, the Q value table selects the current strategy and discards the values in the original Q value table. At this time, the Q value table will be continuously updated and excessive occupation computing resources. Figure 17 shows the average reward value for different learning rates α. It can be seen from the figure that as the learning rate increases, the convergence speed also increases.

6.3.2. Selection of Attenuation Coefficient γ. The attenuation coefficient γ represents the influence of the future strategy on the present under the current strategy, and the value range is $[0, 1]$. In order to study the influence of the attenuation coefficient γ on the learning effect, four different working conditions are set, among which $\gamma = 0.2$, 0.4, 0.6, and 0.8; four attenuation coefficients are used for RL. The number of learning is 1000 times. Table 6 shows the influence of the attenuation coefficient γ on the learning effect. The attenuation coefficient γ in the working condition 6 is 0.2. At this time, the correct rate of action is 99.19%, and the success rate is 99.89%. The attenuation coefficient γ in working condition 7 and working condition 8 is 0.4 and 0.6. At this time, the correct rate and success rate of the action have reached 100.00%. But when the attenuation coefficient γ increases to 0.8, the correct rate is 99.60%, and the success rate is 99.78%. Figure 18 shows the average rewards under different attenuation coefficients γ. It can be seen the attenuation coefficient γ has a greater impact on the average reward value but has a small impact on the calculated convergence speed.

6.3.3. Selection Greedy Strategy ε Value. The greedy strategy ε value is used to balance "experience" and "exploration" in the RL Q-learning. In order to study the influence of the greedy strategy ε value on the learning effect, this section sets 4 working ε conditions for RL. Table 7 shows the influence of the greedy strategy value on the learning effect. The greedy strategy ε value of working condition 10 is 0.1, the correct rate is 81.85%, the success rate is 96.02%, and the learning effect is relatively poor because at this time there is a 90% probability that the agent will act according to the Q value of the learned experience, that is, directly select the action with the largest Q value. It can be seen that, at this time, it is mainly learned through "experience," and the probability of exploring new actions is smaller. As the value of the greedy strategy ε increases, the probability of the agent exploring new actions increases, so the learning effect of working case 11 is improved, and the success rate reaches 100.00%, but the correct rate at this time does not reach 100.00%. The greedy strategy ε value of working condition 12 is 0.9, and the success rate and correct rate at this time are 100.00%. Figure 19 shows the average reward of different greedy strategy values ε. It can be seen from the figure that the value mainly affects the convergence speed of RL. When the value is increased, learning uses a higher probability to explore

Table 5: The influence of learning rate α on the learning effect.

Working condition	Learning rate α	Attenuation coefficient γ	Greedy strategy ε value	Correct rate (%)	Success rate (%)
1	0.2	0.4	0.1	40.73	80.83
2	0.4			81.85	96.20
3	0.6			98.39	99.24
4	0.8			100.00	100.00
5	1.0			100.00	100.00

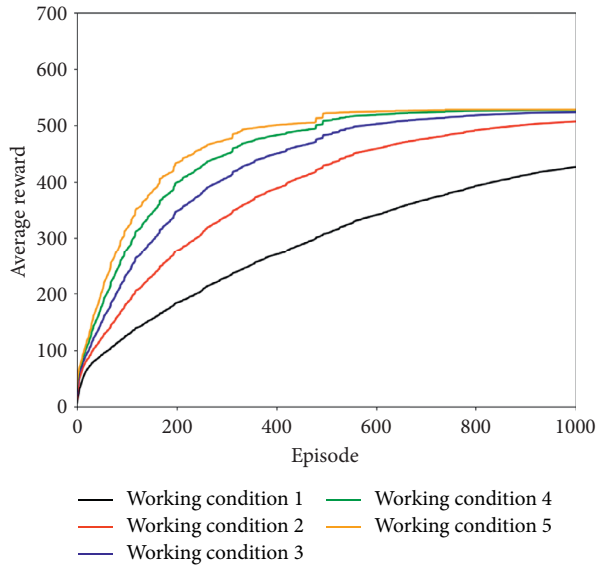

Figure 17: The average rewards at the different learning rates α.

Table 6: The influence of discounting factor γ on the learning effect.

Working condition	Learning rate α	Attenuation coefficient γ	Greedy strategy ε value	Correct rate (%)	Success rate (%)
6	0.8	0.2	0.1	99.19	99.89
7		0.4		100	100
8		0.6		100	100
9		0.8		99.60	99.78

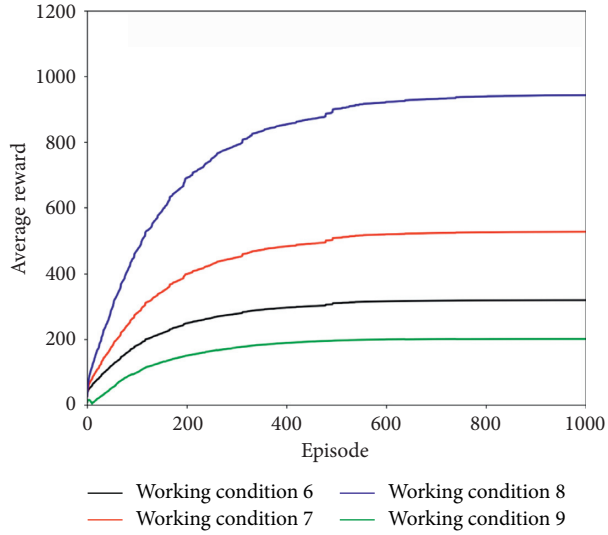

Figure 18: The average rewards at the different discounting factors γ.

Table 7: The influence of values of ε on the learning effect.

Working condition	Learning rate α	Attenuation coefficient γ	Greedy strategy value ε	Correct rate (%)	Success rate (%)
10	0.4	0.4	0.1	81.85	96.02
11			0.5	99.19	100.00
12			0.9	100.00	100.00
13			1.0	99.20	100.00

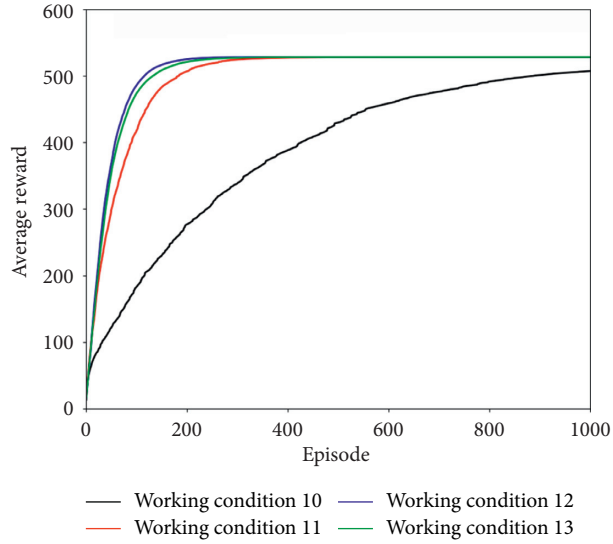

Figure 19: The average rewards at the different values of ε.

Table 8: The effect of vibration reduction for different methods.

Observation index	Vibration reduction method	
	Simple Bang-Bang	Reinforcement learning
Maximum displacement response (mm)	0.016 (80.00%)	0.013 (83.50%)
Maximum speed response (mm/s)	0.970 (77.83%)	0.707 (83.83%)
Maximum acceleration response (m/s^2)	0.061 (74.26%)	0.039 (83.62%)

Note. The value in parentheses is the response reduction rate, that is (no control situation−control situation)/no control situation.

new actions, thereby improving the convergence speed of the calculation. However, when the greedy strategy value ε for operating condition 13 is 1.0, its convergence speed is lower than operating condition 12. This is because if the agent only focuses on exploring new actions, that is, if the agent adopts all random actions, most of the learned actions will be of no value and affect the learning effect.

In summary, in RL Q-learning, a higher learning rate α can improve the accuracy and convergence speed of learning, but a too high learning rate will cause the constant update of the value table to occupy computing resources. Therefore, a larger value can be set at the beginning of learning. As the learning process deepens, the learning rate α can be reduced to reduce the occupancy rate of computing resources; the attenuation coefficient γ represents the impact of future strategies on the present under the current strategy. The attenuation coefficient γ has a greater impact on the average reward value but has a small impact on convergence; the ε value mainly affects the convergence speed of RL. When the ε value is increased, learning uses a higher probability to explore new actions, thereby increasing the calculation convergence speed. However, if the agent only focuses on exploring new actions, that is, the agent adopts all random actions, most of the learned actions will be of no value and affect the convergence speed of learning.

6.4. Vibration Reduction Control Effect of Semiactive Control Strategy Based on RL. This section used simple Bang-Bang and a semiactive control strategy based on the RL Q-learning algorithm to study the vibration reduction control of the two-layer frame structure. During vibration damping control, 4 MR dampers are installed on the upper part of each column on the first floor. It can be seen from Table 8 that, among the two semiactive control strategies, the RL strategy has the best effect, and the maximum displacement, velocity, and acceleration responses are reduced by 83.50%, 83.83%, and 83.62%, respectively. Compared with the simple Bang-Bang control, the maximum displacement, speed,

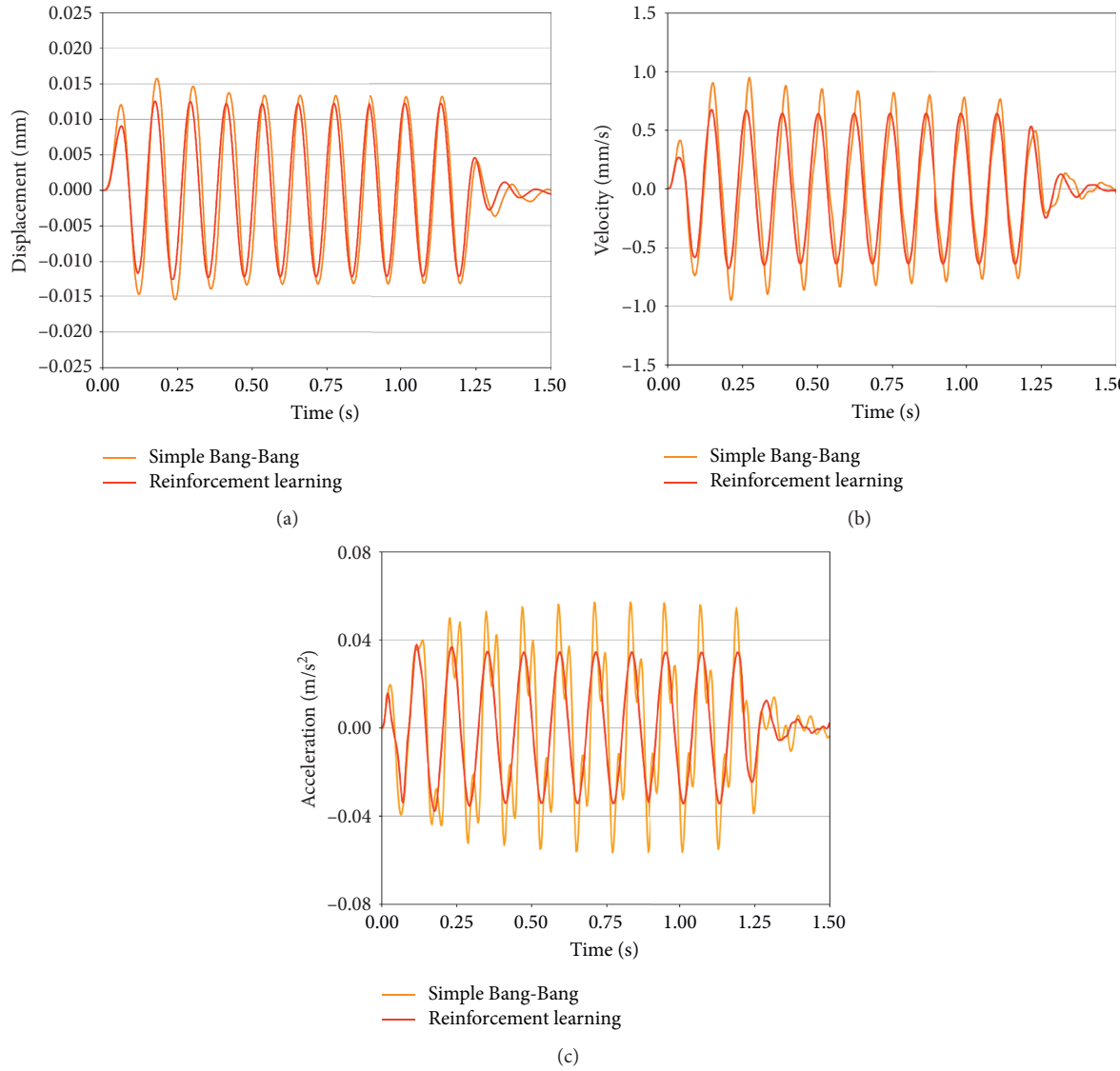

Figure 20: The time history of response in the center point of second floor for different control methods.

acceleration response, and vibration reduction effect are increased by 3.50%, 6.00%, and 9.36%, respectively.

Figure 20 shows the response time history curve of the central point of the second floor of the structure under two semiactive damping strategies. It can be seen from Figure 20 that the damping effect of RL strategy is better than that of simple Bang-Bang for displacement, velocity, and acceleration response. In addition, Figure 20(c) shows that the acceleration time history curve of RL strategy changes more smoothly, and there is no local acceleration mutation like simple Bang-Bang control.

7. Conclusions

This paper proposes a semiactive control strategy of MR damper based on the RL *Q*-learning algorithm. According to the structural response, the corresponding reward evaluation function is established, and the semiactive control strategy of the MR damper based on the RL *Q*-learning algorithm is realized through the secondary development of Abaqus. Taking the two-layer frame structure as an example, the vibration damping control is implemented through a semiactive control strategy based on RL *Q*-learning. At the same time, the simple Bang-Bang control is compared, and the following conclusions are drawn.

This article proposes a semiactive control strategy based on the RL *Q*-learning algorithm. The method continuously learns through the "exploration" method to obtain the optimal action value of each step of the MR damper—the applied current. In the vibration damping control, the optimal action value obtained at each step of the learning is input into the MR damper, so that it can provide the optimal damping force to control the structure vibration. The results show that the semiactive control strategy based on RL *Q*-learning is simple, easy to implement, and robust. By adopting the semiactive control strategy of RL to study the vibration reduction of the two-layer frame structure, the results show that RL is better than simple Bang-Bang

control. In addition, the acceleration time history curve of the RL strategy changes more smoothly.

For the RL algorithm, this paper establishes three reward evaluation functions: displacement reward evaluation function, speed reward evaluation function, and acceleration reward evaluation function. It can be seen from the results of RL and vibration damping control that the vibration damping control effect of the action learned by the velocity reward evaluation function is the best.

This paper discusses the impact of three main RL parameters, learning rate α, attenuation coefficient γ, and greedy strategy ε, on the learning effect. A higher learning rate α can improve the accuracy of learning and the speed of convergence, but a too high learning rate will keep updating the value table and occupy too much computing resources. Therefore, a larger value can be set at the beginning of learning. As the learning process deepens, the learning rate α can be reduced to reduce the occupancy rate of computing resources. The attenuation coefficient γ has a greater impact on the average reward value but has a small impact on convergence. The greedy strategy value ε mainly affects the convergence speed of RL. When the value ε is increased, learning uses a higher probability to explore new actions, thereby increasing the convergence speed of calculation. However, if the agent only focuses on exploring new actions, that is, all the agents adopt random actions, most of the learned actions will be of no value, which will affect the learning speed.

Data Availability

The codes used in this paper are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this study.

Acknowledgments

This study was funded by the National Natural Science Foundation of China (grant no. 51579089).

References

[1] M. Rahman, Z. C. Ong, S. Julai, M. M. Ferdaus, and R. Ahamed, "A review of advances in magnetorheological dampers: their design optimization and applications," *Journal of Zhejiang University-Science A*, vol. 18, no. 12, 2017.

[2] H. Ji, Y. Huang, S. Nie, F. Yin, and Z. Dai, "Research on semi-active vibration control of pipeline based on magneto-rheological damper," *Journal of Applied Sciences*, vol. 10, no. 7, 2020.

[3] A. H. Ghasemikaram, A. Mazidi, M. R. Fazel, and S. A. Fazelzadeh, "Flutter suppression of an aircraft wing with a flexibly mounted mass using magneto-rheological damper," *Proceedings of the Institution of Mechanical Engineers—Part G: Journal of Aerospace Engineering*, vol. 234, no. 3, pp. 827–839, 2020.

[4] Y. Q. Ni, Y. F. Duan, Z. Q. Chen, and J. Ko, *Damping Identification Of Mr-Damped Bridge Cables From In-Situ Monitoring Under Wind-Rain-Excited Conditions*, The Hong Kong University, Hong Kong, China, 2002.

[5] M. Abdeddaim, A. Ounis, N. Djedoui, and M. K. Shrimali, "Pounding hazard mitigation between adjacent planar buildings using coupling strategy," *Journal of Civil Structural Health Monitoring*, vol. 6, no. 3, pp. 603–617, 2016.

[6] R. Stanway, J. L. Sproston, and N. G. Stevens, "Non-linear modelling of an electro-rheological vibration damper," *Journal of Electrostatics*, vol. 20, no. 2, pp. 167–184, 1987.

[7] Y.-K. Wen, "Method for random vibration of hysteretic systems," *Journal of the Engineering Mechanics Division*, vol. 102, no. 2, pp. 249–263, 1976.

[8] B. F. Spencer, S. J. Dyke, M. K. Sain, and J. D. Carlson, "Phenomenological model for magnetorheological dampers," *Journal of Engineering Mechanics*, vol. 123, no. 3, pp. 230–238, 1997.

[9] M. Ismail, F. Ikhouane, and J. Rodellar, "The hysteresis bouc-wen model, a survey the hysteresis bouc-wen model," *Archives of Computational Methods in Engineering*, vol. 16, no. 2, pp. 161–188, 2009.

[10] H. Jae Lee, G. Yang, H. J. Jung, B. F. Spencer, and I. W. Lee, "Semi-active neurocontrol of a base-isolated benchmark structure," *Structural Control and Health Monitoring*, vol. 13, no. 2-3, pp. 682–692, 2006.

[11] S.-Y. Ok, D.-S. Kim, K.-S. Park, and H.-M. Koh, "Semi-active fuzzy control of cable-stayed bridges using magneto-rheological dampers," *Engineering Structures*, vol. 29, no. 5, pp. 776–788, 2007.

[12] V. Bhaiya, S. D. Bharti, M. K. Shirmali, and T. K. Datta, "Genetic algorithm based optimum semi-active control of building frames using limited number of magneto-rheological dampers and sensors," *Journal of Dynamic Systems, Measurements, Control*, vol. 140, no. 10, pp. 101013–101021, 2018.

[13] S. J. Dyke, B. F. Spencer, M. K. Sain, and J. D. Carlson, "Modeling and control of magnetorheological dampers for seismic response reduction," *Smart Materials and Structures*, vol. 5, no. 5, pp. 565–575, 1996.

[14] U. N. Mughal, "State of the art review of semi active control for magnetorheological dampers," *AIP Conference Proceedings*, vol. 1798, no. 1, Article ID 020101, 2017.

[15] K. I. Gkatzogias and A. J. Kappos, "Semi-active control systems in bridge engineering: a review of the current state of practice," *Structural Engineering International*, vol. 26, no. 4, 2016.

[16] A. Bathaei, S. M. Zahrai, and M. Ramezani, "Semi-active seismic control of an 11-dof building model with tmd+mr damper using type-1 and -2 Fuzzy algorithms," *Journal of vibration and control*, vol. 24, no. 13, 2018.

[17] N. K. Hazaveh, J. G. Chase, G. W. Rodgers, and S. Pampanin, "Smart semi-active mr damper to control the structural response," *Bulletin of the New Zealand Society for Earthquake Engineering*, vol. 48, no. 4, 2015.

[18] H.-S. Kim, "Development of seismic response simulation model for building structures with semi-active control devices using recurrent neural network," *Applied Sciences*, vol. 10, no. 11, p. 3915, 2020.

[19] C. Watkins and P. Dayan, "Technical note: *Q*-learning," *Machine Learning*, vol. 8, no. 3-4, pp. 279–292, 1992.

[20] C. Brodley and T. N. Health, "Experimental analysis of simulated reinforcement learning control for active and passive building thermal storage inventory," in *Proceedings of the 21st International Conference on Machine Learning*, vol. 48, no. 4, pp. 413–424, New York, NY, USA, 1958.

[21] C. Brodley, "Learning control systems-review and outlook," in *Proceedings of the 21st International Conference on Machine Learning*, New York, NY, USA, July 2004.
[22] M. L. Littman and L. J. N. Michael, "Reinforcement learning improves behaviour from evaluative feedback," *Nature*, vol. 521, no. 7553, pp. 445–451, 2015.
[23] M. Hara, M. Inoue, H. Motoyama, J. Huang, and T. Yabuta, "Study on motion forms of mobile robots generated by *Q*-learning process based on reward databases," in *Proceedings of the IEEE International Conference on Systems*, Taipei, Taiwan, October 2007.
[24] C. Szepesvari, *Algorithms for Reinforcement Learning*, Morgan and Claypool Publishers, San Rafael, CA, USA, 2010.
[25] S. B. Thrun, *Efficient Exploration in Reinforcement Learning*, Springer, Berlin, Germany, 1992.
[26] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, MIT Press, Cambridge, MA, USA, 1998.
[27] M. Wiering and M. V. Otterlo, *Reinforcement Learning: State of the Art*, Springer, Berlin, Germany, 2012.
[28] C. Andrieu, N. De Freitas, A. Doucet, and M. Jordan, "An introduction to MCMC for machine learning," *Machine Learning*, vol. 50, no. 1-2, pp. 5–43, 2003.
[29] H. Van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-learning," in *Proceedings of the 13th AAAI Conference on Artificial Intelligence*, Austin, TX, USA, February 2015.
[30] N. R. Fisco and H. Adeli, "Smart structures: Part I-Active and semi-active control," *Scientia Iranica*, vol. 18, no. 3, pp. 275–284, 2011.
[31] H. Naderpour, R. Vahdani, and M. Mirrashid, "Soft computing research in structural control by mass damper (a review paper)," in *Proceedings of the 4th International Conference on Structural Engineering*, Tehran, Iran, February 2018.