Control method of power grid topology structure based on reinforcement learning

Yanhong Yang1,3, Wei Pei1, Wei Deng1, Hao Xiao1 and Hongjian Sun2

1 Institute of Electrical Engineering, Chinese Academy of Sciences, Beijing 100190, China;
2 Department of Engineering, Durham University, Durham DH13LE, UK
3 Email: yangyanhong@mail.iee.ac.cn

Abstract. With the rapid development of renewable energy and power electronics technology, uncertainty, complexity and data accumulation in power system continue to increase. Traditional methods often encounter bottlenecks in solving operational optimization decision-making and control problems. The development of artificial intelligence technology provides new methods and means for solving the problem of intelligent control of power grid topology. The deep reinforcement learning (DRL) method is used to learn from the historical experience data of power grid topology control that can improve the control and decision-making and methods of system operating performance, and solve the huge variables in traditional models. Through the method in this paper, the DRL method can effectively improve the real-time optimization and control ability of the grid topology.

1. Introduction

As the penetration rate of renewable energy continues to increase, the power grid is facing increasing challenges. The existing grid control methods are mainly oriented towards generator power control on the power generation side and demand response on the load side. As the carrier of electric energy transmission, the power network can flexibly change the topology by controlling the substation and line switchgear to achieve renewable energy consumption and system network losses are reduced, but the real-time and intelligent control of the grid topology is rarely considered in research and practice [1-2].

At present, the control of the grid topology is mainly limited by modeling methods and computing capabilities. The traditional power grid topology control modeling method is relatively simple, and all switching devices are set as optimized variables. The optimized calculation model contains very huge 0/1 variables which established in the traditional way, and it is difficult to solve based on mathematical optimization algorithms [3-4], and a lot of simplification of the model is required, which affects the calculation accuracy, and the real-time performance is also difficult to guarantee.

Game AI based on reinforcement learning has completely crushed top human players in games such as Go and StarCraft. Control algorithms based on reinforcement learning have been applied to robots, drones and other industrial equipment, and trading algorithms based on reinforcement learning have been deployed on financial platforms and obtain excess returns [5-6]. A smart manufacturing strategy for real industrial friction drilling is proposed to identify the most accurate machine-learning technique to process experimental datasets in paper [7]. A data-driven method based on neural network (NN) and Q-learning algorithm is developed for home energy management in paper [8]. And
deep neural network based model-free reinforcement learning is utilized to manage the energy in a multi-microgrid system in paper [9].

The study of learning to run a power network (L2RPN) has become a current hot spot. The French RTE Power Company and ChaLearn Company organized the first global open competition of "Grid Topology Adjustment and System Optimization Control Based on Reinforcement Learning" in June 2019, and a series of competitions will be continue hold in 2020, for example Sandbox L2RPN, WCCI L2RPN, NeurIPS L2RPN and so on[10-12].

The development of artificial intelligence technology provides new methods and means for solving the above-mentioned intelligent control problems of power grid topology. This paper proposes a modeling and decision-making method for intelligent control of power grid topology. The deep reinforcement learning method is used to learn from the historical experience data of power grid topology control that can improve the control and decision-making and methods of system operation performance. It solves the huge and difficult traditional model variables.

2. Framework of grid topology control based on DRL

This paper uses reinforcement learning to construct the control architecture of the power grid topology, and specifically uses the deep Q network to design the agent. The Q network of dueling DQN has two Q functions: state value function and action advantage function. Because the power grid topology control action space is huge, the application of dueling DQN can learn the important states of the agent, without knowing the influence of each action on each state, and can quickly identify the best action. The power grid topology control architecture based on DRL is shown in Figure 1.

![Figure 1. Power grid topology control architecture based on DRL.](image-url)

The output of the dueling DQN network is obtained by the linear combination of the state value function and the action advantage function. The specific calculation formula is as follows [13]:

\[
Q(s,a;\theta,\alpha,\beta) = f_v(s;\theta,\beta) + \left( f_a(s,a;\theta,\alpha) - \frac{1}{|F|} \sum_{a'} f_a(s,a';\theta,\alpha) \right)
\]  

(1)

In the formula, Q is the Q network output function value of dueling DQN; s is the system state; a is the system action; \( \theta \) represents the public network structure parameter; \( \alpha \) represents the parameter of
the action advantage function fully connected layer; \( \beta \) represents the state value function of the fully connected layer parameter; \( f_s \) is the state value function, which is only related to state; \( f_a \) is the action advantage function, which is related to both state and action.

In this paper, the action advantage function is designed as a single action advantage function minus the average value of all action advantage functions, which can ensure that the relative order of the advantage functions of each action remains unchanged and improve the stability of the algorithm.

3. Multi-energy flow optimization operation model of industrial park

Different from the traditional method of establishing a grid physical control model for grid control problems, this paper adopts an intelligent control method based on reinforcement learning, and this needs to design reasonable action variables, state variables and reward values according to the physical topology of the grid.

3.1. Action variable design

As shown in the Figure 2, the power grid topology includes units such as substations, generators, loads, and power transmission lines, which is the environment for using DRL control. The power grid topology control variables include two types, one is the switch control of power transmission lines, and the other is the control of substations.

![Figure 2: Power grid topology.](image)

Regarding the switch control of power transmission lines, each power transmission line has two states: on and off. If there are \( N \) power lines, the number of actions of all power transmission lines is:

\[
A_p = 2^N
\]

(2)

The other is the control of substations. As shown in the Figure 3, each substation may contain multiple bus bars, and the control structure is flexible and diverse. For a substation containing \( k \) bus bars and \( p \) transmission lines, there are \( k^p(p-1) \) possible topological structures; therefore, if the grid contains \( M \) substations, the number of actions of all substations (optional topology ) are:

\[
A_M = \sum_{m=1}^{M} k_m^{p_m-1}
\]

(3)

In the formula, \( m \) is the \( m \)-th substation, which is the number of bus bars in the substation \( m \); is the number of connected power lines in the substation \( m \).
The above-mentioned two types of grid topology control optional action spaces show an exponential growth with the increase of the grid scale. Table 1 gives several specific examples, which assumes that all substations contain only 2 bus bars. The actual situation is more complicated than this.

Table 1. Examples of grid topology control action space [14].

| Case  | Number of substations | Number of power lines | Number of optional topologies for substations | Number of optional actions for power lines |
|-------|-----------------------|-----------------------|---------------------------------------------|------------------------------------------|
| Case 1| 5                     | 8                     | 31,320                                       | 256                                      |
| Case 2| 14                    | 20                    | 1,397,519,564                                | 1,048,576                                |
| Case 3| 36                    | 59                    | 1.88e+21                                     | 5.76e+17                                 |
| Case 4| 118                   | 186                   | 3.88e+76                                     | 9.81e+55                                 |

As mentioned above, if all power transmission line control and all substation control are considered at the same time, the action space of the grid topology is very huge, and it is difficult to train and solve effectively. Considering that the actual grid is running at the previous moment and the next moment, the state is relatively continuous, no large-scale actions will occur. Here, when designing the action variables of the present invention, the number of power transmission lines and substations that can be operated is selected according to the size of the grid. The specific formula is as follows:

\[ x_N = \text{roundup}(N + 20) \]  (4)

\[ x_M = \text{roundup}(M + 20) \]  (5)

In the formula, \( x_N \) is the number of power transmission lines that can be selected; \( x_M \) is the number of substations that can be selected; roundup is rounding up.

Therefore, the control action variable space of the grid topology structure designed in this paper is:

\[ A = 2^{x_N} + \sum_{m=1}^{x_N} k_m^{p_m-1} \]  (6)

3.2. State variable design

Regarding the operating states and parameters of the power grid at time step \( t \), this paper selects variables that change with the operation of the system as state variables, this state variables as the input of the agent, and does not deal with constant parameters such as impedance in the power grid. The state variables are selected as follows:
1) Active power, reactive power and voltage value of the generator; 
2) Active power, reactive power and voltage value of the load; 
3) Active, reactive and voltage values at the head and end of the power transmission line; 
4) Thermal limit of power transmission line; 
5) The operating status of the power transmission line; 
6) The operating status of the substation.

3.3. Reward value design
The goal of power grid control is to maximize the available transmission capacity for a given continuous time. Therefore, the ratio of the sum of the available transmission capacity of all transmission lines to the maximum transmission capacity of the system is used as the reward value for normal operation of the system; if the system is overloaded and the load loses power, then the system stops running, and -1 is given as the reward value. The above is the single-step operation of the power grid system. Within a certain control period T, due to the coupling relationship between the front and back moments, the final goal of the power grid control period has the largest accumulated reward value for all time steps.

For one time section, the reward value R is the total available transmission capacity, which designed as the equation (7):

\[ R = \begin{cases} -1, & \text{stop} \\ \frac{1}{N} \sum_{n=1}^{N} \max(0, (1 - \frac{p_n}{p_{n\text{max}}})^2), & \text{running} \end{cases} \]  

(7)

In the formula, \( p_n \) is the real-time power of line n, and \( p_{n\text{max}} \) is the maximum power that line n can carry.

4. Agent training and system control

4.1. Training data
The training of the deep neural network needs to rely on a large amount of data. This paper combines historical operation data of the power grid and simulation data to generate training data and test data. The historical operating data of the power grid is homogenized and cannot reflect all operating scenarios. This paper generates operating data for special scenarios through a power grid simulation method, and supplements and perfects a large amount of historical data with a small amount of simulation data. The specific method is as follows:

1) Obtain power grid simulation running data samples \( \{X_1, X_2, X_3, \ldots, X_n\} \);
2) Calculate the w nearest neighbor samples \( \{Y_{i1}, Y_{i2}, Y_{i3}, \ldots, Y_{iw}\} \) of each sample based on the Euclidean distance;
3) Performed random linear interpolation amplification between the power grid simulation running data sample and its w nearest neighbor samples. After each sample \( X_i \) is interpolated, w synthetic samples \( Z_j \) (j=1, 2, ..., w), \( rand(0,1) \) is a random number in the interval (0,1), \( Z_j \) is calculated as follows:

\[ Z_j = X_i + rand(0,1) \ast (Y_{ij} - X_i) \]  

(8)

4) Mix the samples obtained after interpolation and amplification with historical operating data to obtain a new overall data set;
5) To enhance adaptability, randomly select 3% of the overall data set as the test set, and the remaining 97% data as the training set.

4.2. Model parameters
When training the dueling deep Q network described in this paper, you need to adjust the parameters of the neural network according to the test results; the parameters that need to be adjusted include the learning rate, the number of network layers, the number of layer nodes, and the batch size. The optional range of parameters is as follows:
1) Learning rate. The learning rate is mainly related to the magnitude of the update network weight. Too large a learning rate may cause the model to fail to converge. Too small a learning rate will cause the model to converge slowly. The learning rate is set to \([0.01, 0.005, 0.001, 0.0005, 0.0001]\).

2) The number of network layers and the number of nodes need to cooperate with each other, and are closely related to the scale of the power grid system. Here, the number of network layers is designed to be 6-20, and the number of nodes is 500-2000.

3) Batch size. The batch size is the number of samples sent to the model to train the neural network at a time. Large batches can make the network converge faster, but consume memory resources. The batch size is designed here as \([16, 32, 64, 128, 256]\).

4.3. Control decision
The intelligent control and decision-making of the grid topology is divided into two parts: offline system and online system. The offline system library stores historical events and action records of power grid operation, performs offline simulation of power grid operation control, enhances and perfects data through historical data accumulation and samples, trains the agent through the offline training method and updates the agent model and parameters for use by agents running online. The online system is a real-time operating system. The agent calculates output actions and control commands according to the state variables and reward values fed back by the power grid. The power grid runs according to the control commands, and feeds back the updated status and reward values to the online agent, and stores them in offline system library. The online system uses the agent model trained by the offline system to complete the calculation in seconds, ensuring the real-time control of the grid topology. The power grid topology control and decision-making process is shown in Figure 4.

![Figure 4. The power grid topology control and decision-making process.](image)

5. Conclusions
This paper proposes a modeling and decision-making method for intelligent control of power grid topology. The dueling DQN method is used to construct the action, state and reward variables of power grid topology control, and the control and decision process that integrates online control and offline training of the agent is proposed. This method can solve the problem of huge variables in the traditional power grid topology control model, which is difficult to solve in real time. Through the method in this paper, real-time optimization and control of power network topology can be carried out, and deep research simulation results will be performed in the future study.
Acknowledgments
This research was supported by the National Key Research and Development Program of China (2017YFB0903402), as well as the National Natural Science Foundation of China (51507164).

References
[1] Nourollah, Sara, and Gevork B. Gharehpetian 2018 Coordinated load shedding strategy to restore voltage and frequency of microgrid to secure region IEEE Transactions on Smart Grid 10(4) 4360-4368
[2] Rocchetta, Roberto, et al. 2019 A reinforcement learning framework for optimal operation and maintenance of power grids Applied energy 241 291-301
[3] Wu, Jiajing, et al. 2019 Sequential topology recovery of complex power systems based on reinforcement learning Physica A: Statistical Mechanics and its Applications 535 122487
[4] Lan, Tu, et al. 2019 AI-Based Autonomous Line Flow Control via Topology Adjustment for Maximizing Time-Series ATCs. arXiv preprint arXiv:1911.04263
[5] Zhang, Zidong, Dongxia Zhang, and Robert C. Qiu 2019 Deep reinforcement learning for power system applications: An overview CSEE Journal of Power and Energy Systems 6(1) 213-225
[6] Donnot, Benjamin, et al. 2017 Introducing machine learning for power system operation support. arXiv preprint arXiv:1709.09527
[7] Andres B, Gorka U, Perez J M, et al. 2018 Smart optimization of a friction-drilling process based on boosting ensembles[J] Journal of Manufacturing Systems 48 108-121
[8] Xu X, Jia Y, Xu Y, et al. 2020 A Multi-agent Reinforcement Learning based Data-driven Method for Home Energy Management[J] IEEE Transactions on Smart Grid 99 1-1
[9] Y Du and F Li 2019 Intelligent multi-microgrid energy management based on deep neural network and model-free reinforcement learning IEEE Trans. Smart Grid, early access, doi: 10.1109/TSG.2019.2930299
[10] L2rpn challenge. [Online]. https://l2rpn.chalearn.org/competitions. 2020.10.06.
[11] L2rpn competitions. [Online].https://competitions.codalab.org/competitions/24493. 2020.10.06.
[12] Marot, Antoine, et al. 2020 Learning to run a power network challenge for training topology controllers Electric Power Systems Research 189 106635
[13] Wang, Ziyu, et al. 2016 Dueling network architectures for deep reinforcement learning International conference on machine learning
[14] https://github.com/rte-france/Grid2Op/blob/master/getting_started/0_Introduction.ipynb. 2020.10.06