On-Line Optimization of Microgrid Operating Cost Based on Deep Reinforcement Learning

S H Lin¹, H H Yu¹,² and H W Chen¹

¹School of Electric Power, South China University of Technology, Guangzhou 510640, China
E-mail: hohuiyu@163.com

Abstract. In view of the microgrid random optimization scheduling problem, this paper proposes a microgrid online optimization algorithm based on deep reinforcement learning. By using the approximate state-action value function of the deep neural network, the action of the battery is discretized into the decision variables output by the neural network, and then the remaining decision variables are solved by nonlinear programming and the immediate return is calculated. The optimal strategy is obtained by using the Q-learning algorithm. In order to make the neural network adapt to the randomness of wind, photovoltaic and load power, according to the wind, photovoltaic and load power prediction curves as well as wind, photovoltaic and load prediction errors, Monte Carlo sampling was used to generate multiple sets of training curves to train the neural network. After the training is completed, the weight is fixed and the real-time action value of the battery is output according to the real-time status of the microgrid, so as to realize the online optimal dispatching of the microgrid. Compared with day-ahead optimization results under different fluctuations of wind, photovoltaic and load power, the effectiveness and superiority of this algorithm in online optimization of microgrid are verified.

1. Introduction

Due to the non-renewability of traditional energy, the development and effective utilization of new energy has become the focus of research. As an effective access method, the microgrid composed of distributed power sources has begun to be widely used. Compared with traditional energy, wind and solar energy generation is random and intermittent, which brings great challenges to the optimal dispatch of microgrid. Therefore, how to propose a reasonable and effective scheduling plan has become an important topic.

For the optimal dispatching of microgrid, a large number of algorithms have emerged, including traditional classical optimization algorithms [1,2], heuristic algorithms [3-5], stochastic programming [6] and robust optimization [7,8] algorithm etc. Reference [7] considered the uncertainty of renewable energy and load, and established a microgrid optimal dispatch model based on two-stage robust optimization. Reference [9] adopted fuzzy logic algorithm, designed an adaptive droop controller to eliminate the influence of line resistance on power distribution, and realized the reasonable distribution of unbalanced power in DC microgrid. Reference [10] divides the demand side into two levels for optimization. The first level optimizes the output and transferable load of each micro power supply based on the wind and load power forecast curves; the second level further optimizes the load based on the expected supply curve obtained by the dispatching department through historical data and the current operating conditions of the equipment. In reference to the nonlinear multi-constraint
characteristics of the microgrid optimal dispatching model, the reference [11] proposed an improved bat algorithm to solve the model by combining the horizontal cross strategy, the conversion adjustment mechanism and the two-way learning.

Most of the above algorithms are used for day-ahead optimization, and less for intra-day optimization. At present, the online optimization algorithms of microgrid mainly include model predictive control algorithm [12], alternating direction multiplier algorithm [13], dynamic programming algorithm [14] and so on. Reference [12] uses model predictive control algorithms to use state information at the current moment and several periods in the future to achieve rolling optimization. Reference [13] proposed an online optimization algorithm based on the alternating direction multiplier algorithm to solve the problem of high uncertainty in micro power supplies. Reference [14] uses approximate dynamic programming theory, uses piecewise linear function to approximate the value function, and applies it to online optimization of microgrid. As the most popular optimization algorithm recently, artificial intelligence algorithms have also been applied to the optimization of microgrids. Reference [15] uses the Q-learning algorithm to optimize the action of the battery and improve the utilization of solar generators and batteries. Reference [16] sets the actions of batteries and hydrogen storage devices as charging, discharging and no operation, and uses deep reinforcement Q-learning to achieve the optimal action for online decision-making. This paper proposes a microgrid online optimization algorithm based on deep reinforcement Q-learning (Deep Q Network, DQN), which uses the neural network to approximate the state-action value function, and uses the trained neural network to determine the optimal battery action in real time. The remaining decision variables are solved through nonlinear programming, so as to realize real-time online optimization of microgrid. Different scenarios are set to verify the feasibility of the algorithm.

2. Problem formulation

The microgrid structure is shown in Figure 1 below. The microgrid consists of wind turbines (WT), photovoltaic systems (PV), diesel generators (DG), fuel cells (FC), storage battery (BS), loads (LD) and external Grid composition.

![Figure 1 Micro grid structure](image)

2.1. Objective function

Ignore the cost of scenery, and set up an online optimization model with the goal of minimizing the total economic operating cost during the optimization period. The objective function is as follows

\[
F = \min E \sum_{t=1}^{T} (C_t)
\]

\[
C_t = (C_t + C_{t-1} + C_{t-2} + C_{t-3} + C_{t-4}) \Delta t
\]
Where $T$ is the total optimization period; $\Delta t$ is the step length of each period; $E$ is expected value; $C_t$ is the total economic operating cost of period $t$; $C(t)$ is the operation and maintenance cost of each micro power supply in period $t$; $C_{dis}(t)$ is the depreciation cost of the battery in period $t$; $C_{fuel}(t)$ is the fuel cost in period $t$; $C_{env}(t)$ is the environmental governance cost in period $t$; $C_{int}(t)$ is the interaction cost between the microgrid and the external power grid in period $t$.

$$C_{int}(t) = K_{omt} P_t(t)$$  \hspace{1cm} (2)

Where $K_{omt}$ is operation management coefficient of each micro power supply; $P_t(t)$ is active power of each micro power supply in period $t$.

The depreciation cost of a battery $[17]$ is related to its state of charge (SOC). The change characteristics of the depreciation cost of the battery can be expressed by a quadratic function

$$f = a_1 \Delta SOC(t) + a_2 P_{bat}(t) \Delta SOC(t) + a_3 P_{bat}(t)$$  \hspace{1cm} (3)

Where $a_0, a_1, a_2$ and $a_3$ is the variation characteristic coefficient during charging; $b_0, b_1$ and $b_2$ is the variation characteristic coefficient during discharging. $P_{bat}(t)$ is active power of battery in period $t$; when it is greater than 0, it means the battery is charging; when it is less than 0, it means the battery is discharging; $SOC(t)$ is state of charge in period $t$.

$$C_2(t) = f(SOC(t), P_{bat}(t))$$  \hspace{1cm} (4)

When the battery is discharging, the state of charge of the battery is

$$SOC(t) = SOC(t-\Delta t) + \frac{P_{bat}(t) \Delta t}{\eta_c E_c}$$  \hspace{1cm} (5)

When the battery is charging, the state of charge of the battery is

$$SOC(t) = SOC(t-\Delta t) + \frac{P_{bat}(t) \Delta t}{\eta_d E_d}$$  \hspace{1cm} (6)

Where $\eta_c, \eta_d$ indicates the charge and discharge efficiency of the battery respectively; $E_c$ is the capacity of the battery; $SOC(t-\Delta t)$ is the state of charge of the battery in the previous period.

$$C_{int}(t) = C_{int}(P_f(t)) + C_{fc}(P_f(t))$$  \hspace{1cm} (7)

$$C_{int}(P_f(t)) = a_0 P_f(t)^2 + b_0 P_f(t) + c_0$$  \hspace{1cm} (8)

$$C_{fc}(P_f(t)) = C_{fuel} \left( \frac{P_f(t)}{\eta \cdot LHV_f} \right)$$  \hspace{1cm} (9)

Where $C_{fuel}(P_f(t))$ is fuel cost of diesel generators; $C_{fc}(P_f(t))$ is fuel cost of fuel cell; $a_0, b_0$ and $c_0$ is fuel factor of diesel generator; $P_{fuel}(t)$ is active power of diesel generator in period $t$; $P_f(t)$ is active power of fuel cell in period $t$; $C_{fuel}$ is fuel price; $\eta$ is power generation efficiency; $LHV_f$ is electric energy conversion coefficient.

$$C_s(t) = \sum_{i=1}^{n} C_s \gamma_i P_i(t)$$  \hspace{1cm} (10)

Where $C_s$ is environmental governance cost coefficient; $\gamma_i$ is the emission amount of the $i$-th micro-power source pollutant per unit electric energy.

$$C_s = p_{buy}(t) \times P_{buy}(t) - p_{sell}(t) \times P_{sell}(t)$$  \hspace{1cm} (11)

Where $P_{buy}(t)$ is the price of power purchase in period $t$; $P_{sell}(t)$ is the price of power sold in period $t$; $P_{buy}(t)$ is power purchased in period $t$; $P_{sell}(t)$ is power sold in period $t$.

2.2. Restrictions

A. Power balance constraint

$$P_{wt}(t) + P_{pv}(t) + P_{fuel}(t) + P_{bat}(t) + P_{fuel}(t) + P_{bat}(t) - P_{sell}(t) - P_{sell}(t) = 0$$  \hspace{1cm} (12)

Where $P_{wt}(t)$ is active power of WT in period $t$; $P_{pv}(t)$ is active power of PV in period $t$; $P_{fuel}(t)$ is active load in period $t$.

B. Diesel generator constraint
\[ -R_d \Delta t \leq P_g(t) - P_g(t - \Delta t) \leq R_d \Delta t \]  
\[ P_{g,\text{min}} \leq P_g(t) \leq P_{g,\text{max}} \]  
\[ R_d \] is the downward ramp rate of the diesel generator; \( R_u \) is the upward ramp rate of the diesel generator; \( P_{g,\text{min}} \) and \( P_{g,\text{max}} \) is the lower limit and upper limit of active power of diesel generator respectively.

C. Battery constraint

\[ P_{\text{bat, min}} \leq P_{\text{bat}}(t) \leq P_{\text{bat, max}} \]  
\[ \text{SOC}_{\text{min}} \leq \text{SOC}(t) \leq \text{SOC}_{\text{max}} \]  
Where \( P_{\text{bat, min}} \) and \( P_{\text{bat, max}} \) is the lower limit and upper limit of active power of battery respectively; \( \text{SOC}_{\text{min}} \) and \( \text{SOC}_{\text{max}} \) is the lower limit and upper limit of SOC respectively.

D. Fuel cell constraint

\[ P_{\text{fc, min}} \leq P_{\text{fc}}(t) \leq P_{\text{fc, max}} \]  
Where \( P_{\text{fc, min}} \) and \( P_{\text{fc, max}} \) is the lower limit and upper limit of active power of fuel cell respectively.

E. Interaction constraint between microgrid and external grid

\[ 0 \leq P_{\text{ex, min}}(t) \leq P_{\text{ex, max}} \]  
\[ 0 \leq P_{\text{ex, min}}(t) \leq P_{\text{ex, max}} \]  
\[ P_{\text{ex, min}}(t) \times P_{\text{ex, max}}(t) = 0 \]  
Where \( P_{\text{ex, max}} \) is the upper limit of exchange power between microgrid and external grid respectively.

3. Background and preliminaries

3.1. Reinforcement learning

The learning process of reinforcement learning can be described in Figure 2. The agent interacts with the environment and continuously updates its actions. The agent obtains state \( s_t \) from the environment in period \( t \); the agent chooses the action \( a_t \) at according to the state \( s_t \) and calculate the next state \( s_{t+1} \) and immediately return \( r_t \).

The purpose of reinforcement learning is to find an optimal strategy \( \pi^* \) to maximize the cumulative expected value of immediate return. The state transition process is shown in Figure 3. The cumulative return function in period \( t \) can be expressed as

\[ R_t = \sum_{i=0}^{T} \gamma^{-i} r(s_i, a_i) \]  

Where, \( R_t \) is the cumulative return from the current state to the end state; \( \gamma \) is the attenuation coefficient, the farther away from \( t \), the smaller the weight of the immediate return.

\[ \text{Figure 2. Agent and environment interactions} \]  
\[ \text{Figure 3. State transition} \]

Q-learning uses the state-action value function \( Q^\pi(s,a) \) to represent the expected value of the cumulative return obtained by the agent executing the strategy \( \pi \).
\[
Q^\pi(s, a) = E_s[R_s | s, a, s'] = a
\]

\[
\hat{Q}^\pi(s, a) = \max_{a_i} Q^\pi(s, a)
\]

Where, \(\pi^*(s,a)\) is the strategy that maximizes the state-action value function; \(\hat{Q}^\pi(s, a)\) is optimal state-action value function.

Q-learning algorithm obtains the optimal strategy by solving the following Bellman equation with recursive attributes

\[
Q^\pi(s_t, a_t) = E_s[r_t + \gamma \max_{a_{t+1}} Q^\pi(s_{t+1}, a_{t+1})]
\]

\[Q^* = E_s[r_t + \gamma \max_{a_{t+1}} Q^*(s_{t+1}, a_{t+1})]
\]

3.2. Deep Q network

When the state space is too large, the Q-learning algorithm will face the problem of dimensionality disaster. There are too many state action pairs and it is difficult to store in a table format. DQN fit the state-action value function through the deep neural network, and select the action with \(\varepsilon\)-greedy strategy

\[
a = \begin{cases} 
\arg\max_{a \in A} Q(s, a; \omega), \text{rand()} > \varepsilon \\
\text{random choose action, rand()} < \varepsilon 
\end{cases}
\]

Where, \(\varepsilon \in (0,1)\), \(\omega\) is neural network weights.

In order to reduce the correlation between samples, DQN introduces an experience replay mechanism. At each step, \(\{s_t, a_t, r_t, s_{t+1}\}\) is stored in the experience pool, and a certain number of samples are randomly selected during training to update the weight of the neural network, which improves the stability of training. If the same network is used for training, the training will be unstable due to the constant change of Q value, so DQN introduces two neural networks: target network and eval network.

At the beginning, the structure and parameters of the target network and the neural network are the same. The eval network is updated in real time with training, and the target network copies the parameters of the eval network every certain number of steps. The loss function during training is

\[
J(\omega) = E(y_i - Q(s_i, a_i; \omega))^2
\]

\[
y_i = r_i + \gamma \max_{a} Q(s_{i+1}, a, \omega^*)
\]

Where, \(\omega\) is the weight of the eval network; \(\omega^*\) is the weight of the target network.

4. Proposed method

4.1. State and action space

The decision variables of the microgrid model include \(P_{\text{bat}}(t), P_g(t), P_b(t), P_{\text{buy}}(t)\) and \(P_{\text{sell}}(t)\). If all of the above decision variables are used as actions output by the neural network, there will be a problem that the number of discrete action space combinations is too large and the training is too difficult. From the objective function and constraint conditions of the microgrid optimization model, it can be known that the optimization problem of the microgrid model is essentially to find the optimal output arrangement of the battery during the \(T\) period. Therefore, only \(P_{\text{bat}}(t)\) is used as the output action of the neural network \(A = \{P_{\text{bat}}(t)\}\). The remaining decision variables in period \(t\) can be solved by nonlinear programming. The state space is

\[
s_t = \{P_g(t), P_b(t), P_{\text{re}(t)}, P_{\text{sell}(t)}, P_{\text{buy}(t)}, P_{\text{t-end}}, t, P_g(t - \Delta t), \text{SOC}(t)\}
\]

Where, \(P_g(t - \Delta t)\) is the output of diesel generators in the last period.

4.2. Reward function

The economic operating cost in period \(t\) is solved in two stages. In the first stage, the neural network outputs the action \(P_{\text{bat}}(t)\); according to the action, calculates the state of charge of the battery for the next period \(\text{SOC}(t + \Delta t)\), and then calculates the economic operation cost of the battery and the penalty for exceeding the limit of \(\text{SOC}\).
\[ C_{t+1} = C_{bat}(P_{bat}(t)) \Delta t + C_{ro}(P_{bat}(t)) \Delta t \]  \tag{29} 
\[ C_{ro}(P_{bat}(t)) = r_{o} \cdot \max(SOC(t + \Delta t) - SOC_{max}, 0) + r_{o} \cdot \max(SOC_{min} - SOC(t + \Delta t), 0) \]  \tag{30} 

Where, \( C_{bat}(P_{bat}(t)) \Delta t \) is the economic operating cost of battery in period \( t \); \( C_{ro}(P_{bat}(t)) \Delta t \) is the penalty for exceeding the limit of SOC; \( r_{o} \) is penalty coefficient.

In the second stage, \( P_{bat}(t) \) is substituted as a known quantity into the objective function in period \( t \). Use nonlinear programming to solve the single-period objective function to obtain the remaining decision variables: \( P_{A}(t) \), \( P_{E}(t) \), \( P_{bat}(t) \) and \( P_{sell}(t) \). Then calculate the remaining economic operating costs. The deep reinforcement learning algorithm is the algorithm that solves the largest cumulative return, but the microgrid optimization is a minimization problem, so the immediate return is set to \( r_{t} = (C_{bat}(P_{bat}(t)) \Delta t) \). By dividing the calculation of rewards into two stages, the dimensionality of neural network output actions and the number of penalty constraints can be reduced. At the same time, using nonlinear programming to solve the remaining decision variables also increases the accuracy of the algorithm.

4.3. Training process

The specific training process of the microgrid online optimization algorithm based on deep reinforcement learning is as follows:

1. Initialize the eval network and target network weight \( \omega' = \omega \)
2. Initialize the experience replay pool \( D \)
3. According to the forecast information (WT, PV and LD forecast curve), use Monte Carlo sampling to generate \( M \) sets of training curves
4. For episode = 1, M do
5. Get the initial state \( s_{t} \)
6. For \( t = 1, T \) do
7. According to the \( \epsilon \)-greedy strategy select an action \( a_{t} \)
8. Calculate \( SOC(t + 1), C_{bat}(P_{bat}(t)) \Delta t \) and \( C_{ro}(P_{bat}(t)) \Delta t \)
9. Solve the remaining decision variables using nonlinear programming and calculate \( r_{t} \)
10. Calculate \( s_{t+1} \) and store \( \{s_{t}, a_{t}, r_{t}, s_{t+1}\} \) transition in \( D \)
11. Sample random minibatch of transition \( \{s_{t}, a_{t}, r_{t}, s_{t+1}\} \) from \( D \)
12. Calculate the loss function according to formula (26) (27) and update weight \( \omega \)
13. Copy the eval network weight to the target network every certain step \( c \) \( \omega' = \omega \)
14. End for
15. End for

5. Results and discussion

The calculation example uses the micro-grid structure shown in Figure 1. Because the wind and solar energy generation are pollution-free and the operating cost is low, the wind and solar are used for power supply first, and the operating cost of wind and solar energy generation is ignored. The capacity configuration and operating cost coefficient of each micro power supply are shown in Table 1. The pollutant emission coefficients and treatment costs of diesel generators and fuel cells are shown in Table 2. The parameters of the fuel cell are shown in Table 3 [18]. Electricity purchase and sale prices are shown in Table 4 [17].

| Micro Power Supply | Pmin (kW) | Pmax (kW) | \( K_{cost} \) ($/kW) |
|--------------------|----------|----------|---------------------|
| FC                 | 0        | 40       | 0.0293              |
| DG                 | 0        | 30       | 0.088               |
| BS                 | -15      | 15       | 0.05                |
| Grid               | -30      | 30       | 0.10                |
Table 2. Pollutant treatment cost and emission coefficient

| type  | $C_i$/kW | $\gamma$/g/kW | DG | FC |
|-------|----------|---------------|----|----|
| CO$_2$ | 0.0123   | 650           | 490 |    |
| SO$_2$ | 1.25     | 0.21          | 0.0028 |    |
| NO$_x$ | 2.5      | 9.9           |    | 0.0098 |

Table 3. Fuel cell parameters

| $C_{fuel}$/m$^3$ | $\eta$ | $LHV_f$/kW/m$^3$ |
|-----------------|--------|------------------|
| FC              | 4.375  | 80%              |
|                 |        | 9.7              |

Table 4. Purchasing and selling prices

| Price($/kWh) | Valley time | Peak time | Normal time |
|--------------|-------------|-----------|-------------|
| Buy          | 0.43        | 1.21      | 0.69        |
| Sell         | 0.27        | 1.02      | 0.50        |

Fuel cost factor of diesel generator $a_{DG}, b_{DG}$ is 0.0168, 0.168 and 0.3, respectively; initial state of charge of the battery SOC(0) is 0.3; SOC$_{min}$ and SOC$_{max}$ is 0.2 and 0.9, respectively; $\eta_c, \eta_d$ is 0.85 and 0.95, respectively. The day-ahead forecast curve of wind, photovoltaic and load power are shown in Figure 4. Two neural networks with the same structure are used; the number of neurons in the hidden layer of each neural network is 50; the number of hidden layers is 4, and the activation functions are all relu. The attenuation coefficient $\gamma$ is 0.96; the initial value of the gradient update learning rate $\alpha$ is 0.005; the learning rate is adaptively attenuated; the number of iterations is 30000; the size of the experience replay pool D is 9600, and the number of randomly selected samples m is 128. The power of the battery is discrete according to its maximum and minimum power as $[-15, -14, -12, -10, -8, -6, -4, -2, 0, 2, 4, 6, 8, 10, 12, 14, 15]$; the number of output neurons of the neural network is 17.

Assuming that the forecast errors of WT, PV and LD meet a normal distribution, the mean value is the forecast value of each time period, and the standard deviation is 10%, 10%, and 5%, respectively. According to the forecast errors, randomly generate 30,000 training curves through Monte Carlo sampling to train the neural network of DQN. The optimizer selects Adam; the initial value of $\epsilon$ is 0.9; the minimum value is 0.1, and the target network weights are updated every 5 steps.

Figure 4. Cures of PV, WT and Load power

The average cumulative return curve obtained from training is shown in Figure 5. It can be seen from the curve that as the number of training increases, the average cumulative return gradually

Figure 5. Cumulative return average curve
becomes flat, and the system has stabilized. Save the weights and test the optimization performance of the algorithm. In order to verify the adaptation of the algorithm to the randomness of wind, photovoltaic and load power, the following scenarios are set for testing.

Scene A. Optimize the day-ahead forecast curve and compare it with the optimization result of GAMS. The battery optimization results are shown in Figure 6. DQN total cost is 164.59$; GAMS total cost is 162.45$.

Scene B. When wind, photovoltaic and load power all fluctuate, optimize the fluctuation curve and compared it with the day-ahead optimization result of GAMS. Wind, photovoltaic and load power comparison curve before and after fluctuation are shown in Figure 7. The battery optimization results are shown in Figure 8. DQN total cost is 167.96$; GAMS total cost is 169.87$.

![Figure 6. Change curve of SOC of scene A](image1)

![Figure 7. WT PV and Load power comparison curve](image2)

![Figure 8. Change curve of SOC of scene B](image3)

It can be seen from the test results of two scenarios that when the neural network is trained, the neural network can output corresponding actions according to different states. In scenario A, because the discretization of action and the training curve contain randomness, the result of the DQN is slightly larger than the GAMS. From the optimization results of scenarios B, when the wind, photovoltaic and load power fluctuate, DQN algorithm can adaptively adjust the battery action, and the optimization result is better than the day-ahead optimization result of GAMS. The results prove the feasibility of online optimization of this algorithm.
6. Conclusion

This paper introduces artificial intelligence into the optimization of microgrids; taking the minimization of the economic operation cost as the optimization goal, propose a microgrid online optimization algorithm based on deep reinforcement learning. The neural network is used to approximate the state-action value function, and Monte Carlo sampling is used to generate multiple sets of training curves to train the neural network, so that the neural network can adapt to the randomness of wind, solar and load. First, the trained neural network decides the optimal battery action in real time, and then solves the remaining decision variables through nonlinear programming, so as to realize the real-time online optimization of the microgrid. When the wind, photovoltaic and load power fluctuate, DQN algorithm can adaptively adjust the battery action, and the optimization result is better than the day-ahead optimization result of GAMS. It proves the feasibility of online optimization of this algorithm. It has certain practical significance for the online optimized dispatch of microgrid.

References

[1] Igualada L, Corchero C, Cruz-Zambrano M and Heredia F 2014 Optimal energy management for a residential microgrid including a vehicle-to-grid system IEEE Trans. Smart Grid 5(4) 2163–72
[2] Anglani N, Oriti G and Colombini M 2017 Optimized energy management system to reduce fuel consumption in remote military microgrids IEEE Trans. Ind. Appl. 53(6) 5777–85
[3] Tiwari N and Srivastava L 2016 Generation scheduling and micro-grid energy management using differential evolution algorithm International conference on Circuit IEEE pp 1–7
[4] Sun F J, Zhang M J, Zhai X J, Cheng L and Zeng X H 2016 Multi-objective optimization operation based on ant colony algorithm for micro-grid Heilongjiang Electric Power 38(05) 377-379+411
[5] Liu R R, Zhang K S, Zhang G and Liu T 2017 Micro grid optimal operation based on adaptive artificial fish algorithm Power System and Clean Energy 33(04) 71-6
[6] Feng L, Cai Z X, Wang Y and Liu P 2017 Strategy for tie line power fluctuation suppressing of load-energy storage coordinated microgrid considering energy-storage characteristic of load Automation of Electric Power Systems 41(17) 22-8
[7] Liu Y X, Guo L and Wang C S 2018 Economic dispatch of microgrid based on two stage robust optimization Proceedings of the CSEE 38(14) 4013-4022+4307
[8] Zhu J Y, Liu Y, Xu L X, Jiang Z Z and Ma C X 2019 Robust day-ahead economic dispatch of microgrid with combined heat and power system considering wind power accommodation Automation of Electric Power Systems 43(04) 40-51
[9] Yan L F, Liu J, Shi M X, Chen X and Wen J Y 2019 Adaptive power allocation strategy based on fuzzy logic algorithm for hybrid energy storage system in DC microgrid Proceedings of the CSEE 39(09) 2658-70
[10] Mi Y, Li Z Q, Wu Y W, Liu H Y, Song G X and Su X J 2018 Bi-Layer optimal dispatch of grid-connected microgrid based on two-stage demand response Power System Technology 42(06) 1899-906
[11] Sheng Y J and Yang B 2020 Optimal dispatch of microgrid with demand response and an improved bat algorithm Journal of Huazhong University of Science and Technology 48(02) 120-5
[12] Dong L, Chen H, Pu T J, Chen N S and Wang X H 2016 Multi-time scale dynamic optimal dispatch in active distribution network based on model predictive control Proceedings of the CSEE 36(17) 4609-17
[13] Ma W J, Wang J H, Gupta V and Chen C 2018 Distributed energy management for networked microgrids using online ADMM with regret IEEE Trans. Smart Grid 9(2) 847-56.
[14] Shuai H, Fang J K, Ai X M, Tang Y F, Wu J Y and He H B 2018 Stochastic optimization of economic dispatch for microgrid based on approximate dynamic programming IEEE Trans. Smart Grid 10(3) 2440-52
[15] Leo R, Milton R S and Sibi S 2014 Reinforcement learning for optimal energy management of a solar microgrid Global Humanitarian Technology Conference-south Asia Satellite IEEE pp 181-6
[16] Zhang Z D, Qiu C M, Zhang D X, Xu S W and He X 2019 A coordinated control method for hybrid energy storage system in microgrid based on deep reinforcement learning Power System Technology 43(06) 1914–21
[17] Mao X M, Chen S, Wu J K and Guo Z Z 2015 Optimal dispatching of microgrid containing battery under time-of-use price mechanism Power System Technology 39(05) 1192-97
[18] Chen H D. 2019 Research on optimal scheduling of multi-micro network system Xi’an University of Technology