

Research Article

Optimization Method of Power Equipment Maintenance Plan Decision-Making Based on Deep Reinforcement Learning

Yanhua Yang ¹ and Ligang Yao²

¹School of Engineering, Fujian Jiangxia University, Fuzhou 350108, China

²School of Mechanical Engineering and Automation, Fuzhou University, Fuzhou 350116, China

Correspondence should be addressed to Yanhua Yang; mayspring@163.com

Received 11 July 2020; Revised 2 February 2021; Accepted 19 February 2021; Published 15 March 2021

Academic Editor: Pietro Bia

Copyright © 2021 Yanhua Yang and Ligang Yao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The safe and reliable operation of power grid equipment is the basis for ensuring the safe operation of the power system. At present, the traditional periodical maintenance has exposed the abuses such as deficient maintenance and excess maintenance. Based on a multiagent deep reinforcement learning decision-making optimization algorithm, a method for decision-making and optimization of power grid equipment maintenance plans is proposed. In this paper, an optimization model of power grid equipment maintenance plan that takes into account the reliability and economics of power grid operation is constructed with maintenance constraints and power grid safety constraints as its constraints. The deep distributed recurrent Q-networks multiagent deep reinforcement learning is adopted to solve the optimization model. The deep distributed recurrent Q-networks multiagent deep reinforcement learning uses the high-dimensional feature extraction capabilities of deep learning and decision-making capabilities of reinforcement learning to solve the multiobjective decision-making problem of power grid maintenance planning. Through case analysis, the comparative results show that the proposed algorithm has better optimization and decision-making ability, as well as lower maintenance cost. Accordingly, the algorithm can realize the optimal decision of power grid equipment maintenance plan. The expected value of power shortage and maintenance cost obtained by the proposed method is \$71.75\$ \$MW·H\$ and \$496000\$ \$yuan\$.

1. Introduction

In order to ensure the continuous and safe operation of power equipment, it is necessary to do a good job in preventive plan maintenance and repair of equipment and timely find and eliminate potential accidents. The arrangement of power equipment maintenance plan is an important measure to ensure the safe operation of power grid, reliable power supply, and improve the health level of equipment. With the continuous development of the power system, equipment maintenance is facing new challenges [1]. How to optimize maintenance plan, prevent over maintenance of equipment, save the maintenance cost, and improve the reliability of planned maintenance power supply is problem which is still have to be solved. On the one hand, the increasing number of power equipment makes the

maintenance workload heavier, further increasing the maintenance cost. How to optimize the maintenance plan, prevent over maintenance of equipment, and reduce the maintenance cost is one of the goals to achieve the economic operation of the power grid. On the other hand, the economic development and people's daily life put forward higher requirements for power supply reliability, and planned maintenance is one of the reasons for the reduction of power supply reliability. Therefore, the optimization of equipment maintenance plan and reasonable arrangement of equipment maintenance can ensure the safe and reliable operation of power equipment and improve the reliability of power supply.

The maintenance plan of power equipment depends on manual preparation, which is limited by the professional technical ability and working experience of the personnel.

According to the requirements of the power grid operation mode, dispatching plan, and equipment maintenance regulations, maintenance plan preparation personnel prepares maintenance plan to meet the safe and reliable operation of power grid. However, with the development of the power grid, some problems are gradually exposed in the manual maintenance plan; reliability and economy cannot be guaranteed; many factors need to be considered in the maintenance plan formulation, such as heavy workload and low efficiency; manual planning is not easy to analyze data statistics; the quality of maintenance plan is affected by the technical ability and work experience of professional personnel.

Maintenance plan optimization is a multistage dynamic planning process [2]. The solving algorithm for optimization models includes mathematical optimization method and intelligent optimization method. There are two kinds of optimization models: single objective function model and multiobjective function model. In the economic single objective function model, for example, literature [3] maximizes demand reliability by minimizing total square sum of reserves (SSR) and power production cost (mainly fuel cost) and uses the multiobjective simulated annealing method based on dominance to determine the compromise solution of the model. In [4], the optimization model is established based on the index of network loss, aiming at the minimum of monthly network loss. In the reliability single objective function model, for example, literature [5] puts forward an optimization model of distribution equipment maintenance based on risk assessment, which takes the minimum risk of power grid operation as the optimization objective and uses the particle swarm optimization algorithm to solve it. Based on the risk assessment, literature [6] puts forward the coordinated maintenance strategy for UHV receiving power equipment, which provides reference for ensuring the normal operation of UHV. In [7], based on the equipment status and the risk of power grid loss, the optimization model of power distribution equipment maintenance plan is established. In [8], a maintenance rate optimization model of power equipment asset management is proposed, which uses Markov process and explicitly considers the aging of equipment. In the literature, the algorithm used to solve the problem stays in single objective optimization but does not achieve multiobjective optimization; the proposed optimization algorithm only optimizes the single objective and does not coordinate the economy and reliability of the maintenance plan.

In [9], an improved multiobjective differential evolution algorithm for adaptive optimization of control parameters to solve high-speed rail line planning is proposed, as well as a heuristic algorithm to obtain a better initial solution. In [10], according to application of multiagent reinforcement learning to guide multi workflow scheduling on service cloud, an optimization algorithm is proposed, which takes the number of workflow applications and heterogeneous virtual machines as state input and takes the maximum completion time and cost as rewards. In [11], based on the radial basis function neural network approximation model, a

fast multiobjective optimization algorithm for contactor characteristics is proposed, which can speed up the convergence and get optimal solution. In [12], a multiobjective optimization method based on sensitivity is proposed to optimal configuration of thyristor-controlled series capacitors in the transmission network, which can reduce the total reactive power loss and reduce the load capacity of the transmission line. In reference [13], a coordinated stochastic scheduling model based on the multiobjective optimization method is proposed to improve the absorption capacity of wind power on the premise of energy saving and emission reduction. The application of deep learning and reinforcement learning in multiobjective optimization provides a new idea for decision-making optimization of power equipment maintenance plan.

The development and application of AI [9], big data [10], parallel computing, and optimization algorithm and theory [14, 15] create favorable conditions for the optimization of power equipment maintenance plan. Deep learning [16] (DL) is composed of multilayer nonlinear units, which can automatically learn abstract features from training data; reinforcement learning (RL) [17] has strong decision-making ability, high adaptability, and advantages of decision optimization; deep reinforcement learning (DRL) [18] uses the high-dimensional feature extraction ability of DL and the decision-making ability of RL to solve the decision-making problems in high-dimensional state space and high-dimensional action space [19–21]. The application of the deep reinforcement learning method in the power system is mainly in intelligent power generation control, power grid intelligent control, and other fields [1, 22, 23].

Deep reinforcement learning is an end-to-end perception and decision-making system, which has universality. The process of perceptual learning and decision-making optimization is as follows: the agent interacts with the environment to get a high-dimensional observation state s , and uses the deep learning method to get specific feature representation; based on the reward evaluation value function, the current state is mapped to the corresponding action by decision; the environment reacts to the action and gets the next observation. Finally, the optimal decision of the target can be obtained by continuously cycling the above process.

In this paper, a multiobjective mathematical model of power equipment maintenance optimization considering reliability and economy is established, and a kind of deep distributed Q-network is adopted (Q-networks, DDRQN) to realize multiagent deep reinforcement learning solution, to make full use of DRL's optimization ability and decision-making ability, and to realize the intelligent arrangement of power equipment maintenance plan. The comparison experiment has processed the new method and single agent deep reinforcement learning algorithm as well as the particle swarm intelligence algorithm for the deep cyclic Q-network. The results show that the proposed algorithm has higher reliability, lower maintenance cost, and more reasonable maintenance plan.

2. Optimization Model of Power Equipment Maintenance Plan

2.1. Objective Function. Power equipment maintenance plan optimization is a multiobjective and multiconstraint optimization problem. The objective function can be divided into three categories: reliability objective function, economy objective function, and practicability objective function [24]. The reliability objective functions include loss of load probability (LOLP), minimum load loss due to power failure, maximum system reliability index, and minimum expected energy not supplied (Expected Energy Not Supplied, EENS). The economic objective function includes maintenance cost and outage loss. The practical goal is to put forward due maintenance, reduce maintenance outage, and even distribution of maintenance workload from the practical point of view.

EENS is defined as the sum of the power loss caused by the outage of the power equipment. It evaluates the power supply reliability reduction caused by the outage of the equipment. The objective function expression is as follows:

$$f_1 = \sum_{t=1}^T \left(\sum_{x \in S_t} C_x \prod_{i=1}^M P_i^{x_i} (1 - P_i)^{1-x_i} \right) T_t, \quad (1)$$

where t refers to the number of maintenance periods, S_t refers to a collection of states, $x = (x_1, x_2, \dots, x_n)$ means the state vector of the equipment, C_x refers to the load shedding under the fault state, M refers to the number of equipment, $x_i = 1$ refers to the shutdown state, $x_i = 0$ means the equipment is in the running state, P_i refers to the equipment outage probability, T_t refers to the number of unit hours, and the unit is MW·H.

The maintenance cost of the equipment refers to the expenses incurred by the maintenance of the equipment during the maintenance period. The objective function expression is as follows:

$$f_2 = \sum_{i=1}^N \sum_{t=1}^T p_i^t z_i^t u_i^t, \quad (2)$$

where N is the total number of equipment to be overhauled; p_i^t represents the unit time cost; z_i^t represents the maintenance team arranged for the equipment i in the t period; $u_i^t = 1$ represents the outage maintenance of the equipment i in the t period, $u_i^t = 0$ represents the normal operation, and unit is 10000 yuan.

2.2. Constraints

2.2.1. Time Constraints. Any maintenance work shall be completed on time.

$$u_i^t \begin{cases} 1s_i \leq t \leq s_i + m_i, \\ 0, \end{cases} \quad (3)$$

where u_i^t refers to the equipment i maintenance status, $u_i^t = 1$ means the equipment i is powered off for maintenance in the t period, $u_i^t = 0$ means the equipment i is in normal

operation for maintenance in the t period, s_i means the equipment i starts maintenance in the s_i period, m_i is the maintenance period, e_i means the earliest time to start maintenance, and l_i means the latest time to start maintenance.

2.2.2. Maintenance Resource Constraints. In a maintenance cycle, the number of equipment that can be maintained at the same time is limited.

$$\sum_{i=1}^N z_i^t u_i^t \leq Z_{\max}, \quad (4)$$

where Z_{\max} indicates the upper limit of maintenance workload in the t period.

2.2.3. Simultaneous Maintenance Constraints. Equipment maintenance shall avoid repeated power failure at the same load point, and all problems shall be solved during power failure maintenance.

$$s_k = s_i, \quad (5)$$

where s_k is the time for equipment k to start maintenance and s_i is the time for equipment i to start maintenance.

2.2.4. Mutually Exclusive Maintenance Constraints. To prevent the expansion of power outage, some power equipment cannot be arranged for maintenance at the same time.

$$s_j > s_i + m_i + 1, \quad (6)$$

where s_j is the time when equipment J starts maintenance; s_i is the time when equipment i starts maintenance; m_i is the maintenance period of equipment i .

2.2.5. Security Constraints. In order to ensure the safe and stable operation of power grid, the maintenance plan must be checked by power flow calculation.

$$I_i \leq I_{i,\max}, \quad (7)$$

$$V_{i,\min} \leq V_i \leq V_{i,\max}, \quad (8)$$

$$P_i \leq P_{i,\max}, \quad (9)$$

where $I_{i,\max}$, $V_{i,\max}$, $V_{i,\min}$, and $P_{i,\max}$ are, respectively, expressed as the upper and lower limit of current flow, the upper and lower limit of node voltage, and the allowable power limit. I_i is the current flowing through equipment I , V_i is the node voltage amplitude, and P_i is the active power passing through the device.

2.3. Optimization Model. In this paper, the expected value of energy shortage (EENS) is taken as the reliability evaluation index, and the maintenance cost is taken as the economic evaluation index of maintenance plan.

$$\begin{cases} \min f_i(X), \\ \text{s.t. } g_i(X) = 0, \\ h_j(X) \leq 0, \end{cases} \quad (10)$$

where $f_i(X)$ represents the number i objective function of power equipment maintenance plan, X is the n -dimensional decision vector, m means number of objective function, $g_i(X)$ and $h_j(X)$, respectively, represents the equality constraint function and inequality constraint function, p is the number of equality constrained functions, and q is the number of inequality constrained function.

In most cases, the economic objective function and reliability objective function are in conflict with each other. The improvement of one objective may lead to the performance degradation of the other, and it is impossible to make both objectives reach the optimal solution at the same time. It can only be dealt with in the middle of them to make each subobjective reach the optimization as much as possible. The solution of a multiobjective optimization problem is usually a set of solutions that meet the conditions. For the solution set, we can find a set of solutions that are as close to the best as possible.

3. Optimization Model Solving Method

3.1. DDRQN Algorithm. For complex decision-making problems, the decision-making ability of the single agent system is still far from enough. The multiobjective power equipment maintenance planning model established in this paper is competitive or cooperative in economy and reliability. Therefore, under certain conditions, the DRL algorithm needs to be extended to a multiagent system in which multiple agents cooperate or compete with each other. DDRQN allocates the deep recurrent q -networks (DRQN) training module for each agent to build a multiagent system. Its function expression is as follows:

$$L_t^m = [(y_t^m - Q(o_t^m, h_{t-1}^m, m, a_{t-1}^m, a_t^m; \theta_i))^2], \quad (11)$$

where y_t^m is the target Q -value function:

$$y_t^m = \begin{cases} R_t, & \text{if } s_t, \\ R_t + \gamma \cdot Q(o_{t+1}^m, h_t^m, m, a_t^m, a; \theta_i^-), \end{cases} \quad (12)$$

where $Q(o_t^m, h_{t-1}^m, m, a_{t-1}^m, a_t^m; \theta_i)$ is expressed as Q function, o_t^m is the observation of agent with number m at time t , h_{t-1}^m represents the long- and short-term memory hidden layer state of agent number m at time $t-1$, a_{t-1}^m means the action corresponding to the current Q -value function of the agent number m , θ_i is the network weight corresponding to the agent number m in the i round iteration, θ_i^- is the target network weight corresponding to the agent number m in the i round iteration, m is the number of the agent currently processed, a_{t-1}^m is a part of the state action history sequence, γ is the discount factor, and s_t is the status of time t . R_t is the reward function. θ_i and θ_i^- are expressed as follows:

$$\begin{aligned} \theta_{i+1} &= \theta_i + \alpha \cdot \nabla \theta, \\ \theta_{i+1}^- &= \theta_i^- + \alpha^- \cdot (\theta_{i+1} - \theta_i^-), \end{aligned} \quad (13)$$

where $\nabla \theta$ is the gradient value, α is the learning rate, and α^- is the target learning rate.

3.2. Transforming Optimization Objectives into Deep Reinforcement Learning Tasks. A deep reinforcement learning problem must have two conditions: agent and environment. Its learning and decision-making process is as follows: according to agent strategy and environment information, the agent makes corresponding actions to make the reward optimal [25, 26]. Deep reinforcement learning has been successful in other areas, but the decision-making optimization of power equipment maintenance plan is not the same as other environments. Therefore, the key to its successful application is to transform the decision-making optimization problem of power equipment maintenance plan into deep reinforcement learning task.

The decision-making and optimization of power equipment maintenance plan can be used as a dynamic environment, and the equipment to be overhauled can be used as the a_t of an agent. For any t time, the power generation and load of the system can be used as a state, and s_t is defined as follows:

$$s_t = \{P_G, P_{\text{Load}}\}. \quad (14)$$

Under the action of a_t , the reward of environmental feedback can be determined by optimization objective (1). The decision-making optimization problem of power equipment maintenance plan is transformed into solving the optimal solution of the equation set. Then, Build up an action A for the power equipment which need to be overhauled. After that, calculate and get the state s_t and reward a_t of feedback, when the action value is r_t . The memory unit is shown as $d_t = \{a_t, s_t, r_t\}$. By calculating the loss error, the agent learns the strategy corresponding to the power equipment maintenance problem, namely, the maintenance plan. The solution expression is as follows:

$$\begin{cases} L_t^m = [(y_t^m - Q(o_t^m, h_{t-1}^m, m, a_{t-1}^m, a_t^m; \theta_i))^2], \\ y_t^m = \begin{cases} R_t, & \text{if } s_t, \\ R_t + \gamma \cdot Q(o_{t+1}^m, h_t^m, m, a_t^m, a; \theta_i^-), \end{cases} \\ R_t = f(X) = (f_1(X), f_2(X)). \end{cases} \quad (15)$$

The agent learning process is shown in Figure 1.

3.3. Solution Flow. The decision-making and optimization process of the power equipment maintenance plan is shown in Figure 2, which is mainly divided into three parts: data preparation, algorithm solution, and safety check.

Data preparation: acquisition of initial data, including the list of equipment to be overhauled and the date of overhauling, the overhauling plan arranged by the superior, the remaining overhauling plan, and the data of power grid

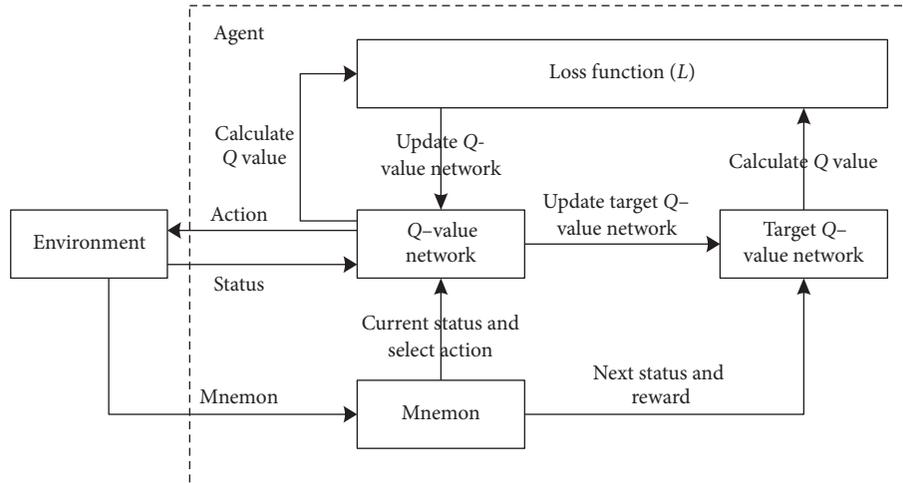


FIGURE 1: Schematic diagram of agent learning process.

operation mode; establishment of the optimization model of power equipment overhauling plan based on the acquired initial data according to expression (15); adjustment of power grid operation mode based on the scheduled overhauling plan, and adjustment of the adjusted operation mode. The mode of safety check is adopted to ensure that the power grid operation is safe and stable at this time. If the safety and stability conditions are not met, the alarm information will be output.

Algorithm solution: DDRQN algorithm solution process is used according to the current power grid operation state to get the power generation P_G and load of the system P_{Load} so as to get the initial state s_t ; at this time, the power grid is safe and stable operation state; build action set A with the power equipment to be overhauled; in the process of performing an action a_t exploration, get the state s_{t+1} and reward r_{t+1} so as to calculate the target Q-value function y_t^m , Q-value function, and loss error L_t^m . Through the gradient descent algorithm, all parameters of the DDRQN network are updated; through one-time solution process, under the condition of equipment I maintenance, EENS and maintenance cost of power grid at this time are obtained; based on a certain strategy, under the condition of state s_{t+1} , the next exploration action a_{t+1} is made to start the next solution process, and the solution is not terminated until all equipment to be overhauled are arranged and the arrangement results meet the constraint conditions. At this time, the strategy is the power equipment maintenance plan.

Safety check: take the maintenance strategy as the section to calculate the power flow and carry out safety check. If the constraint conditions (7), (8), and (9) for safe operation of the power grid are not met, update the parameters a_t , s_t , and r_t of the DDRQN algorithm and solve it again. Carry out $n - 1$ check for the maintenance plan and output the alarm prompt information if the condition for stable operation of the power grid is not met. After a series of iterative solutions, the optimal maintenance strategy is obtained.

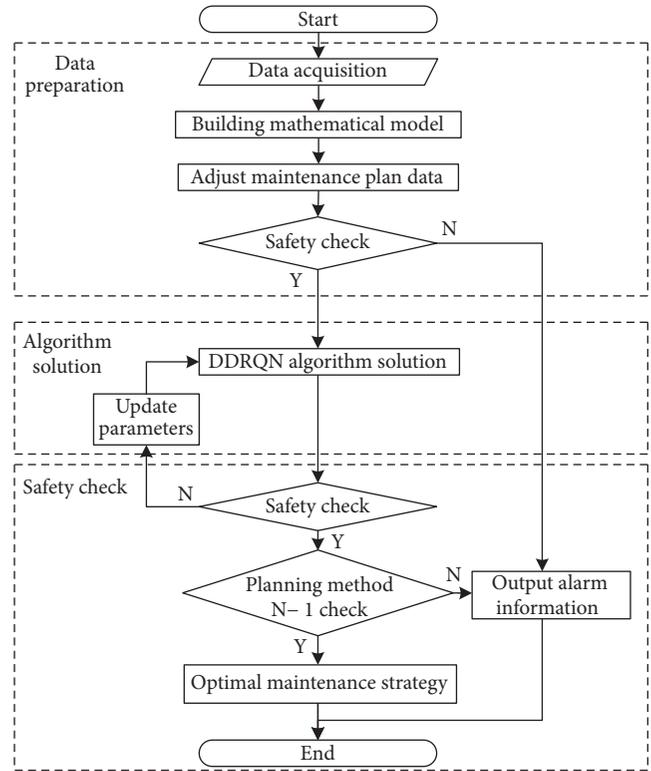


FIGURE 2: Flow chart of decision optimization of the power grid maintenance plan.

4. Case Analysis

In order to verify the distributed deep cycle Q network to achieve the decision-making optimization of power equipment maintenance plan, this paper uses the IEEE-118 node system to verify its network topology as shown in Figure 3, including 54 generators and 132 lines. The verification environment is based on MATLAB R2016 with Intel Core i7-8550U CPU and 8GB RAM computer.

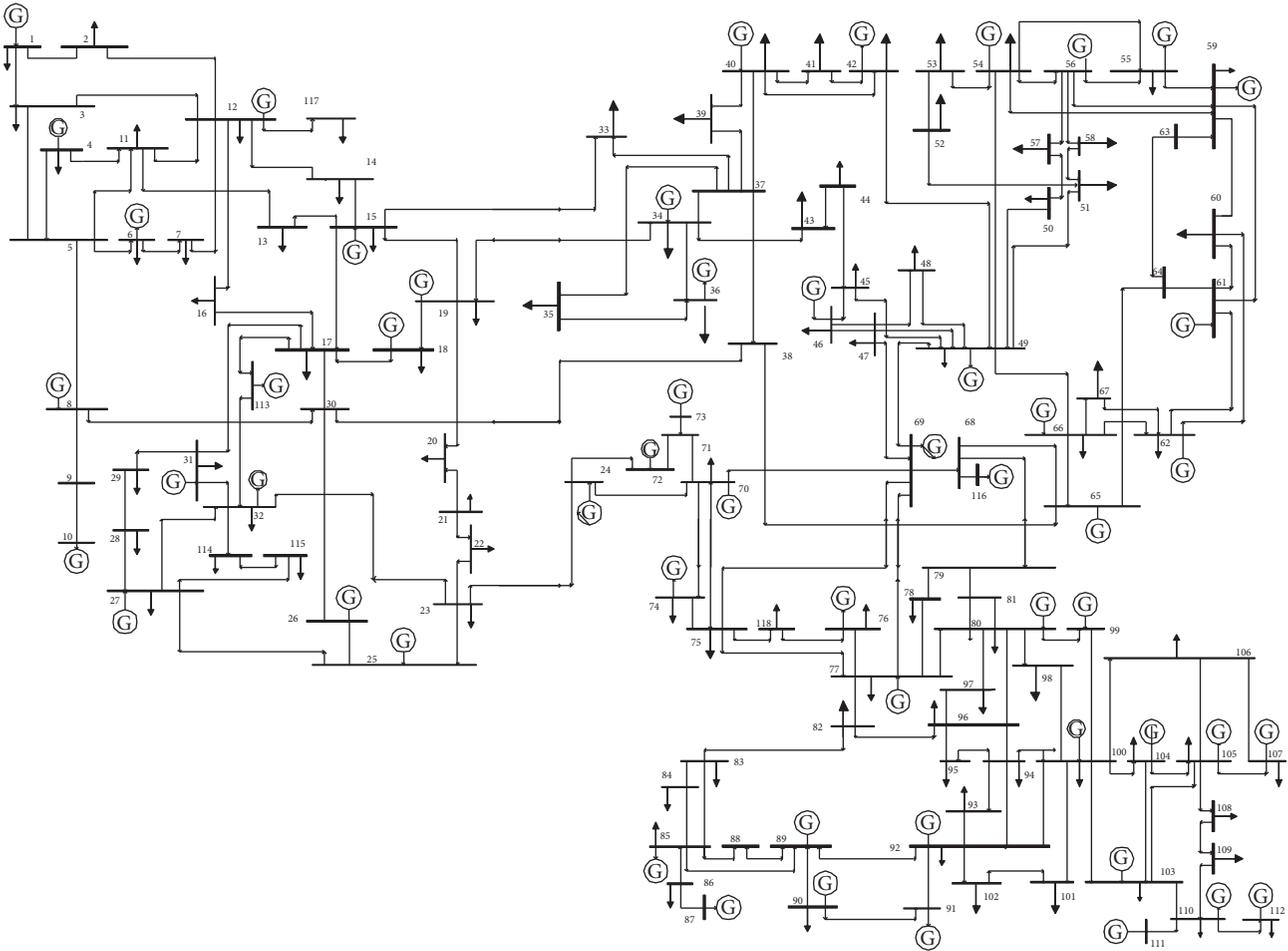


FIGURE 3: Topology of the IEEE-118 node network.

4.1. Parameter Setting. In the process of model solving, three system parameters γ , α , and α^- need to be set, which are usually empirical values. For the learning rate α , the value range is (0, 1); the larger α can speed up the learning speed, and the smaller α can improve the stability of the solution. α^- is the target learning rate, which is used to update the network weight of the target value function. γ is the discount factor, which is used to discount the reward of the target value function. In this paper, we first use a wider search range and a larger step size to find the possible location of the global optimal value and then narrow the search scope and step size in order to find a more accurate optimal value.

4.2. Result Analysis. DDRQN algorithm, deep recurrent Q-networks algorithm, and particle swarm optimization (PSO) algorithm are used to solve the decision-making optimization model of power equipment maintenance plan established in this paper. The results are shown in Figures 4 and 5. The abscissa represents the number of iterations, and the ordinate represents EENS (MW·h) and maintenance cost

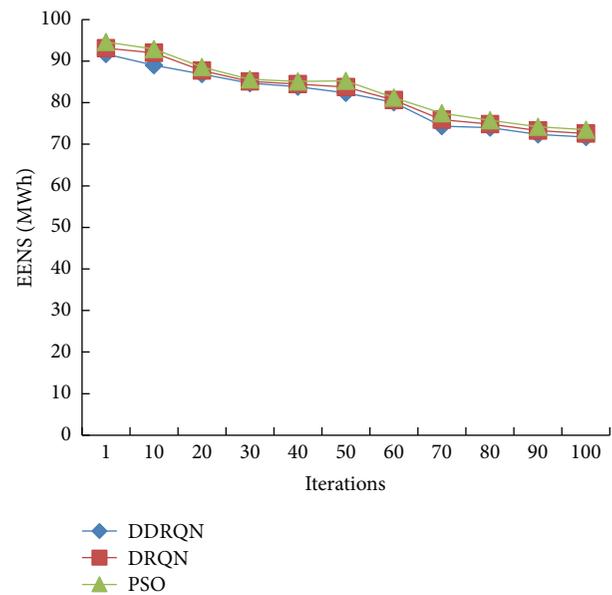


FIGURE 4: Reliability comparison of plan.

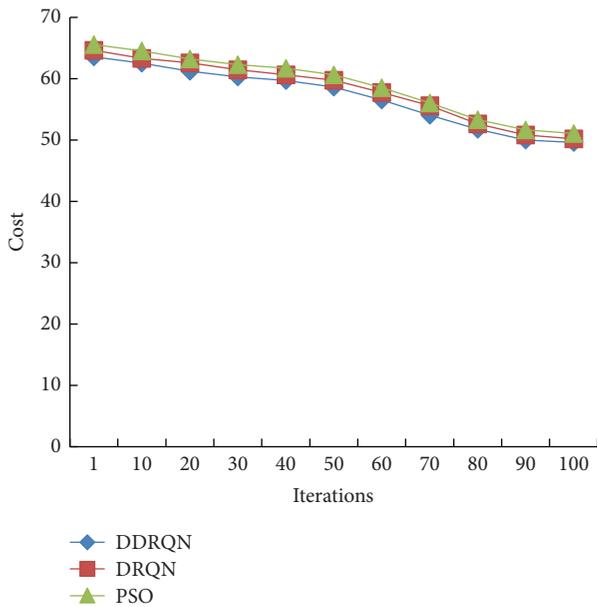


FIGURE 5: Economic comparison of plan.

TABLE 1: Results of different optimization algorithms.

	EENS (MW-h)	Maintenance cost (ten thousand yuan)
PSO	73.48	51.06
DRQN	72.55	50.20
DDRQN	71.75	49.60

(10000 yuan). Table 1 shows the solution obtained by 100 iterations of the optimization algorithm.

As can be seen from Table 1, the results of the optimization algorithm proposed in this paper are better than those of the other two methods because EENS and maintenance costs are less than the other two methods.

It can be seen from the curve in Figure 4 that the DDRQN algorithm proposed in this paper is superior to the PSO algorithm and DRQN algorithm in terms of algorithm convergence performance; the maintenance scheme obtained by DDRQN algorithm, DRQN algorithm, and PSO algorithm, respectively, is superior to the other two algorithms in terms of reliability and economy. The proposed algorithm can effectively obtain the optimal cooperative control and the quality of decision-making in the multiagent system by using its strong comprehensive decision-making ability and coordination interaction ability. Therefore, the overall performance has been greatly improved.

5. Conclusion

In this paper, the maintenance plan optimization model of power equipment is established. The reliability evaluation index EENS is the expected value of electric energy shortage, and the maintenance cost is the economic evaluation index of the maintenance plan. The reliability and economy of power grid operation are fully considered. In this paper, a novel distributed deep loop Q network multiagent deep reinforcement learning algorithm is proposed to obtain the

optimal decision of power equipment maintenance plan and improve the accuracy and intelligence level of making power equipment maintenance plan. The DDRQN algorithm uses its own comprehensive decision-making ability and coordination interaction ability to get the optimal decision among multiagent. Through case analysis, the results show that the DDRQN algorithm has strong learning ability, high adaptability, strong comprehensive decision-making ability, and better optimization performance; it can reasonably plan the maintenance plan, and the maintenance plan obtained fully takes into account the economy and reliability. The practical guiding significance of this algorithm is to realize the accuracy and automation of scheduling in the scheduling of power grid maintenance plan.

Data Availability

The processed data required to reproduce these findings cannot be shared at this time as the data also form part of an ongoing study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] R. Rocchetta, L. Bellani, M. Compare, E. Zio, and E. Patelli, "A reinforcement learning framework for optimal operation and maintenance of power grids," *Applied Energy*, vol. 241, no. 5, pp. 291–301, 2019.
- [2] M. Fattahi, M. Mahootchi, H. Mosadegh, and F. Fallahi, "A new approach for maintenance scheduling of generating units in electrical power systems based on their operational hours," *Computers & Operations Research*, vol. 50, no. 10, pp. 61–79, 2014.
- [3] B. G. Lindner, R. Brits, J. H. van Vuuren, and J. Bekker, "Tradeoffs between levelling the reserve margin and minimizing production cost in generator maintenance scheduling for regulated power systems," *International Journal of Electrical Power & Energy Systems*, vol. 101, no. 10, pp. 458–471, 2018.
- [4] C. Dai, Z. Wang, Y. Dai et al., "Economic optimization method of transmission equipment maintenance scheduling based on declining index of transmission loss," *Northeast Electric Power Technology*, vol. 38, no. 12, pp. 4–9, 2017.
- [5] L. I. Ran, F. Wang, and Z. Li, "Maintenance decision making optimization based on risk assessment for distribution system," *Electric Power Automation Equipment*, vol. 33, no. 11, pp. 1–8, 2013.
- [6] X. Yang, Y. Yin, and X. Kou, "Research on the maintenance strategy of UHV receiving end power grid transmission equipment based on risk assessment," *Electric Power Engineering Technology*, vol. 36, no. 2, pp. 72–75, 2017.
- [7] E. Li, C. Kang, Y. Li et al., "Optimization model of distribution network maintenance plan based on equipment condition evaluation and grid loss risk," *High Voltage Engineering*, vol. 44, no. 11, pp. 3751–3759, 2018.
- [8] B. Xu, S. Xu, and Y. Zhang, "Determining optimal inspection rates of power equipment considering opportunistic maintenance strategy," 2019.

- [9] H. Tian, M. Shuai, and K. Li, "Optimization study of line planning for high speed railway based on an improved multi-objective differential evolution algorithm," *IEEE Access*, vol. 7, no. 7, pp. 137731–137743, 2019.
- [10] Y. Wang, H. Liu, W. Zheng et al., "Multi-objective workflow scheduling with deep-Q-network-based multi-agent reinforcement learning," *IEEE Access*, vol. 7, no. 7, pp. 39974–39982, 2019.
- [11] M. Q. Raza and A. Khosravi, "A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings," *Renewable and Sustainable Energy Reviews*, vol. 50, no. 10, pp. 1352–1372, 2015.
- [12] N. A. Bakar and M. K. M. Desa, "Optimal placement of tcsc in transmission network using sensitivity based method for multi-objective optimization: 2017 IEEE conference on energy conversion," 2017.
- [13] Y. Li, Z. Ni, T. Zhao et al., "Coordinated stochastic scheduling for improving wind power adsorption in electric vehicles-wind integrated power systems by multi-objective optimization approach," 2019.
- [14] S. Twaha and M. A. M. Ramli, "A review of optimization approaches for hybrid distributed energy generation systems: off-grid and grid-connected systems," *Sustainable Cities and Society*, vol. 41, no. 8, pp. 320–331, 2018.
- [15] F. S. Abu-Mouti and M. E. El-Hawary, "Overview of artificial bee colony (ABC) algorithm and its applications," 2012.
- [16] D. Zhang, X. Han, 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.
- [17] Z. Zhang, D. Zhang, and R. C. Qiu, "Deep reinforcement learning for power system Applications: an overview," *CSEE Journal of Power and Energy Systems*, vol. 6, no. 1, pp. 213–225, 2020.
- [18] F. Aznar, M. Pujol, and R. Rizo, "Obtaining fault tolerance avoidance behavior using deep reinforcement learning," *Neuro Computing*, vol. 345, no. 6, pp. 77–91, 2019.
- [19] J. Liu, F. Gao, and X. Luo, "Survey of deep reinforcement learning based on value function and policy gradient," *Chinese Journal of Computers*, vol. 42, no. 6, pp. 1406–1438, 2019.
- [20] L. Wan, X. Lan, H. Zhang et al., "A review of deep reinforcement learning theory and application," *Pattern Recognition and Artificial Intelligence*, vol. 32, no. 1, pp. 67–81, 2019.
- [21] Q. Liu, J. Zhai, Z. Zhang et al., "A survey on deep reinforcement learning," *Chinese Journal of Computers*, vol. 41, no. 1, pp. 1–27, 2018.
- [22] M. Glavic, R. Fonteneau, and D. Erns, "Reinforcement learning for electric power system decision and control: past considerations and perspectives," *IFAC-papers OnLine*, vol. 50, no. 7, pp. 6918–6927, 2017.
- [23] H. Li, R. Cai, N. Liu, X. Lin, and Y. Wang, "Deep reinforcement learning: algorithm, applications, and ultra-low-power implementation," *Nano Communication Networks*, vol. 16, no. 6, pp. 81–90, 2018.
- [24] X. Cui, L. I. Wei, and B. Li, "Review on transmission equipment maintenance planning decision-making," *Power System and Clean Energy*, vol. 31, no. 12, pp. 18–26, 2015.
- [25] Y. Zhang, P. Cai, C. Pan, and S. Zhang, "Multi-agent deep reinforcement learning-based cooperative spectrum sensing with upper confidence bound exploration," *IEEE Access*, vol. 7, pp. 118898–118906, 2019.
- [26] W. U. Jia, S. Chen, X. Chen et al., "Reinforcement learning for model selection and hyperparameter optimization," *Journal of University of Electronic Science and Technology of China (Social Sciences Edition)*, vol. 49, no. 2, pp. 255–261, 2020.
- [27] A.-H. Hossein and M.-R. Hamed, "Power systems big data analytics: an assessment of paradigm shift barriers and prospects," *Energy Reports*, vol. 4, no. 11, pp. 91–100, 2018.