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

Real-time Energy Management of Microgrid Using Reinforcement Learning

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

***Abstract*—Driven by the development and application of smart grid and renewable energy sources (RES) generation technologies, microgrid (MG) plays an important role in environmental protection and optimization of the grid structure by integrating local loads and distributed energy. However, the stochastic and intermittent nature of RES have caused difficulties in the economic energy dispatching of MG. Inspired by reinforcement learning (RL) algorithms, this paper proposes a novel learning-based control MG scheduling strategy. Unlike traditional model-based methods that require predictors to estimate stochastic variables with uncertainties, the proposed solution does not require an explicit model. The proposed method is simulated in the environment composed of realistic data, and the effectiveness of the method is explained and verified.**

***Keywords*-control policy; energy management; microgrid; deep reinforcement learning;**

I. Introduction

Increasingly environmental concerns and flexible trading mechanism have brought new development opportunities and challenges to the operation mode and dispatching strategy of power system. RES as a new energy form to solve energy crisis and environmental problems, has spawned the model of MG which consists of distributed energy, energy storage, and load. The balance between the generating side and the consuming side has always been one of the main problems in power grid energy management. However, due to the inherent stochastic and intermittent nature of various RES and the uncertainty of users' electricity consumption behaviors, the dispatching of MGs has greater stochastic than that of traditional power grids, which is the reason why MGs are usually equipped with an energy storage system (ESS). Meanwhile, the excessive imbalance between supply and demand will consume more reserve and ancillary equipment to ensure the normal operation of the MG, which will significantly reduce its economy[1]. Therefore, it is urgent to carry out an effective energy management strategy to make the economic dispatching.

ESS can not only generate and absorb energy to buffer the flow of energy, but also optimize the power quality and improve the stability of micro-grid, which is the key component of energy dispatching. Traditional control strategies require explicit modeling of microgrids and the optimal strategy is obtained by a solver. For instance, the model predictive control (MPC) has strong robustness due to its predictive and feedback correction ability, and can be effectively applied to the control of complex industrial processes. In [2], an offline algorithm was designed by combining offline solution with sequential optimization based on sliding window. In [3], the strategy of combining optimal generation scheduling with MPC to achieve long-term and short-term optimal planning was proposed. Although the model-based method has advantages and successful applications in the above work, it relies heavily on domain expert knowledge to build MG models and parameters. In addition, complex model with optimization process also bring computational burden in real-time fashion [4]. Motivated by the recent development of reinforcement learning (RL), this paper proposes a model-free strategy for real-time energy dispatching of ESS in MG. It does not need accurate modeling of MG, and can be used for real-time dispatching once the training is completed, which has strong portability [5].

II. Problem Formulation

A. Microgrid Model

1) MG Architecture: Fig. 1 illustrates the microgrid architecture we discussed which consists of a photovoltaic (PV) system, a wind turbine (WT), a battery pack as the energy storage device, a group of local demands, and a main grid connected by point of common coupling (PCC). It should be noted that the purpose of using this microgrid model is to briefly illustrate the proposed control strategy, which can be applied to more complex circumstances due to the portability of data-driven methods.

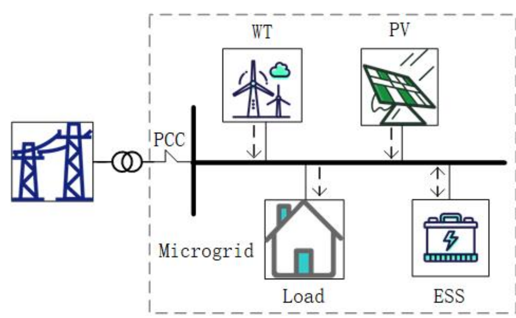

Figure 1. MG architecture includes RES generators, ESS and local load, and is connected to the main grid via PCC.

2) Battery Model: For ESS, We denote the charging or discharging power of the battery at time slot t by $P_b(t)$. Positive $P_b(t)$ means battery charging and negative $P_b(t)$ means discharge. Therefore, the dynamic characteristics of ESS can be expressed by:

$$E(t+1) = \begin{cases} E(t) + \eta_{charge} P_b(t)\Delta_t, P_b(t) > 0 \\ E(t) + \dfrac{P_b(t)}{\eta_{discharge}}\Delta_t, P_b(t) < 0 \end{cases} \quad (1)$$

2473-3636/20/$31.00 ©2020 IEEE
DOI 10.1109/DCABES50732.2020.00019

Where, $E(t)$ represents the energy in the ESS at time t, Δ_t represents the time interval of each scheduling, and η_{charge} and $\eta_{discharge}$ represent the charging and discharging efficiency respectively.

Moreover, the ESS operating parameters, i.e. power and state of charge (SoC), should not exceed the set maximum value, and the constraints are given by:

$$\begin{cases} SoC_{min} \leq SoC(t) \leq SoC_{max} \\ |P_b(t)| \leq P_{max}(t) \end{cases} \tag{2}$$

3) RES model: We assume that there are multiple distributed generators composed of viable RES in the MG. In the above MG structure, PV and WT are used as local energy supply. For each $RES_i, i \in \mathcal{D}$, the generation pattern is uncertain, and its generation power is bounded by:

$$P_{imin} \leq P_i(t) \leq P_{imax}, \forall i \in \mathcal{D} \tag{3}$$

The electricity generated by RES should be maximized to meet the local demand first. If there is excess power, it can be stored in the ESS for subsequent needs. Since the installation costs of distributed generators and auxiliary equipment are paid in a lump sum at the time of installation, and its operation and maintenance costs are relatively small, the generation costs of RES are ignored in this paper.

4) Main Grid Model: In our MG settings, users can pay to get electricity from the main grid if the electricity generated by RES and ESS cannot meet the local demand. The electricity price adopts real-time price (RTP), which means that the control strategy needs to charge and discharge ESS reasonably to maximize the use of RES and reduce the cost of purchasing electricity. At each time slot, the total cost of purchasing power from the main grid can be expressed as:

$$C(t) = \gamma(t) P_g(t) \Delta_t \tag{4}$$

Where, $P_g(t)$ and $\gamma(t)$ are respectively the the power exchange and RTP at time t.

B. Reinforcement Learning

This paper addressed the model-free method based on RL to solve the energy dispatching of ESS as the sequential decision-making problem. RL problems can be expressed as a Markov decision process (MDP), in which the agent learns continuously in interaction with the environment and finally gets the optimal strategy. At each time step, after getting the information passed by the environment, the agent selects the action value $a_t \in \mathcal{A}$ under the current state $s_t \in \mathcal{S}$ according to strategy π, while the environment gives the environment state s_{t+1} and reward value $r_{t+1} \in \mathcal{R}$ for the next time step according to the transition probability $p(s_{t+1}, r_{t+1} \mid s_t, a_t)$.

Solving an RL task means figuring out a strategy that will yield substantial benefits over the long term, which in the microgrid paradigm means minimizing the cost within the total dispatch time horizon. We use a value-based RL method, which uses a value function to evaluate the performance of strategy π. The value function is defined as the expected value of subsequent rewards with a discount factor:

$$Q(s,a) = \mathbf{E}_\pi \left[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s, a_t = a \right] \tag{5}$$

Where, the discount factor $\gamma \in [0,1]$, the closer its value is to 1, the more consideration will be given to future rewards. In other words, the agent becomes more far-sighted.

1) State Space: During the operation of the MG, the operator of the control center monitors the state of the MG in real time through the Supervisory Control and Data Acquisition (SCADA) system. External environmental information, such as RES generation, local load and RTP, is the main factor leading to the power imbalance of the microgrid, so they are the main basis for decision making. In addition, the SoC of the ESS and the hour of the day $\gamma(t)$ are also considered in the state space. The state space is defined as follows:

$$s_t = (P_{PV}(t), P_{WT}(t), P_L(t), SoC(t), \gamma(t), h(t)), s_t \in \mathcal{S} \tag{6}$$

where,$\forall t \in \{0, 1, \ldots, T-1\}$, T is the total scheduling time period. $P_{PV}(t)$ and $P_{WT}(t)$ are respectively PV and WT power generation, $P_L(t)$ is load power. $h(t) \in \{1, 2 \cdots, 24\}$ represents the time information of the scheduling.

2) Action Space: At each dispatch moment, the agent can make decision instructions to ESS based on the state information of the MG. There are three possible actions in the action space, namely idle, charging and discharging, which are expressed as:

$$\mathcal{A} = \{0, 1, 2\} \tag{7}$$

When the charging command is issued, the ESS will absorb the excess power generated by the RES generators as much as possible. When the local power supply is insufficient, the ESS will choose whether to discharge to make up for the lack of energy. The idle command allows the ESS to save the existing power for subsequent use.

3) Reward Function: In the RL paradigm, the rewards are transferred from the environment to the agent, which is used to convey the goals that the control strategy wants to achieve. At each time step, the agent receives the reward value given by the MG environment for buyiny electricity from the main grid:

$$r_t = -C(t) = -\gamma(t) P_g(t) \Delta_t \tag{8}$$

4) Deep Q Network: Common table-based reinforcement learning methods will suffer from "curse of dimensionality" when facing high-dimensional state and action spaces, which increases the difficulty of computing. Through the powerful feature extraction ability of deep learning (DL), artificial neural network (ANN) can be applied to nonlinear approximation of value function [6]. We use a multi-layer full connection layer neural network to fit the optimal value Q^*, with Q^* satisfying the Bellman optimality equation:

$$Q^*(s,a) = r + \gamma \max_{a' \in \mathcal{A}} Q^*(s', a') \tag{9}$$

Finally, the optimal strategy π^* is obtained:

$$\pi^*(s) = \arg\max_{\forall a \in \mathcal{A}} Q^*(s,a) \tag{10}$$

III. Case Study

Since the prediction of the information before the day is beyond the scope of this work, we assume that the predicted value of the external state will be obtained before the actual scheduling one day ahead. The prediction information is used to train the model every day, and the trained model can be used for real-time scheduling of the next day. For each model training, a day is divided into 48 time slots, and RTP are based on real historical data. Moreover, RES generation and user load demand are obtained from real historical data after scaling [7], as illustrated in Fig. 2. For ESS, the maximum capacity is set to 2.4MWh. SoC_{max} is 0.9 and SoC_{min} is 0.2. The maximum charge and discharge power is set to 400kW. The charging efficiency η_{charge} is 0.9, and the discharging efficiency $\eta_{discharge}$ is 0.89. We use an ANN with two hidden layers to approximate the Q value. The two hidden layers have 500 and 200 neurons respectively, and they both use the Relu function as the activation function.

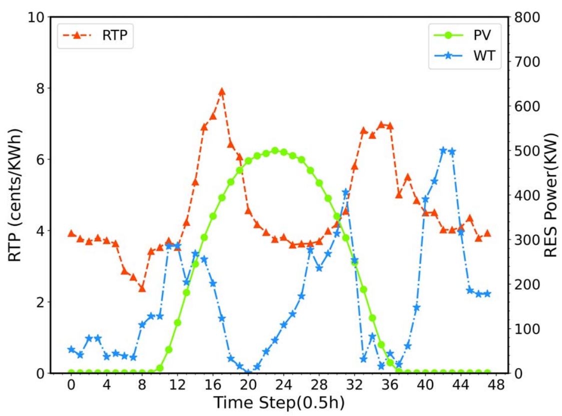

Figure 2. Adopted RTP and RES generation profile for real-time scheduling.

RL algorithms continuously train one day's forecast data as an episode. Initially, the environment is randomly explored and the experience gained is stored in the experience pool. Afterwards, we constantly update Q network through learning, and continuously reduce the exploration of random actions during the training process, and finally take the optimal action completely. We used different random number seeds for multiple training. After 800 episodes of training, the total reward converged, as shown in Fig. 3. For the first 200 episodes, all the action values are random for environment exploration. After 640 episodes, the selection of random actions is turned off, and the algorithm finally converges.

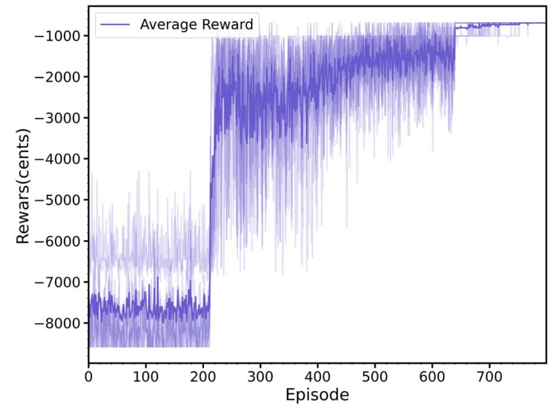

Figure 3. The change of total rewards after training under different random number seeds with training episodes.

We use the trained agent for real-time scheduling under real data. Fig. 4 shows scaling action value and the curve of ESS's SoC during this day and Fig. 5 shows the energy flow. The upper part of the figure shows the energy supply, which is composed of RES generations, ESS discharges and main grid power purchase. The lower part of the figure shows the dynamic changes of load and ESS charging with time.

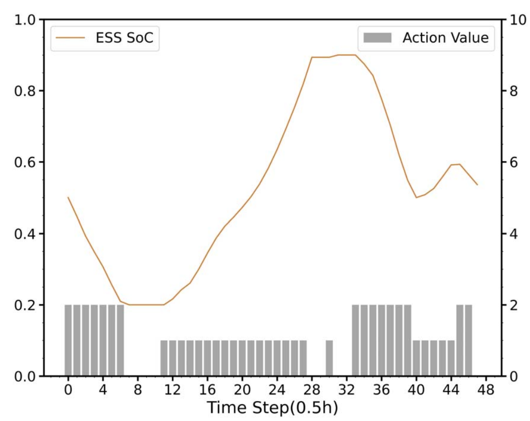

Figure 4. The action value given by the agent and ESS's SoC during the scheduling process.

Based on the analysis of the two figures, at the beginning of the day, the power generation of RES cannot meet the local load demand, so ESS discharges to make up for the insufficient power supply. However, the power initially stored in the ESS is still insufficient, and has reached the minimum SoC of the battery, so users need to purchase additional power. The reason why the electricity purchase behavior is concentrated in the 6-11 time step is because the RTP is low at this time. It can be seen from the figure that the amount of electricity generated by RES generators is more in the middle of this day. At this time, the control strategy chooses to charge the ESS until it reaches the maximum SoC for subsequent use of electricity. In the following time, the optimal strategy

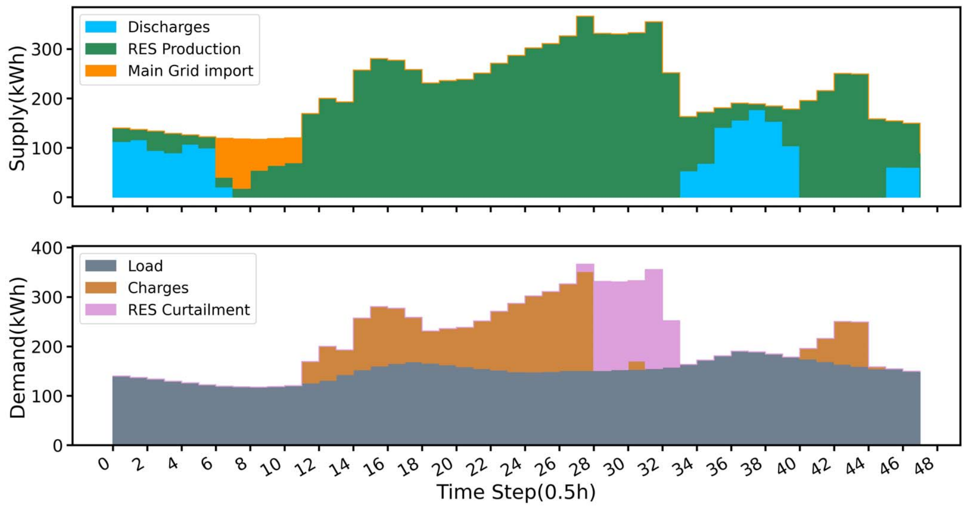

Figure 5. Distribution of energy supply and energy demand during the day's scheduling. The upper and lower parts of the figure represent the different proportions of supply and demand.

will be able to discharge the ESS when the user needs additional power, and charge when the local energy is abundant.

IV. Conclusion

This article proposes a learning-based MG energy real-time scheduling strategy to cope with the dilemma of traditional model-based control methods in the face of uncertainties in new energy generation. Under the reinforcement learning paradigm, by setting the power purchase cost of the MG as a reward for each step of scheduling, the agent can learn the optimal strategy to achieve the goal of minimizing the operating cost of the MG and maximizing the use of local RES in the long term. The simulation results verified on the real data of one day show that, under the model trained the previous day, the charging and discharging instructions issued by the agent can effectively improve the local energy utilization efficiency, give full play to the peak load shaving ability of the ESS, and reduce the additional power purchase cost of the MG operation. In future work, we will consider the use of effective prediction methods to predict the state of the MG and incorporate it into a reinforcement learning framework to enhance its robustness. And a more flexible electricity price transaction mechanism will be considered in the MG environment.

Acknowledgment

This work is supported by the National Key Research and Development Program of China (2018YFB0904800), and the National Natural Science Foundation of China (51777183)

References

[1] P. Zeng, H. Li, H. He, and S. Li, "Dynamic energy management of a microgrid using approximate dynamic programming and deep recurrent neural network learning," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 4435–4445, 2018.

[2] K. Rahbar, J. Xu, and R. Zhang, "Real-time energy storage management for renewable integration in microgrid: An off-line optimization approach," *IEEE Transactions on Smart Grid*, vol. 6, no. 1, pp. 124–134, 2014.

[3] M. Petrollese, L. Valverde, D. Cocco, G. Cau, and J. Guerra, "Real-time integration of optimal generation scheduling with mpc for the energy management of a renewable hydrogen-based microgrid," *Applied Energy*, vol. 166, pp. 96–106, 2016.

[4] S. Zhou, Z. Hu, W. Gu, M. Jiang, and X.-P. Zhang, "Artificial intelligence based smart energy community management: A reinforcement learning approach," *CSEE Journal of Power and Energy Systems*, vol. 5, no. 1, pp. 1–10, 2019.

[5] W. Liu, P. Zhuang, H. Liang, J. Peng, and Z. Huang, "Distributed economic dispatch in microgrids based on cooperative reinforcement learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 6, pp. 2192–2203, 2018.

[6] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[7] Q. Yang and X. Fang, "Demand response under real-time pricing for domestic households with renewable dgs and storage," *IET Generation, Transmission & Distribution*, vol. 11, no. 8, pp. 1910–1918, 2017.