



Article

Research on Energy Management Strategy of a Hybrid Commercial Vehicle Based on Deep Reinforcement Learning

Jianguo Xi, Jingwei Ma, Tianyou Wang and Jianping Gao *

Vehicle and Traffic Engineering College, Henan University of Science and Technology, Luoyang 471003, China; xijianguo@sohu.com (J.X.); mjw321456@163.com (J.M.); 17734812752@163.com (T.W.)

* Correspondence: gaojp@haust.edu.cn

Abstract: Given the influence of the randomness of driving conditions on the energy management strategy of vehicles, deep reinforcement learning considering driving conditions prediction was proposed. A working condition prediction model based on the BP neural network was established, and the correction coefficient of vehicle demand torque was determined according to the working condition prediction results. An energy management strategy and deep reinforcement learning were integrated to build an energy management strategy with deep reinforcement learning based on driving condition prediction. Simulation experiments were conducted according to the actual collected working condition data. The experimental results show that the energy management strategy, i.e., deep reinforcement learning considering working condition prediction, has faster convergence speed and more vital self-learning ability, and the equivalent fuel consumption per 100 km under different driving conditions is 6.411 L/100 km, 6.327 L/100 km, and 6.388 L/100 km, respectively. Compared with the unimproved strategy, the fuel economy can be improved by 3.18%, 3.08%, and 2.83%. The research shows that the energy management strategy, the deep reinforcement learning based on driving condition prediction, is effective and adaptive.

Keywords: index hybrid commercial vehicle; energy management strategy; deep reinforcement learning; working condition prediction



Citation: Xi, J.; Ma, J.; Wang, T.; Gao, J. Research on Energy Management Strategy of a Hybrid Commercial Vehicle Based on Deep Reinforcement Learning. *World Electr. Veh. J.* **2023**, *14*, 294. <https://doi.org/10.3390/wevj14100294>

Academic Editors: Danial Karimi and Amin Hajizadeh

Received: 28 August 2023

Revised: 30 September 2023

Accepted: 12 October 2023

Published: 15 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The power system of a hybrid electric vehicle is composed of multiple power sources, and the power distribution of different power sources is realized through an energy management strategy to improve the vehicle's fuel economy and driving range. As one of the crucial technologies of hybrid electric vehicles, the vehicle control strategy primarily solves plug-in hybrid electric vehicles' energy management and torque distribution. It can be divided into a control strategy based on base rules, a control strategy based on optimization class, and a control strategy based on learning.

The rule-based control strategy can also be called the logic threshold-based control strategy, and its core idea is to ensure that the engine works in the high-efficiency zone. When the engine load is small, the machine stops, and the motor is driven separately; when the engine load is moderate, the engine works in the high-efficiency area, and the engine starts to charge or go; when the engine load is large, the motor can assist so that the machine only works in the high-efficiency area. This control strategy allows the engine to provide steady-state power and the motor to provide transient ability to improve the vehicle's fuel economy. This type of control strategy is simple, reliable, and practical. Ping Li et al. [1] used particle swarm optimization (PSO) to optimize the threshold parameters of the rule-based energy management strategy. To improve the adaptability of the control strategy, multiple historical driving cycles are used to optimize the parameters, resulting in a rule-based energy management control strategy that adapts to unknown driving cycles. Abdoulaye Pam [2] et al. used DP to determine the ideal energy efficiency of the studied

vehicle in a given driving cycle. A rule-based EMS algorithm can be derived by analyzing the DP-EMS results. Charbel J Mansour [3] proposed a strategy optimization method, a rule-based energy management method which takes dynamic programming as the global optimization program to realize the real-time implementation of the energy management strategy of the Prius plug-in hybrid electric vehicle. The optimization process considers the ideal travel the driver selects on the vehicle-mounted global positioning system and is associated with the traffic management system. The control strategy based on rule class is widely used in engineering because of the small amount of calculation required and its fast calculation speed. However, it relies too much on formulation experience and needs better portability; it is challenging to achieve optimal control of vehicle power in application.

The operation of the rule-based control strategy is independent of the working conditions. Although it can improve the vehicle's fuel economy to a certain extent, the results could be more optimal. According to the optimization objectives, the optimization control strategy can be divided into instantaneous and global optimization control strategies. Mansour C [4] et al. proposed a simple adaptive rule strategy based on short-term driving pattern recognition and dynamic programming global optimization program. Mingming Gao et al. [5] proposed an extended-range electric bus energy management strategy based on a convex optimization algorithm, which can be better applied to the REEB energy management system to meet the requirements of the power system. The Radau pseudo spectral knot method (RPM) is proposed by Kegang Zhao [6] to solve the energy management strategy of series-parallel plug-in hybrid electric vehicles based on global optimization to improve computational efficiency. Jian Wu et al. [7] used PSO combined with various driving conditions to optimize the logic threshold parameters of a rule-based energy management strategy with the vehicle dynamic performance index as the constraint condition and the equivalent fuel consumption rate as the optimization objective.

With big data and computer technology development, machine learning is widely used in vehicle energy management strategies. The learning-based control strategy does not depend on the 'expert experience' and the digital model calculation of the controlled object. However, it uses advanced data mining methods and historical/real-time empirical data to obtain prediction results or control strategies. With the help of intelligent algorithms, the state space continuity and state action space continuity of energy management problems are realized, and the discretization problems in the optimization of the DP algorithm are avoided [8–12]. Tawfiq M. Aljohani et al. [13] proposed a real-time, metadata-driven electric vehicle path optimization method to reduce road energy demand. The strategy uses the state-behavior-reward-state-behavior (SARSA) algorithm to learn the maximum travel strategy of electric vehicles as agents. Weihan Li et al. [14] proposed a multi-objective energy management strategy based on cloud-based hybrid architecture. This strategy has a deep deterministic policy gradient, which can improve the system's electrical and thermal safety and minimize the system's energy loss and aging cost. Weihan Li et al. [15] designed a new reward to explore the optimal working range of high-power battery packs without imposing strict charging state constraints. In the training of deep q-learning models, different load curves are randomly combined to avoid over-fitting problems.

Yue S et al. [16] solved the vehicle energy management problem of compound power supply using the sequential difference method. Li Wei et al. [17], from the University of Chinese Academy of Sciences, added the battery life factor into the reward function of the deep reinforcement learning algorithm (DRL) to extend the battery life and verify their strategy's adaptability to working conditions in simulation verification. Tang Xiaolin et al. [18] from Chongqing University used the deep value network algorithm to complete the upper-level tracking control and the lower-level energy management, thus improving the fuel economy of the two vehicles. Zhao Chunling et al. [19], from Chongqing Jiaotong University, applied the DRL algorithm to the energy allocation problem of PHEVs, which not only reduces pollutant emission of diesel engines but also greatly improves the fuel economy of the whole vehicle. Zhang Song et al. [20] took hybrid electric buses as the research object, applied DDQN and TD3 algorithms to vehicle energy management, and

adopted priority empirical playback to optimize the strategy, proving the effectiveness of the strategy.

In actual road driving, the energy management strategy under different working conditions is easily affected by random factors, and it is difficult to achieve a real-time optimal energy management strategy. To solve this problem, this paper first predicts the driving conditions, determines the correction factor of the demand torque distribution based on the prediction results, and corrects the actual demand torque of the vehicle. Finally, the energy management strategy based on TD3 is designed to complete the design and development of an energy management strategy based on condition prediction.

2. Vehicle Power System Construction

As shown in Figure 1, the research object of this paper is the P2 configuration parallel hybrid commercial vehicle produced by a company. The main difference between the P2 configuration and other configurations is that a clutch controls the front and rear sides of the motor, so both the motor and the engine can drive the car independently.



Figure 1. The research object of this study.

Figure 2 shows the structure of the vehicle power system, and its main components are shown successively as follows: diesel engine, motor, power battery pack, clutch, five-speed transmission vehicle controller VCU, etc. The main parameters are shown in Table 1.

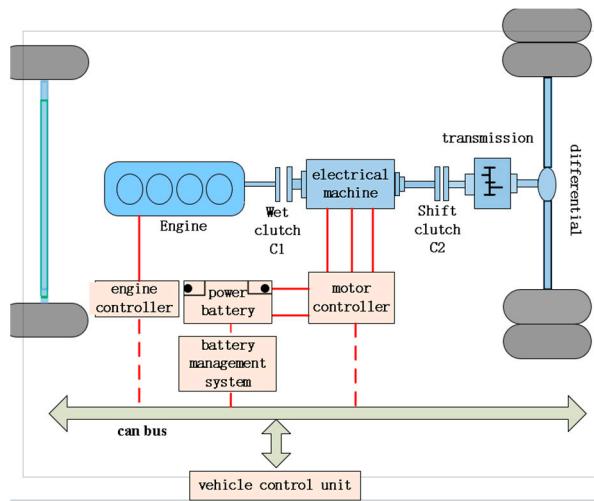


Figure 2. Vehicle structure diagram.

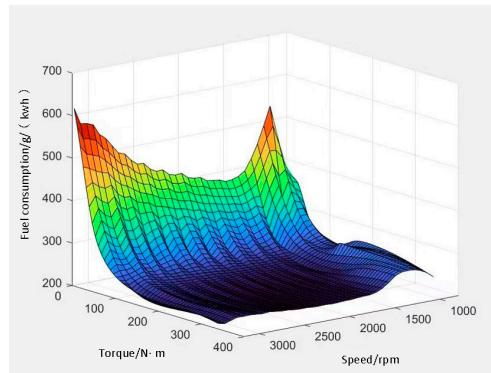
Table 1. Basic parameters of the vehicle and key components.

Category	Basic Parameter	Numerical Value
Vehicle parameters	Quality of preparation/(kg)	2670
	Rolling damping coefficient	0.0132
	Wind resistance coefficient	0.55
	Wheelbase (mm)	3360
	Rolling radius (mm)	369
	Rated power/(kW)	120
Engine	Calibrated speed/(r/min)	4200
	Maximum torque/(N·m)	320
	Rated power/(kW)	25
Motor	Peak power/(kW)	50
	Maximum torque/(N·m)	120
Power cell	Rated capacity/(Ah)	15
	Rated voltage/(V)	330
Main reducer	Transmission ratio	4.33

The experimental modeling method is adopted to model the engine. This paper focuses on the energy management of hybrid commercial vehicles, so the engine model is simplified without considering the instantaneous corresponding characteristics of the system.

The corresponding data, such as torque and speed, were obtained through bench experiments, and the fuel consumption experimental model was obtained using the interpolation Formula (1). The fuel consumption figure is shown in Figure 3.

$$\begin{cases} g_e = f_e(\omega_e, T_e) \\ m_f = \int_0^t f_e(\omega_e, T_e) dt \end{cases} \quad (1)$$

**Figure 3.** Engine fuel consumption MAP.

This paper still adopts the method of experimental modeling, does not consider the internal operation mechanism, and only believes the input and output relationship in building the motor model. By testing the driving motor at different speeds and torque points on the experimental bench, parameters such as speed, torque, and current at the shaft end of the driving motor were recorded, and the motor efficiency (ratio of motor output power to input power) MAP was established, as shown in Figure 4.

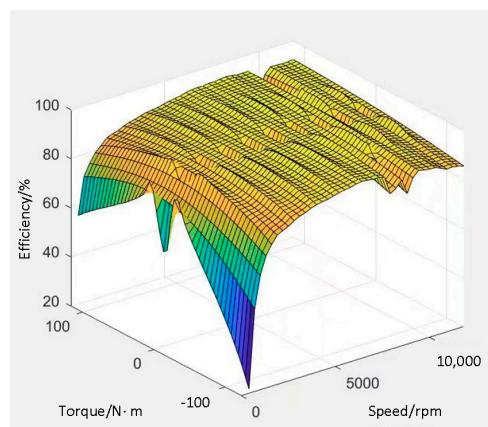


Figure 4. Motor efficiency map.

3. Introduction to Deep Reinforcement Learning

Deep learning originated from the research of artificial neural networks (ANNs). The mathematical model of ANNs is made of layers of neurons. ANNs are distributed parallel information processing algorithms that are used to simulate the behavior of animal neurons. Depending on the system's complexity, ANNs realize the purpose of processing information by adjusting the interconnection between large internal nodes. The so-called deep learning is the neural network composed of multilayer neurons to approximate the function of machine learning. The structure of deep understanding is a multilayer perceptron with multiple hidden layers. By combining low-level features to form more abstract high-level representations of attribute categories or segments, deep learning can discover the distributed characteristics of data [21].

The goal of reinforcement learning is to find the optimal strategy through trial-and-error learning between the agent and the environment to maximize the expectation of cumulative returns.

A reinforcement learning problem involves a decision-maker, the agent, operating in an environment modeled by states $\in S$. The agent can take specific actions at $\in A$ as a function of the current state. After choosing an action at time t , the agent receives a scalar reward $\in R$ and finds itself in a new state that depends on the current state. The chosen action reinforcement learning problem consists of a decision-maker, the agent, operating in an environment modeled by states $\in S$. The agent can take specific actions at $\in A$ as a function of the current state. After choosing an action at time t , the agent receives a scalar reward $\in R$ and finds itself in a new state that depends on the current and chosen actions [22].

At each time step, the agent follows a strategy, called the policy π_t , which is a mapping from states to the probability of selecting each possible action: $\pi(s, a)$ denotes the probability that $a = a_t$ if $s = s_t$.

The objective of reinforcement learning is to use the interactions of the agent with its environment to derive (or approximate) an optimal policy to maximize the total amount of reward received by the agent over the long run [23].

DRL combines the two disciplines of deep learning and reinforcement learning and uses the perceptual advantages of deep learning and the decision-making advantages of reinforcement learning to solve complex control problems belonging to MDP.

DRL combines deep learning and reinforcement learning to form a deep Q-learning network. Deep learning provides learning mechanisms, and reinforcement learning provides learning objectives for deep learning, making deep reinforcement learning capable of solving complex control problems [24].

4. Twin Delayed Deep Deterministic Policy Gradient Algorithm

4.1. Twin Delayed Deep Deterministic Policy Gradient Algorithm

DRL is divided into three main categories: value function-based, policy gradient-based, and search and supervision-based. The algorithm based on the value function uses the value table or value function to estimate the optimal value function reasonably to select the action with the largest value; this method is often applied to discontinuous and discrete environments, and for the large set and continuous action scene, this method is prone to problems such as dimensional catastrophe, and the results of the training are poor. The representative algorithms are Q-learning, DQN. Policy gradient-based algorithms are trained and learn to maximize the reward value of the objective function of the resulting policy to obtain the optimal policy; this algorithm has a better optimization effect compared with the algorithm based on the value function, but it is prone to problems such as local extremes, and the representative algorithms are DDPG. The algorithms based on searching and supervising are the algorithms that add artificial supervision when searching for a strategy to accelerate the learning process and achieve better results. In this paper, the improved TD3 algorithm is applied to energy management strategy development, which is improved based on DDPG and combines the advantages of DDQN and DDPG algorithms [23].

For the DRL algorithm, the twin delayed deep deterministic policy gradient algorithm, the TD3 algorithm, is an improved off-policy deep reinforcement learning algorithm for solving continuous control problems. In essence, the TD3 algorithm integrates the idea of the double Q-learning algorithm into the DDPG algorithm, combines both advantages, and uses delay strategy update and smooth regularization of the target strategy. In the face of complex continuous action space, it can implement efficient output action and effectively solve the overestimation of the Q value.

The TD3 algorithm adopts two critic networks to evaluate the output action-value function and then selects the minimum values of both to update the target Q value, as shown in Equation (2):

$$y = r_{t+1} + \gamma \cdot Q(s_{t+1}, a_{t+1}). \quad (2)$$

For network update mode, the TD3 algorithm also adopts soft update mode to update target network parameters, as shown in Formula (3):

$$\begin{aligned} \omega^* &\leftarrow \tau\omega + (1 - \tau)\omega^* \\ \theta_1^* &\leftarrow \tau\theta_1 + (1 - \tau)\theta_1^* \\ \theta_2^* &\leftarrow \tau\theta_2 + (1 - \tau)\theta_2^*. \end{aligned} \quad (3)$$

When training the algorithm, by increasing the ways to improve the algorithm of random noise and robustness, $\sim \text{clip}(N(0, \sigma), -c, c)$, $c > 0$, and

$$\tilde{a}_{t+1} = \mu^*(s_{t+1}) = \mu(s_{t+1} | \omega_{t+1}^*) + \varepsilon. \quad (4)$$

The loss function of the TD3 algorithm is defined as the error square of the above, as shown in Equation (5):

$$L(\theta_i) = E[(y_t - Q(s_t, a_t | \theta_i))^2]. \quad (5)$$

In the algorithm design, to avoid the correlation between samples, the experience pool playback method is adopted to store the experience data in the experience pool. When selecting samples for network training, the sample is randomly selected to break the correlation between samples and ensure the efficiency of network updating.

The key technology in DRL is the deep neural network to fit the Q value function, whose structure consists of an input, hidden, and output layer. The input layer is composed of states and actions. The selection of hidden layers and neurons in this layer is obtained via trial and error. After many tests, it is concluded that the number of hidden layers is

3, and the number of neurons in each layer is 30, 120, and 120. The relay function is used to activate between hidden layers. As shown in (6), the activation from the hidden layer to the output layer adopts the Tanh function, as shown in (7), and the output layer is the value function.

$$h(t) = \begin{cases} t, & t > 0 \\ 0, & t < 0 \end{cases} \quad (6)$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (7)$$

4.2. Key Parameter Selection

In the construction of an energy management strategy for hybrid commercial vehicles based on a deep reinforcement learning algorithm, the vehicle controller is regarded as an agent, the power system and driving conditions as the environment, and the ultimate purpose of the controller is to find the optimal control strategy.

In hybrid commercial vehicles based on deep reinforcement learning algorithms, the vehicle controller is regarded as an agent, the power system and driving conditions as the environment, and the ultimate purpose of the controller is to find the optimal control strategy.

The key parameters such as system state, action space, and reward signal are set as follows.

State variables: For energy management of PHEVs, the system state reflects the vehicle's characteristics while on the road. In this paper, the normalized acceleration (a), state of charge (SOC), demand torque (Tereq), and vehicle speed (V) are taken as the state variables of the algorithm. The formula can express its state space:

$$S(t) = [a(t), SOC(t), Te_{ice}(t), V(t)]. \quad (8)$$

Action variable: The engine output torque Te_{ice} is taken as the action variable of the algorithm:

$$U(t) = Te_{ice}(t). \quad (9)$$

Reward function: The reward function of the strategy affects the algorithm's convergence. In this paper, the SOC value will be taken as the constraint condition, and vehicle pollutant emission, fuel consumption, and power consumption will be considered as the feedback reward function of the algorithm. See the following formula for details:

$$R(t) = \alpha R_1(t) + \beta((SOC(t) - SOC_{req}(t))^2) \quad (10)$$

$$R_1(t) = \begin{cases} 1/\left(C_{fuel} + CO + HC + NO_X\right) & C_{fuel} + CO + HC + NO_X \neq 0 \cap 0.3 \leq SOC \leq 0.8 \\ 1/\left(C_{fuel} + CO + HC + NO_X + \lambda\right) & C_{fuel} + CO + HC + NO_X \neq 0 \cap SOC < 0.3 \text{ or } SOC > 0.8 \\ 2/\left(min_{C_{fuel}} + min_{CO+HC+NO_X}\right) & C_{fuel} + CO + HC + NO_X = 0 \cap 0.3 \leq SOC \\ -1/\lambda & C_{fuel} + CO + HC + NO_X = 0 \cap SOC < 0.3, \end{cases} \quad (11)$$

where $R(t)$ is the state x at time t ; under the action, x is transferred to the next state to obtain the reward value. R_1 represents the reward reporting function on the instantaneous fuel consumption of the engine and pollutant emission; it represents the instantaneous fuel consumption of the engine; CO, HC, and NO_X represent the emission of automobile pollutants. Because of the difference in dimensionality between them, the normalization method is used to deal with them before summation. λ is the penalty term, which is equal to the sum of the maximum emission and the maximum instantaneous fuel consumption of the engine; β is the penalty factor of SOC change; $SOC_{req}(t)$ is the reference SOC at a certain time; and α is the fuel consumption coefficient. When $SOC > SOC_{req}$, the fuel consumption coefficient is small, with PHEV provided mainly through the motor power;

when $SOC > SOC_{req}$, the penalty factor is set to a larger value to increase the torque distributed by the engine. According to the defined reward function, the reward obtained decreases with the increase in emissions and fuel consumption.

Based on the above introduction of some principles of deep reinforcement learning architecture and the setting of some parameters, the optimal-state action-value function is defined as follows:

$$Q^*(s, a) = \max_{\pi} E[R_t + \gamma Q^*(s_{t+1}, a_{t+1}) | s_t = s, a_t = a]. \quad (12)$$

As the driving condition prediction results affect the vehicle demand torque, an energy management strategy based on driving condition prediction is proposed by combining the driving condition prediction algorithm with the energy management strategy. According to the acceleration probability distribution in Figure 5, the correction factor is determined to correct the required torque of the vehicle. The figure shows that the two curves are more consistent, and the distribution of vehicle acceleration is mostly between -1.5 m/s^2 and 1.5 m/s^2 . Then, we calculate the correlation coefficient, average error, and standard deviation of its predicted demand torque, as well as the actual demand torque to set the correction coefficient: when the acceleration $a < -1.5 \text{ m/s}^2$, the correction coefficient is 0.8; when $-1.5 \text{ m/s}^2 < a < 1.5 \text{ m/s}^2$, the correction coefficient is 1; and when $a > 1.5 \text{ m/s}^2$, the correction coefficient is 1.2.

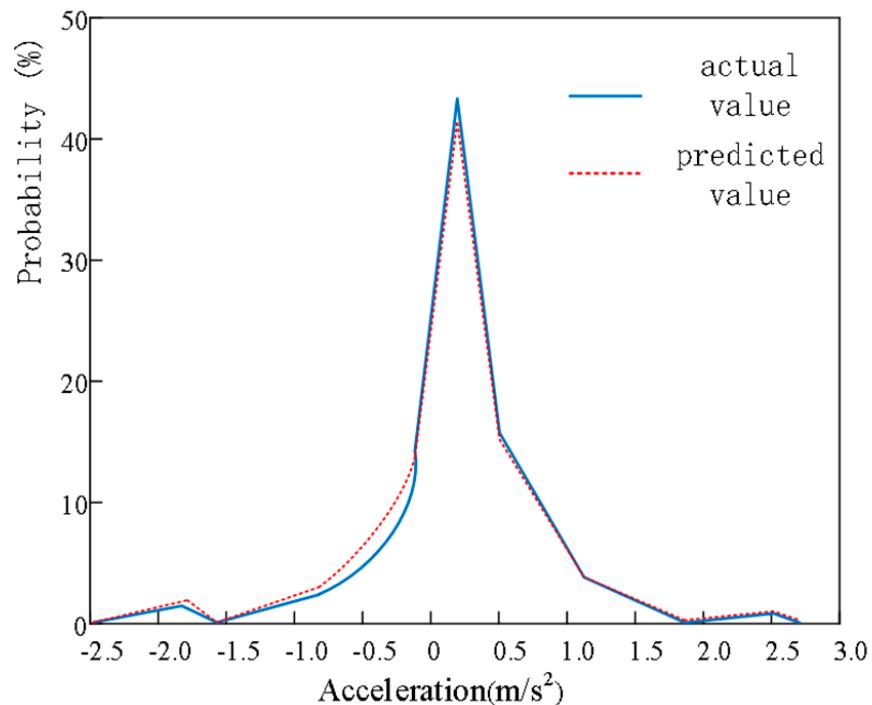


Figure 5. Probability diagram of acceleration distribution.

Figure 6 shows the framework of energy management of deep reinforcement learning based on condition prediction. When the vehicle is running, the BP neural network algorithm is first used to predict the driving conditions, and the vehicle demand torque is corrected according to the predicted value. The corrected vehicle demand torque, speed, battery SOC, and acceleration are state inputs. After the training of the target network and strategy network, the engine output torque with action value is output. We update the status value according to the output action and store the status, action, and reward value in the experience pool.

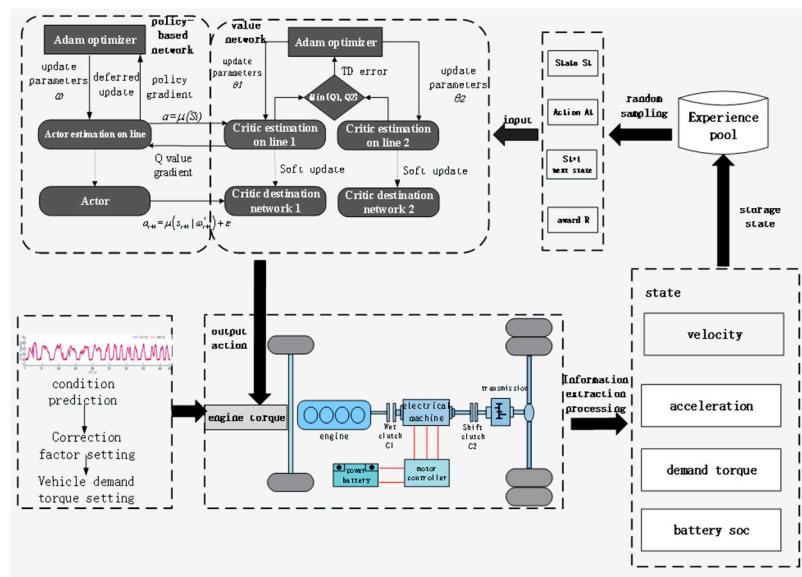


Figure 6. Framework diagram of deep reinforcement learning energy management strategy considering working condition prediction.

5. Simulation Verification

5.1. Subsection Validity Verification

To verify the effectiveness and adaptability of the proposed strategy, this section applies the constructed vehicle model and simulation environment to train the strategy. For the design of the simulation model, the initial SOC value of the strategy is 0.8, and the final SOC value is 0.3 (Table 2).

Table 2. Algorithm parameters.

Parameter Name	Value (m)
Minimum sample set n	64
Discount factor	0.9
Renewal coefficient	0.001
Sample number of experience pool	118,000
The actor estimates the network learning rate	0.001
Delayed updated	3
Estimate the network learning rate of Critic	0.001

To verify the effectiveness of the proposed strategy, a driving condition is selected as the simulation condition of the strategy, and the energy management strategy based on rule control is selected as the evaluation benchmark strategy based on the DRL strategy. The strategy is determined according to the optimal interval of engine operation and the upper and lower limits of SOC. The three strategies are simulated under the same working conditions. The deep reinforcement learning algorithm takes the TD3 algorithm as an example. The simulation results of the three strategies are shown in Figure 7.

As shown in Figure 7a, the TD3 energy management strategy considering condition prediction is continuous in the state space and can realize continuous control of throttle opening. In driving condition 1, the strategy based on DRL can make the vehicle run smoothly. When the vehicle starts to run, the engine and the motor work together. The motor's output torque is greater than the engine's output torque. With the increase in the speed, the demand for torque increases gradually. It can be seen from the distribution diagram of engine operating points that the improved strategic operating points are distributed in a reasonable range and work in the high-efficiency zone.

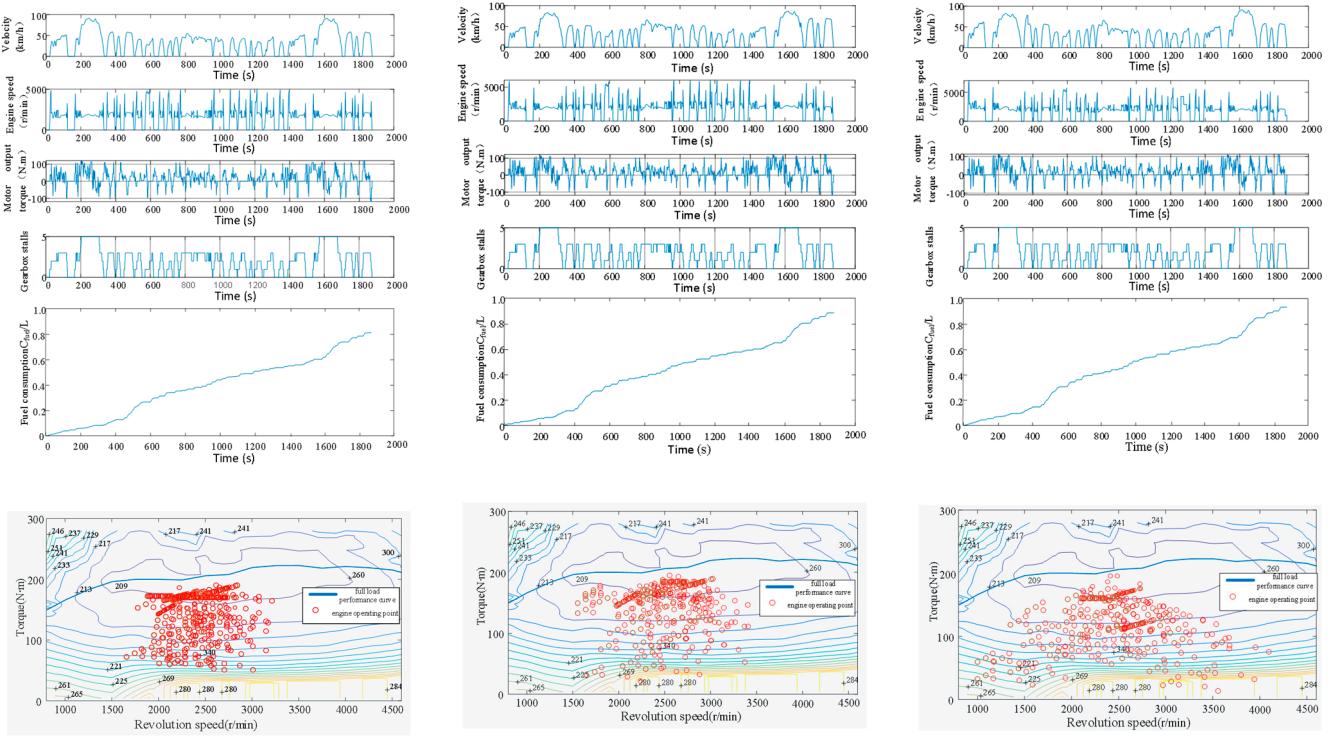


Figure 7. Simulation result.

As shown in Figure 7b, the engine working point of the TD3-based energy management strategy is similar to the TD3 strategy, considering working condition prediction to some extent. However, compared with the unimproved TD3 strategy, the engine working point of the improved TD3 strategy is more located in the high-efficiency zone with low fuel consumption.

As shown in Figure 7c, the rule-based control strategy has an engine working area of 50 N·m–240 N·m, compared with the TD3 strategy, which has a more extensive operating range of 150 N·m–200 N·m. The improved TD3 strategy has lower fuel consumption and continuous control.

As can be seen from Figure 8, the strategy of TD3, considering working condition prediction, has a higher and fuller utilization rate of the motor. When the vehicle starts at the early stage, the driving motor starts to work and gives play to its characteristics of low speed and large torque to avoid the engine working in the inefficient zone. During braking, the braking energy is recovered. The final SOC value of the three strategies fluctuates around 0.3. As can be seen from the simulation results, the fuel consumption of the TD3 energy management strategy considering working condition prediction is reduced by 3.18% and 7.63% compared with the TD3 energy management strategy and rule control strategy. The simulation results are shown in Table 3.

Table 3. Comparison of simulation results.

Control Strategy	Equivalent Fuel Consumption (L/100 km)	Final SOC
Consider the TD3 of the condition prediction	6.411	0.33
TD3	6.622	0.32
Rule-based control	6.941	0.31

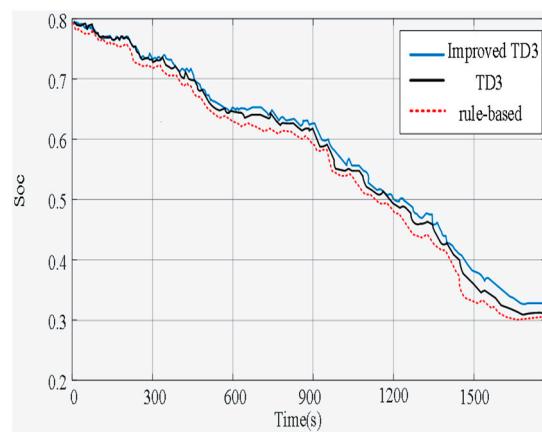


Figure 8. Soc change diagram under driving condition 1.

5.2. Adaptability Verification

In this paper, we utilize MATLAB, Python, PyCharm, and other software to conduct joint simulations. The deep reinforcement learning considering driving condition prediction is verified by comparing the other two driving conditions and the adaptability of the energy management strategy. The simulation model still adopts the above training model. The simulation results are shown in Figures 9 and 10, and the comparative data of SOC and fuel consumption are shown in Tables 4 and 5.

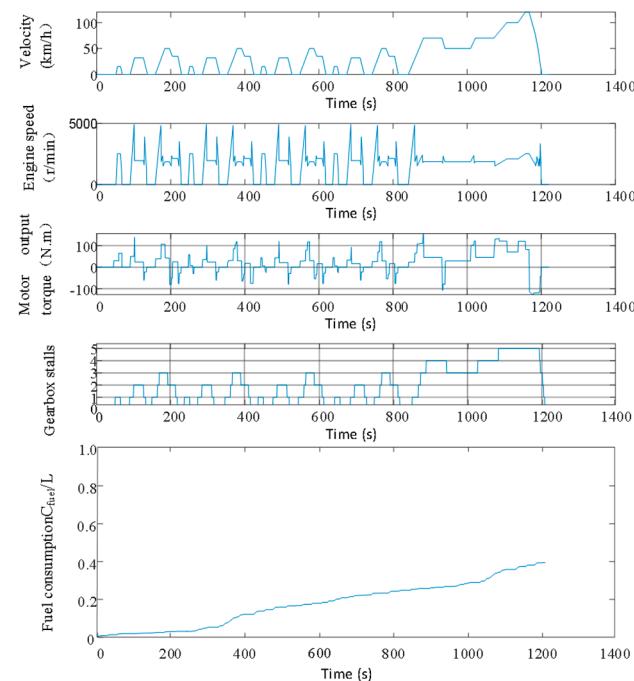


Figure 9. Simulation results of TD3 strategy considering driving mode prediction under mode 2.

Table 4. Comparison of simulation results.

Control Strategy	Equivalent Fuel Consumption (L/100 km)	Final SOC
Consider the TD3 of the condition prediction	6.327	0.35
TD3	6.512	0.34
Rule-based control	6.884	0.34

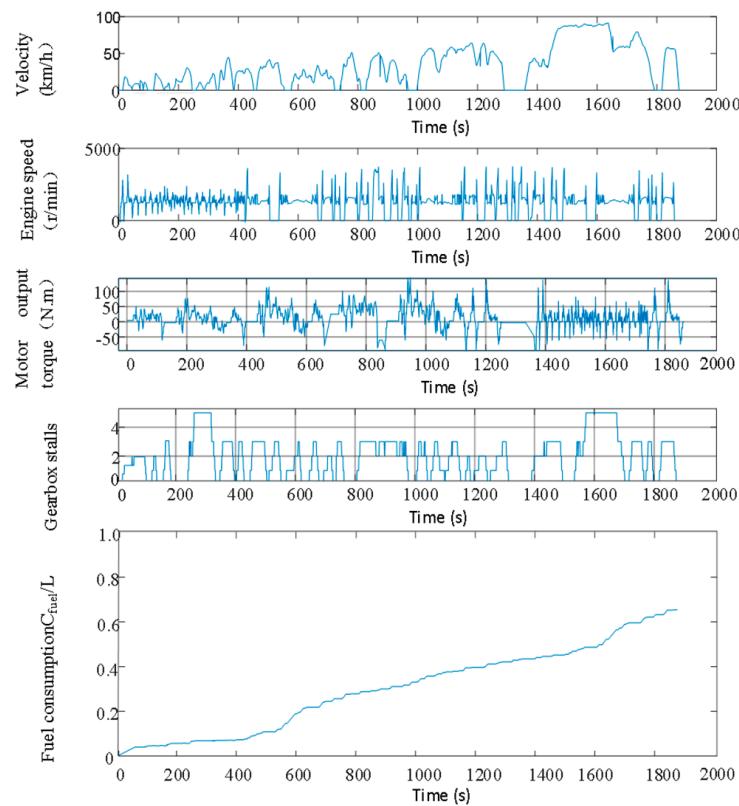


Figure 10. Simulation results of TD3 strategy considering driving mode prediction under driving mode 3.

Table 5. Comparison of simulation results.

Control Strategy	Equivalent Fuel Consumption (L/100 km)	Final SOC
Consider the TD3 of the condition prediction	6.388	0.32
TD3	6.575	0.32
Rule-based control	6.924	0.33

As can be seen from Figures 9 and 10, the proposed strategy can adapt to three different driving conditions and show good fuel economy. The final SOC value of the battery can be stable at around 0.3. Compared with the results, when the final SOC value is roughly the same, compared with the rule control strategy, the TD3 strategy considering driving condition prediction reduces the fuel consumption under different driving conditions by 8.32% and 7.74%, respectively. Compared with the TD3 strategy, the introduction of driving condition prediction reduces fuel consumption by 3.08% and 2.84%, respectively.

6. Conclusions

Based on condition prediction results and the deep reinforcement learning algorithm, this paper proposes an energy management strategy: deep reinforcement learning considering condition prediction. The BP neural network algorithm is used to predict the speed information in the next 5 s, and the obtained correction factor corrects the required torque of the vehicle. The simulation results show that under different training conditions, the proposed strategy can make full use of the characteristics of the drive motor, make the engine work in the optimal range, and make the condition adaptability strong. The introduction of condition prediction effectively reduces the fuel consumption of the energy management strategy with deep reinforcement learning and has a more vital self-learning ability.

Author Contributions: Conceptualization, J.X., J.M., T.W. and J.G.; methodology, J.X., J.M., T.W. and J.G.; software, J.X. and J.M.; validation, T.W. and J.G.; data curation, J.X. and J.M.; writing—original draft preparation, J.M.; writing—review and editing, J.X. and J.M.; visualization, J.X.; supervision, J.G. and T.W.; project administration, J.X.; funding acquisition, J.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Central Plains Technological Innovation leading talents, grant number [224200510014], and The APC was funded by [224200510014].

Data Availability Statement: Not applicable.

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

References

- Li, P.; Li, Y.; Wang, Y.; Jiao, X. An intelligent logic rule-based energy management strategy for power-split plug-in hybrid electric vehicle. In Proceedings of the 2018 37th Chinese Control Conference (CCC), Wuhan, China, 25–27 July 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 7668–7672.
- Pam, A.; Bouscayrol, A.; Fiani, P.; Noth, F. Rule-based energy management strategy for a parallel hybrid electric vehicle deduced from dynamic programming. In Proceedings of the 2017 IEEE Vehicle Power and Propulsion Conference (VPPC), Belfort, France, 14–17 December 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–6.
- Mansour, C.J. Trip-based optimization methodology for a rule-based energy management strategy using a global optimization routine. *Proc. Inst. Mech. Eng. Part D J. Automob. Eng.* **2015**, *230*, 1529–1545. [[CrossRef](#)]
- Mansour, C.; Salloum, N.; Francis, S.; Baroud, W. Adaptive energy management strategy for a hybrid vehicle using energetic macroscopic representation. In Proceedings of the 2016 IEEE Vehicle Power and Propulsion Conference (VPPC), Hangzhou, China, 17–20 October 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1–7.
- Gao, M.; Du, J. Design method of energy management strategy for range-extended electric buses based on convex optimization. In Proceedings of the 2016 11th International Forum on Strategic Technology (IFOST), Novosibirsk, Russia, 1–3 June 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 286–290.
- Zhao, K.; Bei, J.; Liu, Y.; Liang, Z. Development of global optimization algorithm for series-parallel PHEV energy management strategy based on radar pseudospectral knotting method. *Energies* **2019**, *12*, 3268. [[CrossRef](#)]
- Wu, J.; Cui, N.-X.; Zhang, C.-H.; Pei, W.-H. PSO algorithm-based optimization of plug-in hybrid electric vehicle energy management strategy. In Proceedings of the 2010 8th World Congress on Intelligent Control and Automation, Jinan, China, 6–9 July 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 3997–4002.
- Mason, K.; Grijalva, S. A review of reinforcement learning for autonomous building energy management. *Comput. Electr. Eng.* **2019**, *78*, 300–312. [[CrossRef](#)]
- Lee, H.; Song, C.; Kim, N.; Cha, S.W. Comparative analysis of energy management strategies for HEV: Dynamic programming and reinforcement learning. *IEEE Access* **2020**, *8*, 67112–67123. [[CrossRef](#)]
- Sun, H.; Fu, Z.; Tao, F.; Zhu, L.; Si, P. Data-driven reinforcement-learning-based hierarchical energy management strategy for fuel cell/battery/ultracapacitor hybrid electric vehicles. *J. Power Sources* **2020**, *455*, 227964. [[CrossRef](#)]
- Wu, Y.; Tan, H.; Peng, J.; Zhang, H.; He, H. Deep reinforcement learning of energy management with continuous control strategy and traffic information for a series-parallel plug-in hybrid electric bus. *Appl. Energy* **2019**, *247*, 454–466. [[CrossRef](#)]
- Lee, H.; Cha, S.W. Reinforcement learning based on equivalent consumption minimization strategy for optimal control of hybrid electric vehicles. *IEEE Access* **2020**, *9*, 860–871. [[CrossRef](#)]
- Aljohani, T.M.; Mohammed, O. A Real-Time Energy Consumption Minimization Framework for Electric Vehicles Routing Optimization Based on SARSA Reinforcement Learning. *Vehicles* **2022**, *4*, 1176–1194. [[CrossRef](#)]
- Li, W.; Cui, H.; Nemeth, T.; Jansen, J.; Ünlübayir, C.; Wei, Z.; Feng, X.; Han, X.; Ouyang, M.; Dai, H.; et al. Cloud-based health-conscious energy management of hybrid battery systems in electric vehicles with deep reinforcement learning. *Appl. Energy* **2021**, *293*, 116977. [[CrossRef](#)]
- Li, W.; Cui, H.; Nemeth, T.; Jansen, J.; Ünlübayir, C.; Wei, Z.; Zhang, L.; Wang, Z.; Ruan, J.; Dai, H.; et al. Deep reinforcement learning-based energy management of hybrid battery systems in electric vehicles. *J. Energy Storage* **2021**, *36*, 102355. [[CrossRef](#)]
- Yue, S.; Wang, Y.; Xie, Q.; Zhu, D.; Pedram, M.; Chang, N. Model-free learning-based online management of hybrid electrical energy storage systems in electric vehicles. In Proceedings of the IECON 2014-40th Annual Conference of the IEEE Industrial Electronics Society, Dallas, TX, USA, 29 October–1 November 2014; IEEE: Piscataway, NJ, USA, 2014.
- Li, W.; Zheng, C.; Xu, D. Research on Energy Management Strategy of Fuel Cell Hybrid Electric Vehicle Based on Deep Reinforcement Learning. *J. Integr. Technol.* **2021**, *10*, 47–60.
- Tang, X.; Chen, J.; Liu, T.; Li, J.; Hu, X. Research on Intelligent Following Control and Energy Management Strategy of Hybrid Electric Vehicle Based on Deep Reinforcement Learning. *Chin. J. Mech. Eng.* **2021**, *57*, 237–246.
- Zhao, C. Study on Integrated Optimization Control Strategy of Fuel Consumption and Emission of PHEVs. Master’s Thesis, Chongqing Jiaotong University, Chongqing, China, 2022.

20. Zhang, S.; Wang, K.; Yang, R.; Huang, W. Research on the Energy management strategy of Deep Reinforcement Learning for hybrid electric Bus. *Chin. Intern. Combust. Engine Eng.* **2021**, *42*, 10–16+22.
21. Xu, L.; Wang, J.; Chen, Q. Kalman filtering state of charge estimation for battery management system/based on a stochastic fuzzy neural network battery model. *Energy Convers. Manag.* **2012**, *53*, 33–39. [[CrossRef](#)]
22. Tan, F.; Yan, P.; Guan, X. Deep reinforcement learning: From Q-learning to deep Q-learning. In *Neural Information Processing, Proceedings of the 24th International Conference, ICONIP 2017, Guangzhou, China, 14–18 November 2017; Proceedings, Part IV 24*; Springer International Publishing: Berlin/Heidelberg, Germany, 2017; pp. 475–483.
23. Agostinelli, F.; Hocquet, G.; Singh, S.; Baldi, P. From reinforcement learning to deep reinforcement learning: An overview. In *Braverman Readings in Machine Learning. Key Ideas from Inception to Current State, Proceedings of the International Conference Commemorating the 40th Anniversary of Emmanuil Braverman’s Decease, Boston, MA, USA, 28–30 April 2017; Invited Talks*; Springer International Publishing: Berlin/Heidelberg, Germany, 2018; pp. 298–328.
24. Lyu, L.; Shen, Y.; Zhang, S. The Advance of reinforcement learning and deep reinforcement learning. In Proceedings of the 2022 IEEE International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA), Changchun, China, 25–27 February 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 644–648.

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