International Journal of Computer Science and Information Technology
ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 2, Number 1, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v2n1.22
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Energy Management for Hybrid Energy Storage System in Electric Based on Deep Deterministic Policy Gradient

Shuai Xia*, Chun Wang

School of Mechanical Engineering, Sichuan University of Science and Engineering, Yibin, 644000, Sichuan, China

ABSTRACT

In this paper, an intelligent control system design scheme based on deep deterministic policy gradient (DDPG) algorithm is proposed for the complex continuous action space problem in the hybrid energy storage system of electric vehicles. Firstly, the basic principle and internal logic of DDPG algorithm are introduced, including key elements such as Actor-Critic architecture, experience playback, target network, reward signal, policy gradient and value function update. Then, how to apply the DDPG algorithm to the industrial control system is described in detail. The Actor network learns the optimal strategy, the Critic network evaluates the value of the state-action pair, and uses the experience playback and the target network to improve the system stability and performance. Finally, the effect of the intelligent control system based on DDPG algorithm in complex environment is verified by simulation experiments. The results show that the system can effectively optimize the control strategy, improve the response speed and stability of the system, and has a good engineering application prospect.

KEYWORDS

Hybrid energy storage system, Energy management, Deep reinforcement learning, DDPG

1. INTRODUCTION

Vigorously developing new energy vehicles is an important means of China's "double carbon" goal. It is well known that lithium batteries have the characteristics of high specific energy. Although there is great technological progress in power characteristics compared with the past, large rate and high frequency current have a great impact on the internal electrochemical characteristics of the battery. The supercapacitor is mainly used to provide short-term peak power and high-frequency power, playing the role of 'peak shaving and valley filling'. Therefore, the core of energy management of electric vehicle hybrid energy storage system (HESS) is to efficiently allocate the power output of lithium battery and super capacitor on the basis of considering the power demand and key state information of actual driving conditions, so as to realize the efficient collaborative optimization and control of HESS. In general, the energy management strategy of hybrid power supply can be divided into three categories: rule-based energy management strategy (EMS), optimization-based energy management strategy and artificial intelligence-based energy management strategy. Among them, the rule-based energy management strategy has the characteristics of low algorithm complexity, small amount of calculation and simple structure. However, its dynamic adjustment efficiency is poor, and it has a strong dependence on the experience of engineers. Therefore, it is rarely used alone in composite power supply, and is mainly used to compare and evaluate the efficiency of various optimization algorithms. Liu Yonggang [1] improved the rules by hierarchical clustering and achieved good results. However, the rules used in this strategy are relatively simple and cannot

guarantee the optimal performance output of the system under other complex conditions. Rodriguez [2] used fuzzy logic prediction algorithm to predict the load of FC-HEV, which improved the operation efficiency of the HESS. K. V. Singh [3] introduced Elman neural network on the basis of fuzzy control, with minimum fuel consumption and battery life as the optimization objectives. The experimental results show that compared with the traditional strategy, the proposed strategy has better fuel economy and faster response. However, if there is no reasonable fuzzy rule base, the control effect of fuzzy rules will be greatly reduced.

Compared with the rule-based energy management strategy, the optimization-based energy management strategy can be closer to the optimal control, can guide other energy management rule control, and is often used as a standard to evaluate the advantages and disadvantages of other control strategies. However, the optimization-based control method has a large number of complicated calculations, which will greatly increase the computational burden and even cause 'dimension disaster'. Q. Zhang [4] proposed a real-time energy management strategy composed of neural network, wavelet transform and fuzzy control. Among them, the wavelet transform is used to decompose the required power to form an offline data set for training, the neural network is used to predict the required power of the vehicle online, and the fuzzy controller is used to control the voltage of the supercapacitor within a suitable range. Finally, based on the developed real-time simulation platform of battery-supercapacitor HESS, the effectiveness of the proposed energy management strategy is verified. Aiming at the energy management problem of fuel cell, battery and supercapacitor HESS, A. U. Rahman [5] designed a nonlinear controller based on super-twisting sliding mode control. The simulation results show that the proposed strategy can effectively reduce fuel consumption by 29%. In addition, the results of the robustness test verify that the controller changes the external and internal parameters. A. Prasanthi [6] used the improved butterfly algorithm to optimize the management strategy with the goal of improving work efficiency. Zhang Qiao [7] combined wavelet transform, neural network and fuzzy logic to realize real-time and efficient power distribution of composite power supply. With the in-depth development of Internet technology and machine learning algorithms, more and more researchers have devoted themselves to integrating machine learning into the research of composite power supply. This artificial intelligence algorithm can make the optimal energy management strategy of composite power supply system configuration under the condition of unknown system structure and parameters. Ouyang Minggao [8] aimed at the hybrid structure of DDPG (deep deterministic strategy gradient), relying on deep reinforcement learning to reduce the energy loss and aging cost of composite energy storage system. He Hongwen [9-10] also proposed an energy management strategy for hybrid electric vehicles based on DQN and improved DDPG algorithm. The simulation training results show that the fuel economy gap between the strategy and the dynamic programming is reduced to 6.4 %. Xu [11] used the parallel computing SAC algorithm for energy management of the power system. The results show that the SAC-based strategy reduces the energy loss by an average of 7 % compared with the DQN-based and DDPG-based strategies. W.H. Li [12] integrated a new reward condition into deep reinforcement learning to manage the energy of the composite power supply. This algorithm improves the robustness of the strategy, but whether the operation effect of the strategy under other conditions and working conditions is up to standard still needs further research. P. Wu et al. [13] proposed an energy management strategy based on adaptive deep reinforcement learning, using an improved two-layer Q-learning method. The training results show that the strategy can significantly reduce the training time and battery cost of the data. Aiming at the problem of long learning time in DQN, R. Lian [14] proposed an energy management strategy based on DDPG. The strategy embeds the battery characteristics and the best brake fuel consumption curve into the DDPG to improve the training efficiency. The simulation results show that the strategy can accelerate the learning process and has better fuel economy.

2. THE MODEL OF HESS

2.1. The topology of HESS

The topology of the composite power supply has a direct impact on the effect of the energy management strategy. At present, the topology of composite power supply can be divided into passive structure, semi-active structure and fully active structure. In the passive topology, the battery is directly connected in parallel with the super capacitor, which has the advantages of simple structure and low cost. However, the output power of the battery and the super capacitor cannot be decoupled, which is not conducive to the efficient play of the composite power supply system. In the fully active topology, the battery and the super capacitor are connected in series with the bidirectional DC/DC converter respectively, and then connected in parallel. The advantage is that the output power of the battery and the super capacitor can be directly controlled, but too many DC/DC converters will reduce the energy efficiency of the composite power system.

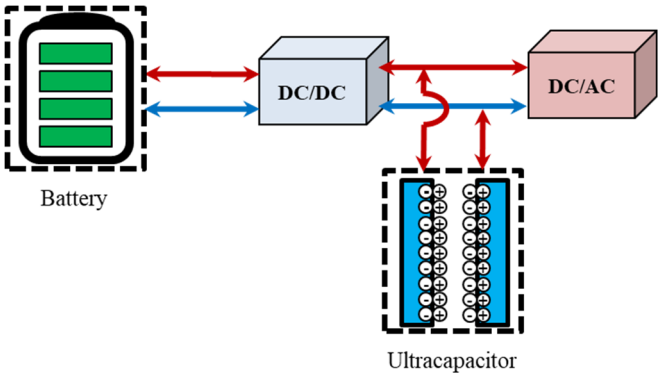

Figure 1 The topology of HESS

2.2. The model selection of HESS

At present, there are three main types of lithium-ion battery and supercapacitor models: electrochemical model, equivalent circuit model and neural network model. Among them, the electrochemical model can describe the electrochemical characteristics and dynamic behavior of supercapacitors more accurately. However, the complexity of the model is high, which requires a large number of parameters to be identified. The huge amount of calculation requires high-performance microcontrollers and the excessive amount of calculation limits its real-time application. The neural network model can simulate the input and output of highly nonlinear mapping, which is very suitable for the modeling of batteries and supercapacitors. However, the establishment of neural network models requires a lot of data for training, and how to ensure the quality of training data is also a key issue. The equivalent circuit model is widely used in the simulation of electric vehicles because of its good adaptability to various working states of power batteries and supercapacitors. Considering the complexity and accuracy of the equivalent circuit model, the Thevenin model and the Rint model are selected to model the lithium battery and the supercapacitor respectively. The specific structure is shown in Figure.2.

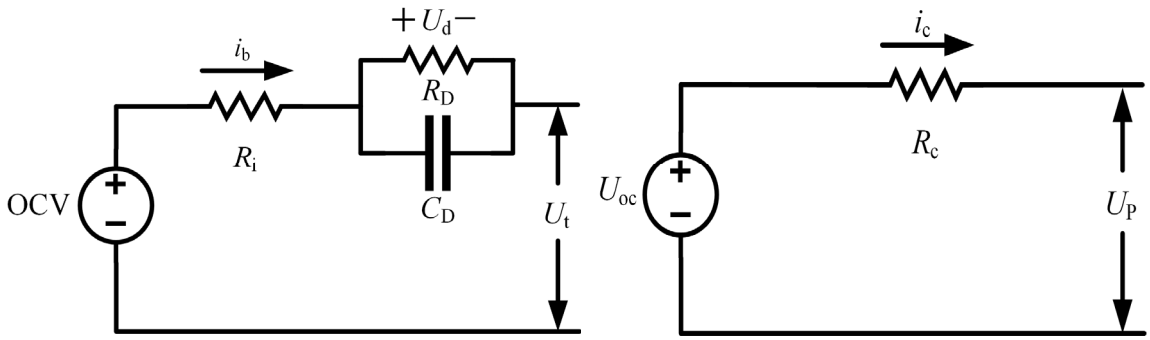

Figure 2 The model of Battery and Ultracapacitor. (a) Thevenin; (b) Rint.

Where OCV represents the open circuit voltage of the battery, Ri represents the DC internal resistance, RD and CD represent the polarization internal resistance and polarization capacitance of the battery, respectively, Ud and Ut represent the voltage divider and terminal voltage of the RC network. Uoc, Rc and UP represent the open circuit voltage, DC internal resistance and terminal voltage of the supercapacitor, respectively. In addition, the state of charge (SOC) of the HESS is defined as:

$$SOC(t)=SOC_{ini}(t_0)-\frac{\int_{t_0}^{t}\eta i_L dt}{C_a} \quad (1)$$

In the formula, SOC(t) represents the estimated value of SOC at time t, SOCini (t0) represents the initial value of SOC, represents the charge and discharge efficiency, and Ca is the rated capacity of the battery or supercapacitor.

2.3. Dynamic experiment test

HPPC (Hybrid Pulse Power Characterization) experiment is a standardized experimental method for battery system characteristic test. By applying pulse current to simulate the dynamic working conditions of the battery in actual use, the response characteristics of the battery at different charge and discharge rates are evaluated, including voltage response, internal resistance change, etc. In addition, the energy storage and release capacity of the battery at different charge and discharge rates, as well as its maximum power output capacity, can be obtained, which provides an important reference for the design and application of the battery system. Through repeated HPPC experiments, the change of battery capacity with the number of cycles can be monitored, so as to predict the life and attenuation trend of the battery, and provide a basis for the optimization of the battery management system. The experimental results of the dynamic characteristics of the battery and the supercapacitor at 25°C are shown in Figure 3.

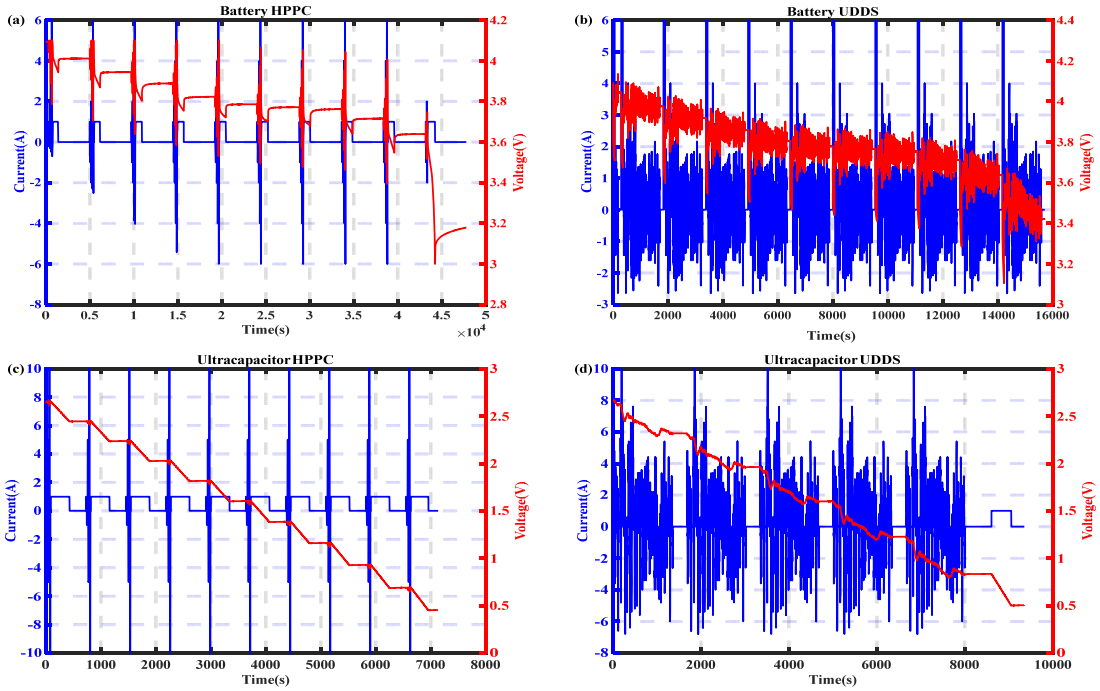

Figure 3 The results of HPPC and UDDS experiments. (a)(b) Battery; (c)(d) Ultracapacitor.

3. DDPG-BASED ENERGY MANAGEMENT STRATEGY

In reinforcement learning, the state variables and action variables of the agent are discrete. However, in most engineering problems, the state variables and action variables of the engineering problems to be solved are continuous. In addition, when the state variables and action variables are discretized in reinforcement learning, the number of discretization's needs to be carefully considered to avoid too large state-action space, which will make the algorithm difficult to converge. Therefore, the powerful representation ability of deep learning is used to fit the Q table to solve the ' dimension disaster ' problem and the continuous state-action space problem caused by variable dispersion. The traditional deep reinforcement learning algorithm uses the deep network to characterize the value function on the basis of the reinforcement learning algorithm, and refers to the Q-Learning in reinforcement learning to continuously provide the target value for the deep network until the entire deep network converges. However, the traditional deep reinforcement learning algorithm has two problems that need to be improved. First, it can only deal with discrete and low-dimensional action space, and the discrete action set is not conducive to the learning of optimal strategies. Second, due to its value-based learning method, it is easy to fall into local optimum. For the first problem, the deterministic policy gradient algorithm transforms the policy into a policy function, and maps the state to a certain action to solve the problem of discontinuous action space. The deterministic policy gradient algorithm is combined with the traditional DQN algorithm to form a deep deterministic policy gradient algorithm. With its actor-critic architecture, end-to-end learning is achieved directly from the original data. Therefore, in order to overcome the discrete output problem based on the DQN method, the deep deterministic strategy gradient is proposed to improve the above phenomenon. In addition, the DDPG algorithm can also improve the stability of agent learning.

3.1. DDPG algorithm

DDPG is an algorithm for solving deep reinforcement learning problems in continuous action space. It combines the advantages of DQN (Deep Q-Network) and deterministic policy gradient method, and can effectively deal with the problems of high-dimensional continuous action space and high-dimensional state space. The DDPG algorithm uses a deep neural network to approximate the value

function and the policy function. The value function is used to evaluate the value of the state-action pair, and the strategy function is used to generate the action. The update of the value function adopts the idea of Q-learning, and the update of the policy function adopts the policy gradient method. We use the policy gradient method to update the parameters of the Actor network. By maximizing the gradient of the action value function to the policy function, the parameters of the policy function are updated to enable the agent to learn a better strategy. In order to improve the stability and convergence of the algorithm, the DDPG algorithm uses the Experience Replay mechanism. In the training process, Agent stores the information such as state, action, reward, and next state in the experience pool, and randomly samples them as training data to reduce the correlation between samples. DDPG algorithm introduces target networks. The target network is a copy of the Critic network and the Actor network, but its parameter update is slower than the original network. The parameters of the target network are gradually updated by soft update to reduce the variance of parameter update and improve the stability of the algorithm. The Critic network updates the parameters by minimizing the loss function of the value function to reduce the gap between the predicted value of the value function and the target value, so as to better evaluate the value of the state-action pair. In the Actor network of DDPG, the goal of the Actor network is to maximize the action value function Q(s, a; θQ). The parameter θμ of the Actor network is updated by the policy gradient method, so that the selection of action a can maximize the value function Q(s, a; θQ), and its policy gradient is updated as :

$$\nabla_{\theta_\mu} J \approx \mathbb{E}_{s_t \sim \rho, a_t \sim \mu}[\nabla_{\theta_\mu} \mu(s_t; \theta_\mu) \nabla_a Q(s, a; \theta_Q) |_{s=s_t, a=\mu(s_t)}] \tag{2}$$

where si is the state, ai is the action, and ri is the reward. γ denotes the discount factor used to measure the importance of future rewards. Q(s, a|θQ) represents the value function, which is used to evaluate the value of a given state s and action a, where θQ represents the parameter of the value function. μ(s) denotes the policy function, which is used to generate actions under a given state, where θμ denotes the parameter of the policy function. α and β represent the learning rate, which is used to control the step size of the parameter update. N denotes the batch size at each update.

This step expresses the gradient update direction of the Actor network parameter θμ, by maximizing the value function Q(s, a; θQ) on the strategy function μ (s; ω) to update the parameters of the Actor network. Moreover, the update parameters of the Actor network are:

$$\theta_\mu \leftarrow \theta_\mu + \beta \nabla_{\theta_\mu} J \tag{3}$$

The objective of the Critic network is to minimize the loss function L(θQ) of the value function, and the loss function measures the gap between the predicted value of the value function and the target value. Among them, the target value of the Critic network is updated to:

$$y_i = r_i + \gamma Q'(s_{i+1}, \mu'(s_{i+1} \mid \theta_{\mu'}) \mid \theta_{Q'}) \tag{4}$$

The calculation method of the target Q value is :

$$L(\theta_Q) = \frac{1}{N} \sum_i (Q(s_i, a_i \mid \theta_Q) - y_i)^2 \tag{5}$$

Update Critic network parameters :

$$\theta_Q \leftarrow \theta_Q - \alpha \nabla_{\theta_Q} L(\theta_Q) \tag{6}$$

The parameters θμ' and θQ' update the target network step by step through soft update. The target network can improve the stability and convergence of the algorithm by slowly updating the parameters of the target network.

3.2. The selection of variables and reward function

Under the deep reinforcement learning framework, the power allocation strategy of the power management system is obtained by the interaction between the DDPG agent and the external environment of the vehicle and the key operating state of the composite power system. The agent can only improve the strategy by interacting with the environment. Therefore, the selection of agent environment and action variables is very important for the learning of optimal strategy. Like reinforcement learning, the required power Preq, battery SOCbat and ultracapacitor SOCbuc are selected to form state variables. Since deep reinforcement learning can deal with continuous state space problems, the polarization voltage of the Thevenin model can be excluded from the state variables. Therefore, under the framework of energy management strategy based on DDPG, the state variable is, and the action variable is the output current Ib of the battery pack. In addition, the reward function also plays a vital role in the deep reinforcement learning algorithm. A good reward function can accelerate the convergence speed of the neural network in the agent, which can greatly shorten the time required for agent training. In the energy management strategy based on DDPG algorithm, the reward function is consistent with the reward function of the energy management strategy based on reinforcement learning, as shown in the formula.

$$r_t = -\left(L_{\text{bat}} + L_{\text{uc}} + L_{\text{dc}} + G_{\text{uc}}\right) \tag{7}$$

Among them, the specific loss of each part of the component is calculated by the following formula:

$$\begin{cases} L_{\text{bat}}(k) = I_{\text{bat}}^2(k) \times R_{\text{c}}(k) + U_{\text{R}}^2(k) \,/\, R_{\text{u}}(k) \\ L_{\text{uc}}(k) = I_{\text{uc}}^2(k) \times R_{\text{p}}(k) \\ L_{\text{dc}}(k) = P_{\text{bat}}(k) \times \left(1 - \varepsilon_{\text{dc}}(k)\right) \varepsilon_{\text{dc}}^{-\text{W}}(k) \\ G_{\text{uc}}(k) = \beta\left(SOC_{\text{ucref}} - SOC_{\text{uc}}\right)^2 \end{cases} \tag{8}$$

3.3. Training settings

The initial SOCbat of the battery will be reset after each scenario training is completed to expand the adaptability of the DDPG agent in the full SOCbat interval. The parameter setting and network structure of the DDPG agent are shown in Table 1. In addition, in order to speed up the convergence of the DDPG algorithm, when the state variable reaches the system limit, the training process will end immediately, and a large penalty is added to the reward function to accelerate the elimination of incorrect learning experience.

Table 1 Agent parameters and network settings

Parameters	Values
Actor networks	32/32/16
Critic networks	32/32/16
Actor learning rates	0.0005
Critic learning rates	0.002
Discount factor	0.995
Minibatch size	256
Target smooth factor	0.001

4. SIMULATION RESULTS AND DISCUSSION

In order to fully illustrate the advantages of the proposed DDPG-based energy management strategy, the proposed strategy is compared with the rule-based strategy. The battery SOCbat, super capacitor SOCuc and their corresponding current comparison results under different temperature and different energy management strategies are shown in Figure 4. In addition, in order to compare the effects of

different energy management strategies in more detail, some characteristic parameters of the HESS, such as the terminal SOCbat and the terminal SOCuc, are summarized in Table 2.

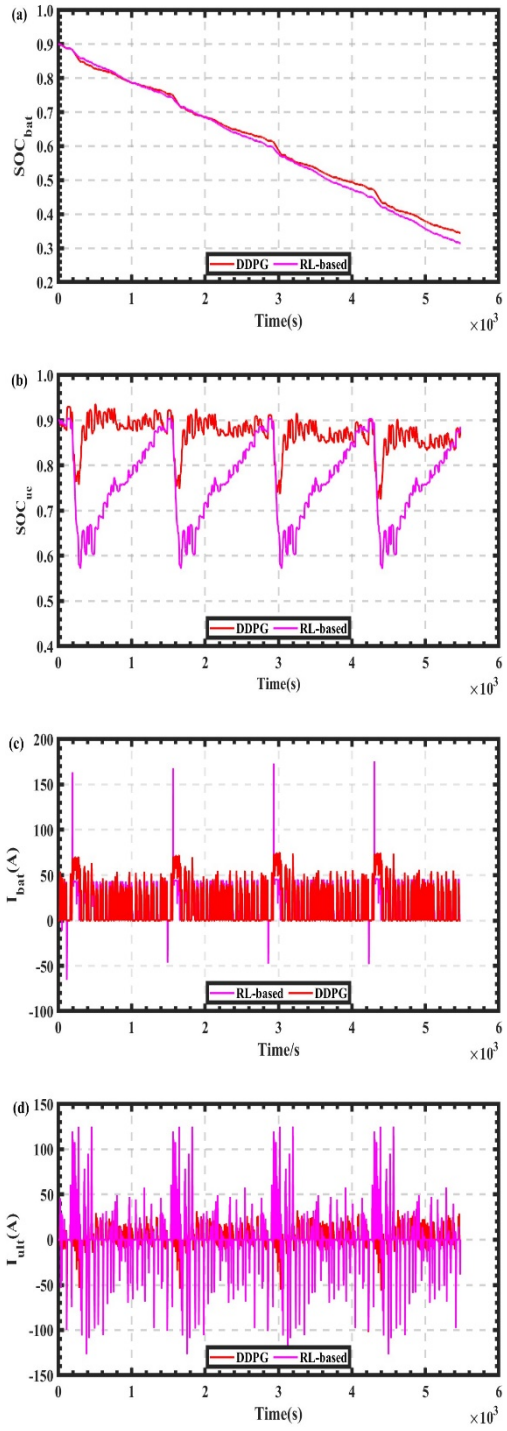

Figure 4 The results of Battery and Ultracapacitor

Table 2 Characteristic parameters in different control strategies

	Rule-based	DDPG-based
Battery SOC_{bat}	0.2894	0.3452
Ultracapacitor SOC_{uc}	0.8856	0.8461
Max Battery I_{bat}	175.6661	75.1178
Max Ultracapacitor I_{uc}	125.4566	101.9870
Max Ultracapacitor I_{uc}	125.4566	101.9870

From Table 2, it can be seen that the SOC at the end of the battery and the supercapacitor is 0.2894 and 0.3452, respectively, which clearly reflects that the energy management strategy based on DDPG has better performance characterization and can be better completed. The purpose of extending the battery life. In addition, the SOC values based on the deterministic rule and the DDPG-based energy management strategy are 0.8856 and 0.8461 at 25°C, respectively, which indicates that the DDPG-based EMS can give full play to the excellent characteristics of supercapacitors that can cut peaks and fill valleys. Among them, the EMS based on DDPG is more economical than that based on deterministic rules. Specifically, compared with the deterministic strategy, the DDPG-based strategy improves the economy by 19.28%. The DDPG-based strategy has made significant improvements in state perception and action output, and solved the problem of discontinuous state action space, which is also one of the important reasons for the economic improvement of the rule-based energy management strategy.

5. CONCLUSION

In this study, the energy management strategies of HESS based on deterministic rules and deep deterministic policy gradient (DDPG) are compared and analyzed. In summary, the DDPG algorithm, as a reinforcement learning method that combines deterministic policy gradient and experience playback technology, performs well in solving the problem of continuous action space. Through the approximation function and strategy function of neural network, DDPG algorithm can effectively learn complex strategies, and has the characteristics of strong adaptability and high training stability. This strategy has better energy economy and robustness, and can respond to changes in actual operating conditions more quickly.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support by The Innovation Fund of Postgraduate, Sichuan University of Science & Engineering (Grant No. Y2022061).

REFERENCES

[1] Y.G. Liu, J.J. Liu, Y.J. Zhang, et al. Rule learning based energy management strategy of fuel cell hybrid vehicles considering multi-objective optimization[J]. Energy, 2020, 207:118212.

[2] R. Rodriguez, J.P.F. Trovão, J. Solano. Fuzzy logic-model predictive control energy management strategy for a dual-mode locomotive[J]. Energy Conversion and Management, 2022, 253:115111.

[3] Singh K V, Bansal H O and Singh D. Fuzzy logic and Elman neural network tuned energy management strategies for a power-split HEVs [J]. Energy, 2021. 225.

[4] [26] Zhang Q, Wang L, Li G. A real-time energy management control strategy for battery and supercapacitor hybrid energy storage systems of pure electric vehicles [J]. Journal of Energy Storage, 2020. 31.

[5] [27] Rahman A U, Zehra S S, Ahmad I. Fuzzy supertwisting sliding mode-based energy management and control of hybrid energy storage system in electric vehicle considering fuel economy [J]. Journal of Energy Storage, 2021. 37.

[6] A. Prasanthi, H. Shareef, M. Asna, et al. Optimization of hybrid energy systems and adaptive energy management for hybrid electric vehicles[J]. Energy Conversion and Management, 2021, 243:114357.

[7] Q. Zhang, L.J. Wang, G. Li, et al. A real-time energy management control strategy for battery and supercapacitor hybrid energy storage systems of pure electric vehicles[J]. Journal of Energy Storage, 2020, 31:101721.

[8] W.H. Li, H. Cui, T. Nemeth, et al. Cloud-based health-conscious energy management of hybrid battery systems in electric vehicles with deep reinforcement learning[J]. Applied Energy, 2021, 293:116977.

[9] H.E. He, J.F. Cao, X. Cui. Energy optimization of electric vehicle's acceleration process based on reinforcement learning[J]. Journal of Cleaner Production, 2020, 248:119302.

[10] Y.C. Li, H.W. He, A. Khajepour, et al. Energy management for a power-split hybrid electric bus via deep reinforcement learning with terrain information[J]. Applied Energy, 2019, 255:113762.

[11] D.Z. Xu, Y.D. Cui, J.Y. Ye, et al. A soft actor-critic-based energy management strategy for electric vehicles with hybrid energy storage systems[J]. Journal of Power Sources, 2022, 524:231099.

[12] W.H. Li, H. Cui, T. Nemeth, et al. Deep reinforcement learning-based energy management of hybrid battery systems in electric vehicles[J]. Journal of Energy Storage, 2021, 36:102355.

[13] P. Wu, J. Partridge, E. Anderlini et al. Near-optimal energy management for plug-in hybrid fuel cell and battery propulsion using deep reinforcement learning[J]. International Journal of Hydrogen Energy, 2021, 46(80):40022-40040.

[14] Lian R, Peng J, Wu Y. Rule-interposing deep reinforcement learning based energy management strategy for power-split hybrid electric vehicle [J]. Energy, 2020. 197.