2020 19th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)

Dueling Double Q-learning based Real-time Energy Dispatch in Grid-connected Microgrids

Yuankai Shu, Wenzheng Bi, Wei Dong, Qiang Yang*
College of Electrical Engineering, Zhejiang University
Hangzhou, 310027 China
qyang@zju.edu.cn

***Abstract*—This paper presents a real-time scheduling strategy based on deep reinforcement learning (DRL) algorithm aiming to realize economic dispatch of microgrid energy storage considering operational uncertainties. Making the scheduling decision of microgrid is a non-trivial task due to the random fluctuations of new energy power generation systems and loads. In order to solve this problem, the double deep Q-learning algorithm with the dueling structure is investigated to ensure the reliability of the microgrid while considering the real-time electricity prices. The agent is tested on the actual data and the results show that the proposed algorithm can get small operation cost of the microgrid in complex situations.**

***Keywords*-Microgrid; Energy storage system (ESS); Dueling DQN; Markov decision process (MDP);**

I. Introduction

Microgrids incorporated with renewable energy units and energy storage system (ESS) devices are viewed to play important roles in the smart grid [1]. On the basis of balancing supply and demand, it provides a systematic way to take full advantage of renewable energy. Unfortunately, the renewable energy power generation system has the characteristics of randomness and intermittent while the microgrid is small in scale and the smoothing effect of load aggregation is weak, resulting in large load fluctuations in the microgrid. ESS is regarded as an important unit to smooth the influence of random fluctuations on the source and load sides in MG real-time dispatch. As energy storage system needs to be optimised across multiple-time steps, considering the changes in real-time electricity prices, its economic dispatch is complex. To ensure the economic and reliable operation of the microgrid with uncertainties, the ESS's real-time scheduling strategy needs to be further exploited.

For the energy management and optimization control problems in the microgrid, there have been many related studies. In [2], the Lyapunov algorithm was introduced into the microgrid energy management, and real-time scheduling of the microgrid system was realized through rigorous mathematical reasoning. This kind of method relies on a clear target expression, and due to the random fluctuation of renewable energy and load, the optimization decision-making scenario of the microgrid is difficult to abstract into a clear mathematical expression. The study in [3] used mixed integer linear programming to solve the stochastic scheduling problem of microgrid consisting of renewable energy and battery system. The authors in [4] presented a smart energy management system (SEMS) and optimise the operation of the microgrid through the genetic algorithm applied to ESS. The heuristic algorithm can obtain a local optimal solution with a certain probability, which is helpful to solve the problem of large data scale and complicated scene.

With the rise of artificial intelligence in recent years, research on optimization and control of reinforcement learning (RL) in power systems are currently being conducted [5]. Reinforcement learning algorithm is a model-free method for solving sequential decisions. The combination of deep learning (DL) with RL has led to a new field of research, called DRL, which use DL to enable RL to deal with larger dimensional state problems. This paper applies Dueling DQN algorithm to the economic dispatch of energy storage in the microgrid. The microgrid energy storage scheduling problem is described as MDP, the state of each unit of the microgrid is used as the state space, and the optimal strategy is obtained through constant interaction between the agent and the environment. Finally, the performance of the algorithm of this paper has been verified on the actual data.

II. Microgrid System Model

The microgrid (Fig. 1) considered in this work consists of wind-turbine generators, a battery storage system, loads and this paper assumes the MG runs in a grid-connected mode and can buy electricity from the real-time electricity market, when there is insufficient energy in the microgrid

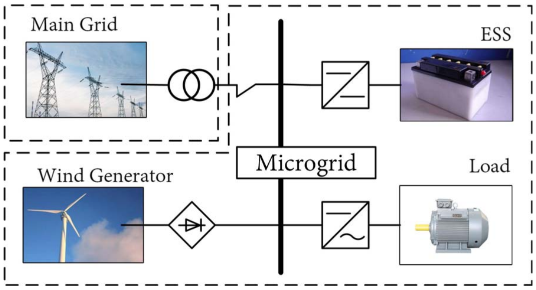

Figure 1: Microgrid system structure.

A. ESS Model

In this paper, a dynamic model is used to represent the ESS model, in which the energy storage battery capacity is represented by s_t^B, and the battery's operating mode includes three states: charging, discharging, and idle. We

2473-3636/20/$31.00 ©2020 IEEE
DOI 10.1109/DCABES50732.2020.00020

denote p_t^{B-dis} and p_t^{B-ch} as the discharging or charging power of the considered ESS. The power p_t^B and the energy level s_t^B of the ESS are constrained by

$$s_{\min} \leq s_t^B \leq s_{\max} \tag{1}$$

$$0 \leq p_t^B \leq p_{\max}^B \tag{2}$$

$$s_{t+1}^B = s_t^B + \eta p_t^{B-dis} \Delta t + \xi p_t^{B-ch} \Delta t \tag{3}$$

where p_{max}^B represents the maximum charging or discharging power; s_{max} and s_{min} are the maximum and minimum capacity of the ESS, respectively; η and ξ are the discharging and charging efficiency. Within the unit time, the battery is only charged or discharged.

B. Energy Dispatch Model

In this paper, the MG system can purchase electricity from the main grid in a real-time electricity market or transmit energy to main grid. Under the condition of satisfying power balance and equipment constraints, the microgrid energy storage scheduling model studied in this paper is to minimize the operation cost of microgrid.

The power balance of the microgrid system can be expressed as

$$\Delta p = p_t^W + p_t^B - p_t^{load} \tag{4}$$

Here, $\Delta p > 0$ represents that the microgrid transmits energy to main grid and $\Delta p < 0$ represents that the microgrid purchases electricity from the main utility grid. Let R_t denote the real-time electricity price in the microgrid, then the operating cost of the microgrid at time step t is computed by

$$C_t^{MG} = \Delta p \cdot R_t \cdot \Delta t \tag{5}$$

C. Modeling ESS scheduling as a MDP

This paper applies a DRL technology to tackle the sequence decision problem involving the operational planning of a battery in the microgrid. In a reinforcement learning context, the fundamental elements for the MDP model are defined as follows.

1) State Space S_t: For the MG system above, we define its state S_t at time step t by:

$$S_t = [p_t^W, p_t^{load}, E_t^{SOC}, R_t, t] \tag{6}$$

which consist of the power of wind power p_t^W at time t, the state of the battery soc E_t^{SOC} at t, the active load power demand p_t^L at t, electricity purchase price of microgrid R_t at time t, time step t of the day.

2) Action Space A_t: At each scheduling time point, the reinforcement learning agent takes discrete actions, then the action space is set as follows :

$$A_t = [0, 1, 2] \tag{7}$$

where $A_t = 0$ represents the battery does not operate, $A_t = 1$ represents battery discharge, and $A_t = 2$ represents battery charge.

3) Reward: In the interaction between the agent and the environment, you will get an immediate reward. From the perspective of energy scheduling of the microgrid, the reward function is set to the operating cost of the microgrid C_t^{MG} at each scheduling moment.

III. DRL Algorithm Framework

Reinforcement learning is based on the Markov decision process (MDP), that is, the state of the system at the next moment is only related to the state of the current moment. The energy scheduling decision problem of microgrid can be regarded as an MDP model. RL can support sequential decision making under uncertainty without prior knowledge.

A. DQN Algorithm

DQN is an algorithm that combines neural networks and Q-learning [6], which takes states or states and actions as input to the neural network, uses the neural network to calculate all action values, and selects the maximum value as the output. Under the strategy π, the agent performs the action a at the state s, and transits to the next state s' with the probability P, while receiving feedback r from the environment. We define $Q(s,a)$ as an state-action value. The current optimal strategy π for the agent to perform actions is to choose an action that maximizes the objective function $r + \gamma Q^*(s', a')$, namely:

$$Q^*(s,a) = E_{s'}[r + \gamma \max_{a'} Q^*(s', a') | s, a] \tag{8}$$

In DQN, the deep neural network with weighted θ is used as a function approximator to estimate the state-action value function, namely: $Q(s,a;\theta) \approx Q^*(s,a)$. Then the Q network can be trained through the loss function

$$L_i(\theta_i) = E_{s,a}[(y_i - Q(s,a;\theta_i))^2] \tag{9}$$

$$\begin{aligned} \nabla_{\theta_i} L_i(\theta_i) = E_{s,a;s'}[(r + \gamma \max_{a'} Q(s', a'; \theta i - 1) \\ -Q(s,a;\theta)) \nabla_{\theta_i} Q(s,a;\theta i)] \end{aligned} \tag{10}$$

In (9), $y_i = [r + \gamma \max_{a_{i+1}} Q(s_{i+1}, a_{i+1}; \theta_{i-1}) | s, a]$ is the goal of iteration i.

B. DDQN With Dueling Network Architectures

DQN includes the step of selecting the maximum estimate when estimating the action value, so it may lead to overestimation during the learning process. In order to enhance the generalization ability of the agent, the structure of double DQN (DDQN) is adopted for the reinforcement learning algorithm of the microgrid. Since the model learning is based on measured data, in order to effectively use the data, this paper uses an improved DQN algorithm based DDQN, that is, using the Dueling DDQN algorithm, which allows the agent to learn the value of state and action separately.

1) Double Q-learning: There are two neural networks in DDQN, one is the target network Q_t and the other is the main network Q_m. The working principle of DDQN is shown in Fig. 2.

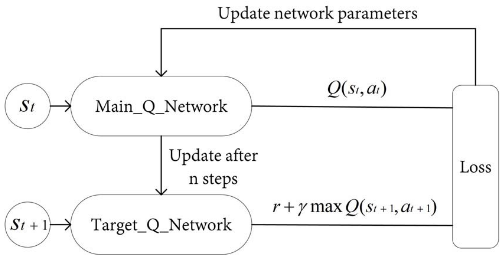

Figure 2: The training process of DDQN

In each update, The main network obtains the action a with the highest value in the next state, and obtains the Q value of the action a from the target network. Then the goal of learning in ddqn is

$$y_i = r + \gamma Q_t(s_{t+1}, \arg maxQ_m(s_{t+1}, a; \theta_m); \theta_t) \quad (11)$$

2) Dueling DQN: The network structure of Dueling DQN [7] is shown below:

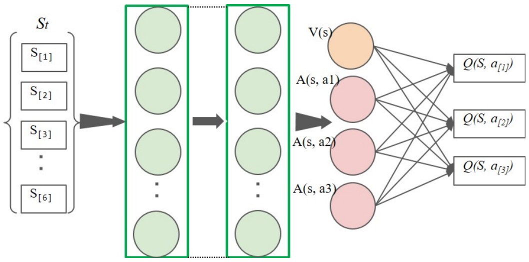

Figure 3: The structure of Dueling DQN

The Q value of each action in Dueling DQN is determined by the following formula

$$Q(s, a; \theta, \alpha, \beta) = V(s; \theta, \beta) + A(s, a; \theta, \alpha) \quad (12)$$

In the formula, the value function $V(s; \theta, \beta)$ indicates the degree of the state. The advantage function $A(s, a; \theta, \alpha)$ indicates the degree of a certain action relative to other actions in this state. The sum of $V(s; \theta, \beta)$ and $A(s, a; \theta, \alpha)$ represents the value of the certain action determined in this state. In the constructed network structure, different actions have different bias, and the value function is a scalar quantity, adding $V(s; \theta, \beta)$ and $A(s, a; \theta, \alpha)$ directly will result in a poor learning effect. In order to improve this method, the average value of the advantage function is usually used for calculation

$$\begin{aligned} Q(s, a; \theta, \alpha) = V(s; \theta, \beta) + [A(s, a; \theta, \alpha) \\ -\frac{1}{|A_t|} \sum_{a'} A(s, a'; \theta, \alpha)] \end{aligned} \quad (13)$$

IV. Case Studies

This paper uses the above microgrid structure as the analysis object, and the user side contains an ESS unit, wind power generation device, and load unit. For all simulations, a day is divided into 24 time slots, the baseline load profile is adopted from AEMO, and the wind generation power data is obtained by scaling down the realistic wind generation statistics, ranging from 0 to 20 kW. Part of the data is shown in Fig. 4. This paper also considers the interaction between the microgrid and the main grid.

It can be seen from the figure that the overall fluctuation of the load curve data is not large, but the typical fluctuation characteristics of each day can still be seen from the curve. The wind power curve obviously has daily differences and day-night gaps, and its fluctuation is very obvious.

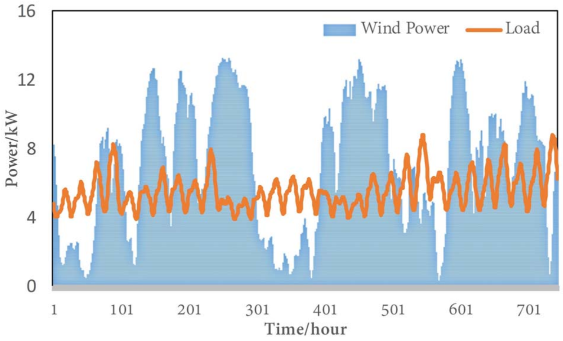

Figure 4: Wind power generation curve and load curve about 4 weeks

In this paper's reinforcement learning algorithm, the initial value of the learning rate is set to 0.0001, and the initial value of the discount factor γ is set to 0.99. The capacity of the battery is set to 100 kW/h.

The following discusses the adaptability of the reinforcement learning strategy proposed above in the microgrid. In actual microgrid dispatching, we expect to achieve the lowest operating costs. In the following we will mainly examine the application of the strategy used in the microgrid's one-day operation. Using prediction data with noises as training set and let the agent traverse all the data in the training set and update the agent's neural network according to the reward. The error between the predicted data and the actual data satisfies the Gaussian distribution.

The change of the agent's reward during the training process is shown in Fig. 5. It can be seen from the curve in the figure that for training data of one day, a stable state can be reached.

Then test the trained model with actual data, the Fig. 6 and Fig. 7 shows the test results. Fig. 6 shows the changes of wind power generation, load and battery soc. According to the curve, it can be seen that the trained agent can respond in real time according to the power change of the microgrid. At the same time, we also considered the real-time electricity price. The power gap between wind power and load demand ,the electricity price and the rewards of every time slot are shown in Figure 7. In order to minimize operating costs, the agent will consider store part of the energy when the renewable energy generation is larger, and consider releasing the stored energy to ensure the load demand when the load is larger.

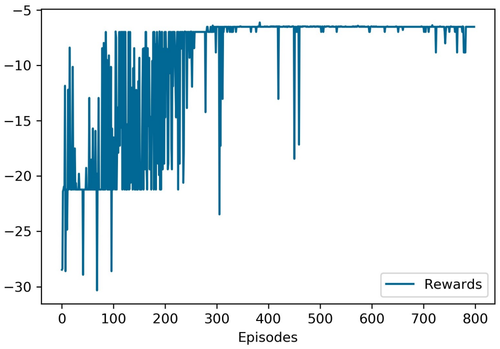

Figure 5: Single-day reward and iteration number of energy storage scheduling strategy

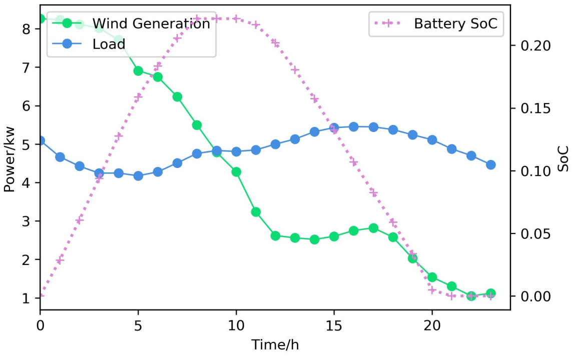

Figure 6: The change of battery soc under the actual fluctuation of wind power and load power

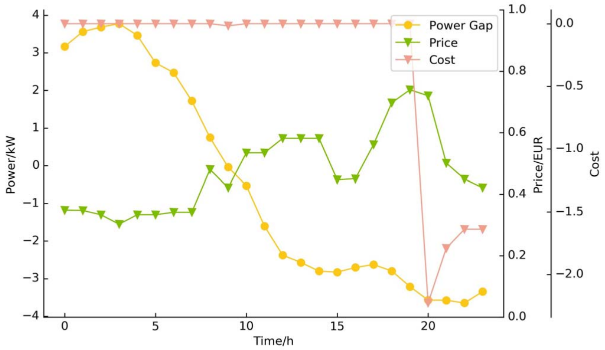

Figure 7: Changes in real-time electricity prices, differences between wind power generation and load demand and the rewards of every time slot in a day

It can be seen from the above results that in the case of the actual data, the ESS applying reinforcement learning strategy will select charge and discharge actions to achieve small operating cost when faces with multiple uncertainties.

V. Conclusion

This paper studies the energy storage scheduling problem of microgrid with wind power generation, considering the impact of electricity price. Specifically, the real-time scheduling of MG is modeled as an MDP and the objective is to find an optimal scheduling strategy to minimize the daily operating cost of the MG. A double DQN method with dueling network structure is developed to solve this problem. In the proposed approach, the dueling DQN are used to approximate the value of the state and the advantage function of the action in the state. The ESS selects the charging and discharging action based on the observation of the microgrid. We analysis the method with actual data, and the results show that the ESS applying this strategy can choose the appropriate action under the influence of multiple uncertain factors. In future research, we will also consider the impact of load demand response and consider the coordination between multiple agents.

Acknowledgment

This work is supported in part by the Fundamental Research Funds for the Central Universities (Zhejiang University NGICS Platform), the Industrial Internet innovation and development project (2019) - Project of Time Sensitive Network (TSN) Technology based Programmable Logical Controller (PLC) (TC190A449), and the National Natural Science Foundation of China (51777183).

References

[1] A. Chaouachi, R. M. Kamel, R. Andoulsi, and K. Nagasaka, “Multiobjective intelligent energy management for a microgrid,” *IEEE transactions on Industrial Electronics*, vol. 60, no. 4, pp. 1688–1699, 2012.

[2] W. Shi, N. Li, C.-C. Chu, and R. Gadh, “Real-time energy management in microgrids,” *IEEE Transactions on Smart Grid*, vol. 8, no. 1, pp. 228–238, 2015.

[3] M. Zachar and P. Daoutidis, “Microgrid/macrogrid energy exchange: A novel market structure and stochastic scheduling,” *IEEE Transactions on Smart Grid*, vol. 8, no. 1, pp. 178–189, 2016.

[4] C. Chen, S. Duan, T. Cai, B. Liu, and G. Hu, “Smart energy management system for optimal microgrid economic operation,” *IET renewable power generation*, vol. 5, no. 3, pp. 258–267, 2011.

[5] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, “Playing atari with deep reinforcement learning,” *arXiv preprint arXiv:1312.5602*, 2013.

[6] D. Zhang, X. Han, and C. Deng, “Review on the research and practice of deep learning and reinforcement learning in smart grids,” *CSEE Journal of Power and Energy Systems*, vol. 4, no. 3, pp. 362–370, 2018.

[7] Z. Wang, T. Schaul, M. Hessel, H. Van Hasselt, M. Lanctot, and N. De Freitas, “Dueling network architectures for deep reinforcement learning,” *arXiv preprint arXiv:1511.06581*, 2015.