INTEGRATED ACTIVE-REACTIVE POWER DISPATCHING MODEL BASED ON DEEP REINFORCEMENT LEARNING

Zizhao LIN[1*], Xuefei LIU[2], Qidong TIAN[3]

The integrated active and reactive power dispatching model based on deep migration reinforcement learning is proposed. Reinforcement learning can guide the agent to acquire optimal scheduling knowledge in the source task, and the migration network can realize the application of the knowledge learned in the source task to new operating scenarios. Through the simulation of IEEE 14 buses, the design method of key elements of agent is given, including the construction of state, action and reward function. By calculating the similarity of the scene, the dispatching strategy under the new scene can be quickly obtained.

Keywords: artificial intelligence; optimal distribution of active power; automatic voltage control; deep reinforcement learning; DDPG algorithm

1. Introduction

Power system is a nonlinear artificial system with multivariable coupling and complex operation behavior. Power grid dispatching control center is the "control center" for safe, reliable and economic operation of power system [1][2]. For a long time, the power grid regulation operation is mainly based on the staff monitoring the operation state of the power system through equipment to assist in analysis, and then making dispatching decisions according to personal judgment [3]. Every link in the decision-making depends more on the knowledge level and experience of the staff, and the regulators need to participate and lead, which is subjective and different. At the same time, the power grid structure is becoming more and more complex due to distributed energy access, expansion and other factors, the requirements for operation dispatching mode are becoming more and more strict, and the complexity of multivariable coupling is increasing [4][6]. It is urgent to find an intelligent method to assist in solving the traditional active and reactive power integrated dispatching method that depends on Mechanism Analysis and dispatcher's personal decision-making, so as to support the economic and safe operation of the power grid.

In the power systems daily operation, there are classical nonlinear programming methods and heuristic algorithms for the day ahead scheduling

[1] Shenzhen Power Supply Co., Ltd., Shenzhen 518000, China, e-mail: linzizhao@126.com
[2] Shenzhen Power Supply Co., Ltd., Shenzhen 518000, China.
[3] Shenzhen Power Supply Co., Ltd., Shenzhen 518000, China.

problem of active power allocation and reactive power allocation [7][10]. However, due to the discreteness of variables, the objective function and constraint conditions of the power system are discontinuous. At the same time, the power system is a multivariable coupled nonlinear artificial system. Many local optimal solutions often exist in the optimization solution. The traditional optimization methods such as Newton's method and Interior Point Method are easy to fall into local optimum when solving the day ahead scheduling [11][12]. And the classical mathematical methods are often difficult to solve because of the optimization model presents the discontinuity, non-derivation, multi extremum and multi segment constraints of the optimization model presents [13].

On the other hand, traditional heuristic algorithms such as Particle Swarm optimization, Genetic Algorithm, artificial bee colony algorithm and Gray Wolf Optimizer rely less on mathematical models than traditional optimization methods, can solve nonlinear and discontinuous optimization problems, and have successful application scenarios in various optimization problems of power system [14]-[17]. However, the solution time of such algorithms is too long to meet the timeliness of scheduling, especially the real-time optimization task of short time scale in large-scale power system [18]. In addition, the classical nonlinear programming methods and heuristic algorithms do not have the ability of "knowledge learning" and "migration application" [19]. When optimizing the scheduling of new scenarios, the solution process must be run from scratch, and the optimization experience cannot be accumulated from the solution history [20]. As a result, the optimization solution of this kind of algorithm is independent every time, and it cannot learn quickly according to the past experience, so it is difficult to move and apply to other scenarios. Not only the solution time is too long, but also the scope of application is limited by the training set, which is difficult to meet the optimal scheduling requirements of increasingly complex power systems.

With the rapid development of artificial intelligence technology and its recent success in many fields including autopilot and game, the ability of autonomous learning and decision making for deep reinforcement learning in complex dynamic systems is reflected. Experts and scholars in the field of power have always paid attention to the application of artificial intelligence technology in power system [21][23]. Artificial intelligence algorithm has been preliminarily applied to active or reactive power dispatching in power system. In recent years, some scholars have tried to apply artificial intelligence algorithms such as DQN, AC, A2C, PG, DDPG to the research of power grid AVC [24], AGC [25] and load side regulation [26], but they only consider the separate optimization of active or reactive power. Reference [27] analyzes the current situation and existing problems of the application of artificial intelligence in power system, gives the design criteria and general framework of regulation system based on artificial intelligence, and expounds the technical difficulties of the application of artificial intelligence in power grid regulation. Reference [28] proposed an economic generation scheduling and control framework, which can obtain smaller control error, smaller total cost and fewer reverse regulation times, so as to solve the coordination problem of multi-scale economic scheduling and generation control in power system. Reference [29] takes the prediction mechanism of neural network as the action selection mechanism of reinforcement learning to solve the problem of random disturbance caused by large-scale access of distributed energy to the power grid. The above literatures are all successful scenarios of deep reinforcement learning in the field of power system, but there are generally low data utilization efficiency. It is necessary to retrain an agent for different types of tasks, which often requires a lot of interaction with the environment, which limits the application of reinforcement learning in some practical scenarios. In addition, the scheduling optimization algorithm based on reinforcement learning has insufficient generalization ability in unfamiliar scenes. If it is retrained in new scenes, it will cost a lot of time. Therefore, how to combine the continuously developed advanced algorithms with the needs of the power system field and better solve the practical needs of the energy field based on the advantages of algorithms is still in a primary exploration stage and quite challenging.

The innovation of this paper is to apply deep reinforcement learning to improve the DDPG algorithm to solve the problem of active and reactive power integration and coordinated operation in power dispatching. Different from the discrete action interval in other articles, the action interval in this paper is mainly continuous action interval; compared with the traditional DDPG method, the accuracy of its application in power grid dispatching is improved; compared with genetic algorithm and other heuristic algorithms, the accuracy of the application of "knowledge transfer" of the improved DDPG algorithm is confirmed, and it has significant timeliness.

2. Active and reactive power integrated dispatching framework of power system

2.1 Basic power system dispatching problems

The basic task of power system dispatching is to control the operation mode of power system so that it can meet the requirements of safe, economic and high-quality power supply under normal or accident conditions.

He key problem of traditional power grid optimal dispatching is that the model solution is complex. Some functional decoupling and model simplification have to be carried out in order to meet the needs of multi scenario rapid solution. Therefore, the operation performance of power grid is difficult to be more optimized. In recent years, artificial intelligence methods represented by deep reinforcement learning have made remarkable achievements in solving high-dimensional nonlinear optimization problems, which makes it possible to train agents to effectively deal with the scheduling knowledge of various scenarios, and then use the knowledge for integrated scheduling.

Using the knowledge extraction ability of reinforcement learning method, this paper regards the optimal dispatching problem of power system as two optimization objectives considering power grid operation cost and voltage level, takes the active output of generator, the given voltage of generator and dimmer, the compensation of transformer tap and reactive power compensation equipment as control variables, and takes power components and network constraints as basic constraints, The active and reactive power distribution of power grid is coordinated and optimized, and the migration learning method is introduced to deal with the complex and changeable operation scenarios of power grid. Under this intelligent scheduling architecture, more objectives or complex nonlinear constraints can be easily considered in combination with engineering requirements. However, in order to facilitate the comparison of the income increment brought by intelligent scheduling and traditional scheduling methods, this paper defines the integrated scheduling problem as above.

2.2 The basic method of applying active and reactive power integrated dispatching based on reinforcement transfer learning

For the problem of active and reactive power coordination and optimization, the traditional reinforcement learning usually includes: observing the power grid environment to determine the state quantity, the content of the action quantity, designing the corresponding objective function according to the actual problems to be solved, the agent takes actions according to the state to interact with the environment to obtain returns, taking different actions and the obtained returns in the same state to store as experience Apply the learned experience to practice. When applying transfer reinforcement learning, the difference lies in the clear division of the scope of the source problem and the transfer problem, as well as the transfer method. When a new agent obtains a new scene, it first judges whether it is an "old problem" under the source problem domain. If the source problem domain has this problem label, the existing knowledge is adopted; if the label error exceeds the threshold, it is judged as an expansion problem and transfer learning is carried out. The whole process is shown in Fig. 1.

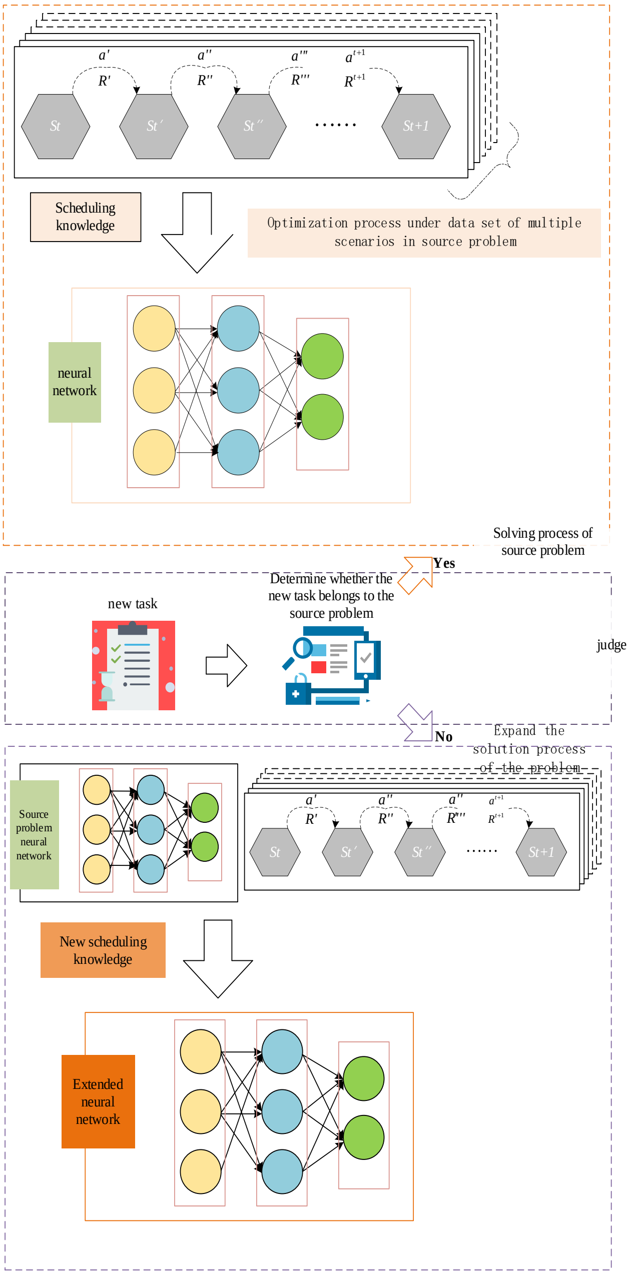
a'
a''
a'''
a^{t+1}
R'
R''
R'''
R^{t+1}
St
St'
St''
……
St+1
Scheduling knowledge
Optimization process under data set of multiple scenarios in source problem
neural network
Solving process of source problem
Yes
Determine whether the new task belongs to the source problem
new task
judge
No
Expand the solution process of the problem
Source problem neural network
a'
a''
a''
a^{t+1}
R'
R''
R'''
R^{t+1}
St
St'
St''
……
St+1
New scheduling knowledge
Extended neural network

Fig. 1. Schematic diagram of the different approaches to the source and expansion problems

3. Active and reactive power coordination optimization modeling based on Reinforcement Learning

3.1 Introduction to deep transfer reinforcement learning

A reinforcement learning solution process can be represented by Markov decision process. Usually, a Markov decision process includes state matrix, return matrix, action matrix, P,γ. The state matrix consists of all States s in the environment that affect the decision-making of agent; the action matrix is composed of a set of decision actions a that can be taken by the agent. The return matrix is composed of the short-term return r, which return by the environment according to s and a, to judge the impact on the s; γ is uncertainty coefficient of long-term return, avoid the agent relying too much on future rewards. When the dimension of state action matrix of reinforcement learning is very large, the exploration stage of finding the optimal strategy is time-consuming; at the same time, large data samples for training are often difficult to obtain in practical problems, and retraining is very time-consuming. Therefore, scholars have studied the possibility of applying the advantages of transfer learning to reinforcement learning, hoping to transfer knowledge from source tasks to new tasks to improve performance. According to the differences of application methods, there are three kinds of transfer settings in reinforcement learning.

1. Instance Transfer. The simplest migration algorithm collects samples from different tasks and reuses them in the learning of target tasks.

2. Representation Transfer. Each RL algorithm uses specific representations for tasks and solutions, such as neural networks, or a set of basis functions that approximate the optimal value function.

3. Parameter Transfer. The migration method changes and adjusts the algorithm parameters according to the source task, so as to speed up the learning process.

In this paper, in order to speed up the accuracy and timeliness of online application, the idea of parameter transfer is used to optimize the traditional reinforcement learning.

3.2 Reward function design

Artificial intelligence method is very important for the setting of reward function. The setting of reward is the driving force for agents to move in a better direction. Setting a good reward function can make the training converge quickly and accurately. In this paper, the following objective function is set to solve the active and reactive power coordination optimization problem:

$$\max = k_1 \times a^{-\frac{\sum_{i=1}^{n} f(P_{Gn})}{c_1}} + k_2 \times b^{-c_2 \times (\sum_{i=1}^{n} \mathrm{Relu}(U_i^{load} - 1.05))} + k_3 \times b^{-c_3 \times (\sum_{i=1}^{n} \mathrm{Relu}(0.95 - U_i^{load}))} \quad (1)$$

The control problem of day ahead scheduling is often the problem of minimizing the objective function. Reinforcement learning is often trained in the direction of maximizing the Q value, so the objective function is slightly processed here to make it more consistent with the reinforcement learning mode of thinking. The Relu function in equation (1) is a primary function with only a positive half axis, and the negative half axis is 0. k_1, k_2, k_3, a, b, c_1, c_2, c_3, are constants, k_1, k_2, k_3 is the proportion coefficient, which represents the proportion between different targets, because generally speaking, the economy of generator cost is more concerned than voltage level. *a, b* is the base of the exponential function. The purpose of using the exponential function is to normalize different objective functions to the interval of [0,1].

Through experiments, it is found that the numerical value is easier to stabilize and converge than the value return training neural network between cells. Therefore, in order to map different values to the appropriate interval, this paper sets the amplification and reduction coefficient c1, c2, c3 to adjust the size of the interval mapped to the exponential function. If the generator cost and voltage deviation are used to train the network directly, there is a large gap between the values of different objective functions, so this paper chooses the form of exponential function to normalize it.

3.3 Constraint equation of active and reactive power coordination optimization

Because the environment of agent interaction is power grid, there are certain constraints. The constraint equations include: power balance equation of power flow. The upper and lower limits of generator active power and reactive power, the upper and lower limits of condenser reactive power, transformer transformation ratio, reactive power compensation capacity, and so on, are inequality constraints, which are also defined in the action interval.

$$\begin{cases} P_i = v_i \sum_{j=1}^{n} v_j (G_{ij} \cos\theta_{ij} + B_{ij} \sin\theta_{ij}) \\ Q_i = v_i \sum_{j=1}^{n} v_j (G_{ij} \sin\theta_{ij} - B_{ij} \cos\theta_{ij}) \\ P_{G_m \min} \le P_{G_m \max} \le P_{G_m \max} \\ Q_{G_m \min} \le Q_{G_m} \le Q_{G_m \max} \\ k_{T_m \min} \le k_{T_m} \le k_{T_m \max} \\ Q_{C_m \min} \le Q_{C_m} \le Q_{C_m \max} \end{cases} \tag{2}$$

Where, *i,j* is the node number; *n*, is the number of nodes;*Pi*, *Qi* is active power and reactive power injecting into the node *i*; *vi* is the voltage amplitude of the node.*θij* Is the voltage phase angle difference between nodes *i* and *j*;*Gij* and *Bij* are the elements in the admittance matrix; P_{Gm} is the active power output of the generator *m*; Q_{Gm} is the reactive power output of the generator *m*; k_{Tm} is the selection of tap on load tap changer *m*; Q_{Cm} is the selection of capacity of capacitor *m*.

3.4 Network structure design and hyperparametric optimization

In addition to the conventional actor network and Critic network, this paper also sets up σ network to control the variance of action during exploration. After a large number of experiments and relevant scene references [18], the parameter is set as: the network input of execution action is N×1 matrix, the number of hidden layers is 2, the number of neurons is 1280 and 256 respectively, and the number of neurons in the output layer is 256. The activation function uses tanh uses mean square error (MSE), and the optimization algorithm is RMSprop. The evaluation network consists of s observed by the agent and a taking corresponding action. The number of hidden layers is the same as that of the execution network. The number of neurons is 1280 and 256 respectively. After adding, 128 neurons are connected. The number of neurons in the output layer is 1, which is the Q value obtained by the evaluation network taking a certain a under this s. The activation function of the output layer is relu, and the other activation functions use tanh. The network structure diagram is shown in Fig. 2.

When transfer reinforcement learning is used to train different scenarios, the "common knowledge" of active and reactive power integrated scheduling in historical scenarios is stored in the neural network. In order to transfer the knowledge learned from the source problem to the expansion problem, this paper fixes the parameters of the hidden layer in the original Actor and Crtic network except the last layer, and only updates the gradient of the last hidden layer during training. The reason for this is that the underlying network of neural network often extracts common features, and the knowledge differences with different characteristics are often reflected in the last layer of neural network. Appendix A gives some super parameters that usually need to be set in transfer reinforcement learning network training. The meaning of super parameters is shown in Appendix A1.

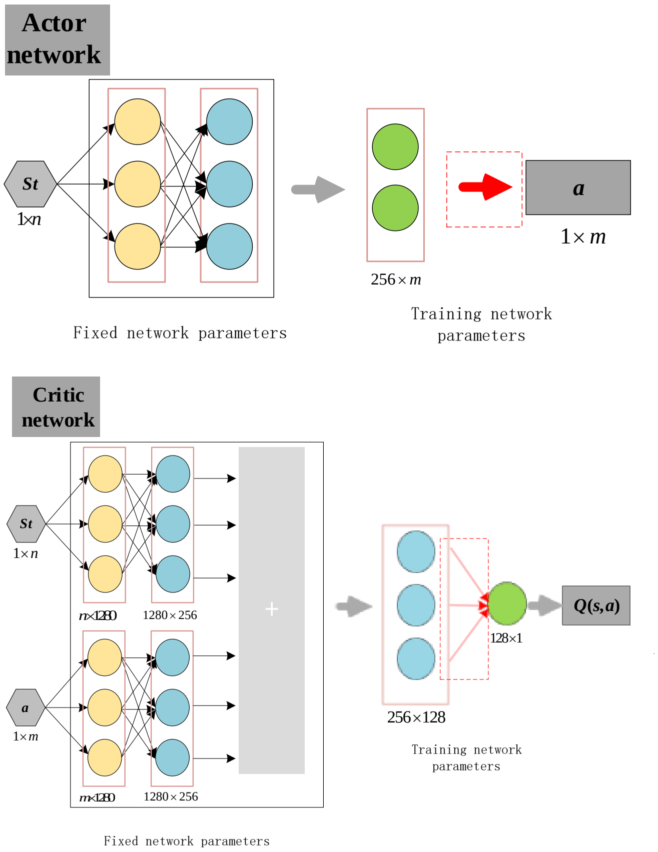

Fig. 2. Schematic diagram of Actor-Critic network structure

4. Case study

For the problem of active and reactive power integrated dispatching, the division of source problem and expansion problem is very important.In this example IEEE-14 node system. The computer configuration used in this experiment is: CPU: Intel i7-7700hq;GPU:NVDIA GTX 1060Ti;Memory: 8G DDR4;python 3.6; Tensorflow 1.7.1; Tensorboard 1.7.0;pytorch 1.2.0;pypower 5.1.4. The electrical wiring diagram of IEEE-14 node is shown in Fig. 3. IEEE-14 node system consists of 5 generators, 11 loads, 17 lines and 3 transformers. Node 1 is a balance node, and nodes 3, 6 and 8 are dimmers that only provide reactive power.The consumption characteristic curve of nodes 1 and 2 is a quadratic function curve, and the parameters are shown in the appendix. Continuous reactive power compensation equipment is set on nodes 1, 2, 3, 6, 8 and 14, and its value range is shown in Table 1. Refer to the objective function in expression (1), in this paper $k_1 = 2$, $k_2 = k_3 = 0.5$; $a = 2$, $b = 5$; $c_1 = 8000$, $c_2 = c_3 = 100$. The dimensions and descriptions of State and Action are shown in Table 2, and some parameter settings in DDPG are shown in appendix. In order to verify the effectiveness and accuracy of the proposed method, in the first experiment, the results of traditional DDPG algorithm and various heuristic algorithms under 96 loads on the test day are studied. In the second study, the computational effect and experimental time between DDPG algorithm and DDPG migration algorithm are compared.

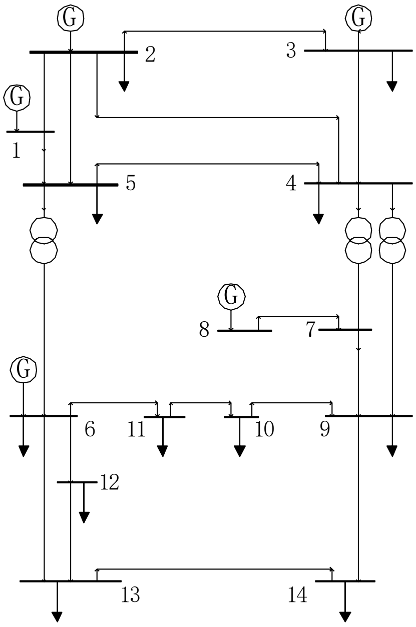

Fig. 3. IEEE-14 node system electrical diagram

4.1 An example of using DDPG method to solve active and reactive power integrated dispatching

As shown in Fig. 4, a total of 2880 sets of data of 96 points per day for 30 days in a month are simulated as source problems for training, in which the maximum value of total load is no more than 220MW and the minimum value is no less than 160MW. 96 points per day with great differences are used as migration training and expansion problems for testing during training. As shown in Fig. 5, the maximum value of total load is 252MW and the minimum value is 157MW. Fig. 6 shows the load curve of each node on the test day.

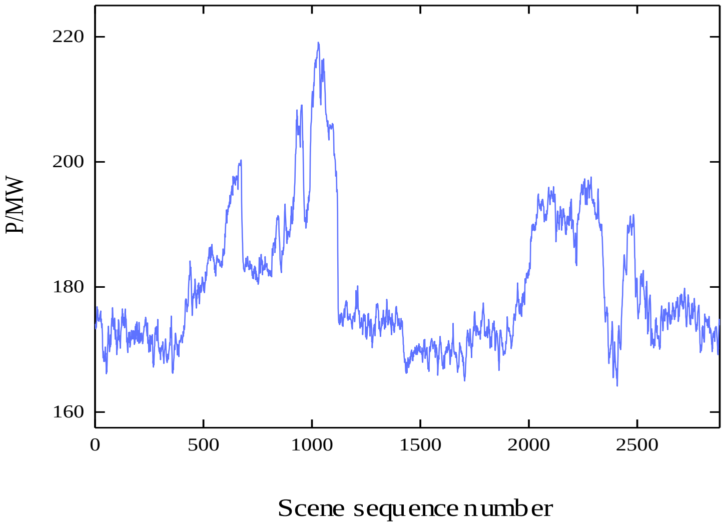

Fig. 4. Total load curve for source problem

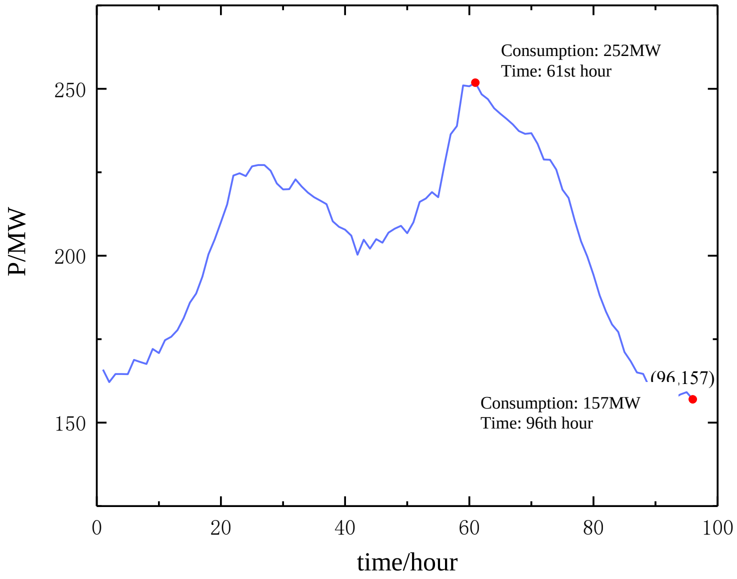

Fig. 5. Test day total load curve

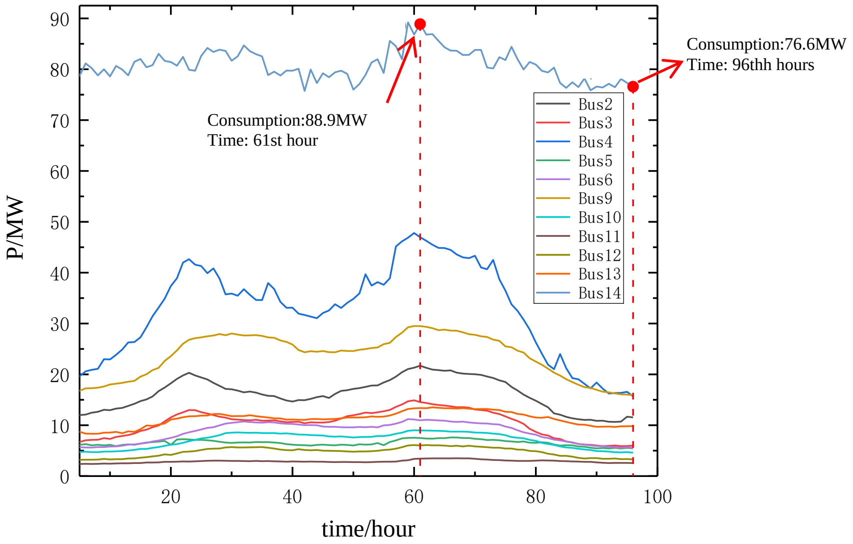

Fig. 6. Load curve of each bus on the test day

Table 1

Operating range of reactive power compensation equipment

Node number	Value range (B)	Node number	Value range (B)
1	[-10,10]	6	[0,25]
2	[0,20]	13	[0,60]
3	[0,20]	14	[0,100]

Firstly, the traditional DDPG algorithm is used to train on the historical data set. A total of about 1000 rounds of training are carried out. There are 200 iterations in each round, which is actually equivalent to 2 × 105 iterations. The setting of super parameters is shown in Appendix B2. The loss function of the training process is shown in Fig. 7, which shows the last training loss value of each round. The loss value of actor is the opposite number in the figure.

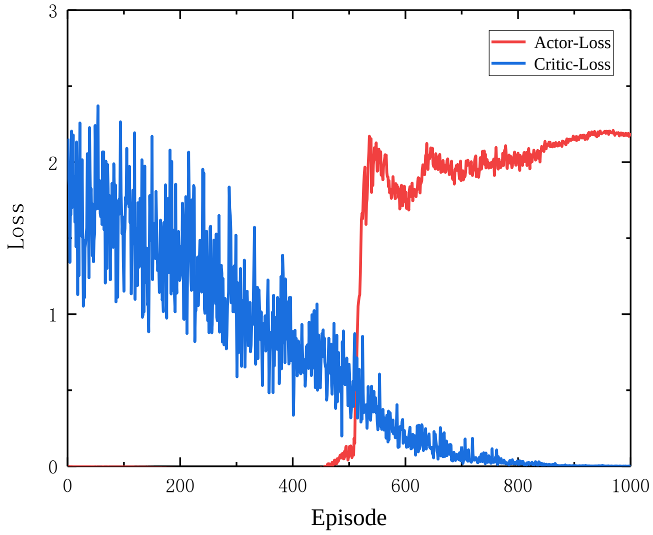

Fig. 7. Actor-Critic Loss during source problem training

Input 96 scenarios on the test day into the trained DDPG model to obtain the generator cost curve and voltage curve under each scenario, as shown in Figs. 8 and 9.

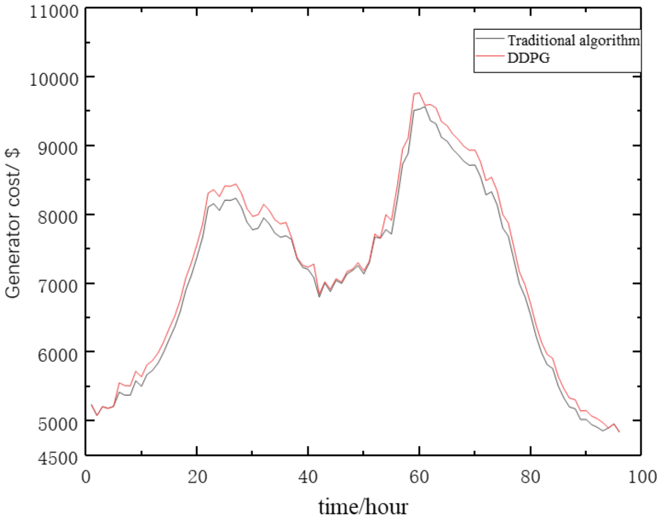

Fig. 8. DDPG algorithm-based generator cost at each bus in each time period under the test day

Table 2

State and action design

	Dimension	Detailed information
State	27	$[P_{G1},P_{G2},V_1,V_2,V_3,V_4,V_5,T_1,T_2,T_3$ $,B_1,B_2,B_3,B_4,B_5,B_6,$ $P_{L1}, P_{L2}, P_{L3}, P_{L4}, P_{L5}, P_{L6},P_{L7}, P_{L8}, P_{L9}, P_{L10}, P_{L11}]$
Action	15	$[\Delta P_{G2},\Delta V_1,\Delta V_2,\Delta V_3,\Delta V_4,\Delta V_5,$ $\Delta T_1,\Delta T_2,\Delta T_3,\Delta B_1,\Delta B_2,\Delta B_3,\Delta B_4,\Delta B_5,\Delta B_6]$

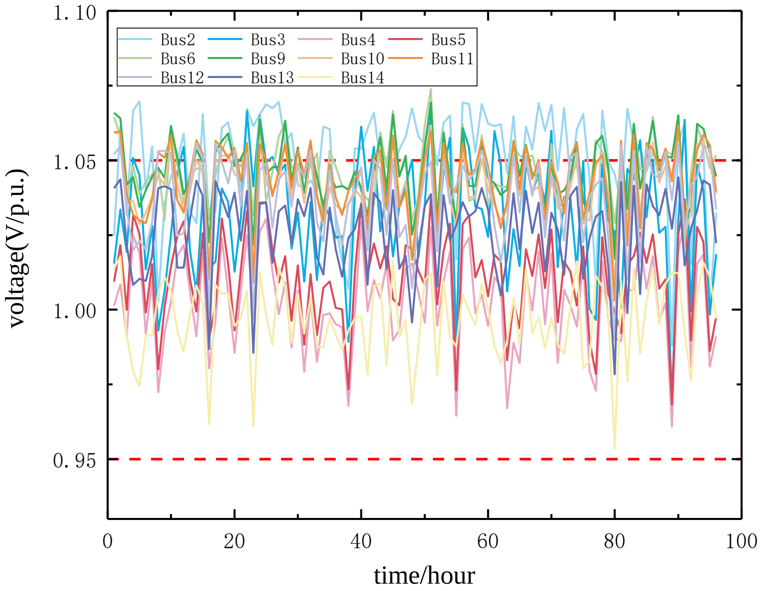

Fig. 9. DDPG algorithm-based voltages at each bus in each time period under the test day

Table 3

Error of DDPG algorithm

Error name	Before migration
Maximum error	239.75$/h
Maximum error rate	2.5154%
Average error	166.7445$/h
Average error rate	2.32%

4.2 An example of using DDPG migration method to solve active and reactive power integrated dispatching

Based on the trained DDPG model in example 1, the migrated DDPG algorithm is used in example 2 to explore and train the knowledge of integrated active and reactive power scheduling under 96 loads in the test day. A total of about 500 trainings are carried out for 20 iterations each time. The super parameters of the training are shown in Appendix, and the loss function of the training is shown in Fig. 10. Fig. 10 shows the last training loss value of each round. Similarly, the loss value of Actor is the opposite number in the figure.

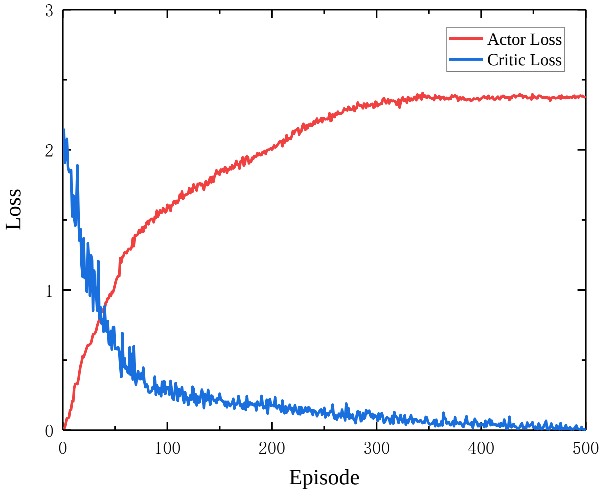

Fig. 10. Actor-Critic Loss under Transfer Reinforcement Learning

This paper compares the optimal solutions of the other methods in three different cases. The comparison of errors before and after transfer learning is shown in Table 4.The generator cost under 96 loads before and after migration is shown in Fig. 11, and the voltage level under 96 loads after migration is shown in Fig. 12.

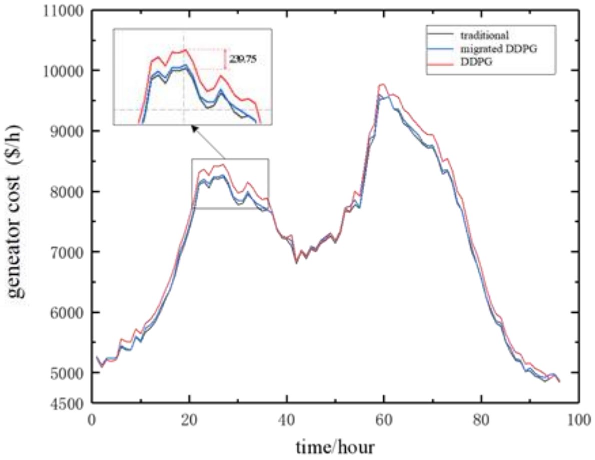

Fig. 11. Comparison of generator costs under different scenarios

Table 4

Comparison of errors before and after transfer learning

Error name	Before migration	After migration
Maximum error	239.75$/h	98.92$/h
Maximum error rate	2.5154%	1.1325%
Average error	166.7445$/h	38.4160$/h
Average error rate	2.32%	0.5576%

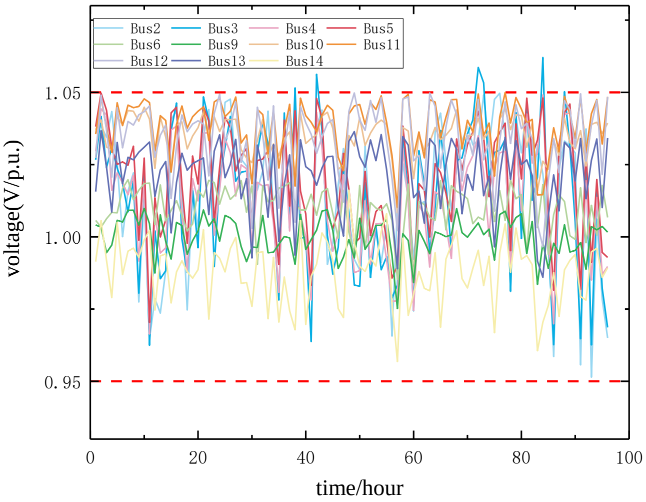

Fig. 12. Voltage diagrams for each time period at each bus after Transfer Reinforcement Learning

The comparison of cumulative time consumption between migration DDPG algorithm and genetic algorithm (GA), particle swarm optimization (PSO) and biogeographic optimization algorithm (BBO) in 96 comparative experiments is shown in Fig. 13.

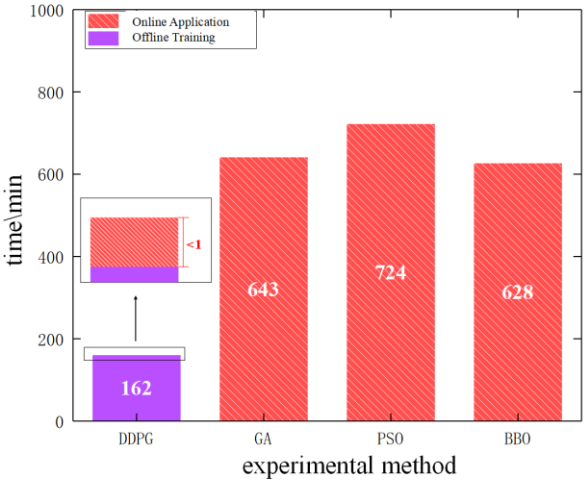

Fig. 13. Comparison of calculation time required for different method experiments

The migration DDPG algorithm needs 12 minutes and 31 seconds for offline training under 96 loads, but it has obvious advantages over other comparison algorithms. When the migration DDPG algorithm is applied online, the calculation time is less than 1 s. It can be seen that heuristic algorithms such as genetic algorithm and particle swarm optimization algorithm do not have off-line training time, but they need to be recalculated to deal with different situations each time, which takes a lot of time to apply.

Because the parameters of the shallow neural network are fixed, the unnecessary exploration range is greatly reduced when re exploring different load levels. It can be considered that the migration training gives a closer initial solution to the expansion problem on the basis of the source problem, so on the one hand, the convergence is guaranteed, on the other hand, the calculation time is shortened. In addition, the results of transfer learning are improved regardless of generator cost or voltage amplitude. This experiment has good results, and provides a new idea and research method for active and reactive power integrated scheduling with a large number of samples.

5. Conclusion

Aiming at the problem of active and reactive power coordination and Optimization in power system, based on the deep transfer reinforcement learning algorithm, this paper proposes a scheme architecture that can be quickly applied online. Because this method can solve high-dimensional, continuous and nonlinear problems, it is a new idea and a new test in the field of active and reactive power coordination optimization. Artificial intelligence and reinforcement learning algorithms accumulate experience on the basis of experiments. After a period of training, the neural network makes the parameters of random initialization converge to the feasible region. The requirements of this method for hardware computing power become more stringent with the increase of problem dimension. Therefore, it has always been people's desire to speed up the network convergence speed, reduce the training time and improve the operation efficiency of intelligent algorithms. This paper provides an integrated scheduling and migration model of active and reactive power, which provides a new idea and mode for training in multi load scenario.

REFERENCES

[1]. *S. X. Fan, L. X. Li, S. Y. Wang, et al.,* Application analysis and exploration of artificial intelligence technology in power grid dispatch and control. Power System Technology, 2020, 44(2):401-411. (in Chinese).

[2]. *H. R. Guo,* Current situation and technology of smart grid dispatching control system. Electrical age, 2021, (08):65-68.

[3]. *W. Wei, E. Wang, Z. Liu,* Real-Time Probabilistic Model Based Risk Assessment of Dispatching Operations Process. Power System Technology, 2013, 37(12):3509-3514. (in Chinese).

[4]. *X. Shan, X. Lu, M. Y. Zhai, et al.,* Analysis of Key Technologies for Artificial Intelligence Applied to Power Grid Dispatch and Control. Automation of Electric Power Systems, 2019, 43(01):49-57. (in Chinese).

[5]. *M. H. Lin, Z. F. Hou, C. Wang, M. J. Liu, T. Xu, Y. Guo,* Coordinated optimal dispatching of transmission and distribution based on distributed interior point method. Renewable energy, 2020, 38(09):1226-1232.

[6]. *Y. S. Zhong,* Large power grid intelligent regulation system framework under big data and artificial intelligence. China new communications, 2021, 23(01):63-64.

[7]. *W. Wang, L. Liu, J. Liu and Z. Chen*, Energy management and optimization of vehicle-to-grid systems for wind power integration, in CSEE Journal of Power and Energy Systems, vol. 7, no. 1, pp. 172-180, Jan. 2021.

[8]. *Layth AL-BAHRANI, Murtadha AL-KAABI, Mohammed AL-SAADI,* Optimal power flow based on differential evolution optimization technique. UPB Scientific Bulletin, Series C: Electrical Engineering and Computer Science; Volume 82, Issue 1, 2020, Pages 247-258.

[9]. *Layth AL-BAHRANI, Virgil DUMBRAVA*, Optimal power flow based on particle swarm optimization. UPB Scientific Bulletin, Series C: Electrical Engineering, Vol. 78, Iss. 3, 2016, pp. 253-264, ISSN 2286-3540.

[10]. *H. Wang, Z. Lei, X. Zhang, J. Peng, H. Jiang,* Multiobjective Reinforcement Learning-Based Intelligent Approach for Optimization of Activation Rules in Automatic Generation Control, in IEEE Access, vol. 7, pp. 17480-17492, 2019.

[11]. *K. Y. Wang, X. J. Luo, R. Jia, C. W. Zhou,* Coordinated short-term optimal dispatching method of wind storage, water and fire by giving full play to the complementary role of multiple energy. Power grid technology, 2020, 44(10):3631-3641.

[12]. *K. W. Wang, D. Y. Zhang,* A summary of reactive power optimization algorithm in power system. Electrical Measurement & Instrumentation, 2016, 53(10):73-79+98. (in Chinese).

[13]. *Y. Li, H. Tang, K. Lv, K. Wang,* Optimization of Dynamic Dispatch for Multiarea Integrated Energy System Based on Hierarchical Learning Method. IEEE Access, vol. 8, pp. 72485-72497, 2020.

[14]. *Y. Ai, M. Du, Z. Pan and G. Li,* The optimization of reactive power for distribution network with PV generation based on NSGA-III. CPSS Transactions on Power Electronics and Applications, vol. 6, no. 3, pp. 193-200, Sept. 2021.

[15]. *H. B. Zhai, H. Z. Cheng, G. Y. Lv, et al.,* Transmission Network Planning Based on Schema Recording Parallel Ant Colony Algorithm. Proceedings of the CSEE, 2005(09):17-22. (in Chinese).

[16]. *M. Basu,* Modified particle swarm optimization for nonconvex economic dispatch problems. International Journal of Electrical Power and Energy Systems, 2015, 69.

[17]. *Y. Ai, M. Du, Z. Pan and G. Li,* The optimization of reactive power for distribution network with PV generation based on NSGA-III, in CPSS Transactions on Power Electronics and Applications, vol. 6, no. 3, pp. 193-200, Sept. 2021.

[18]. *B. Yang, Z. Y. Chen, W. J. Ruan, et al.,* Demand response dispatching strategy for intelligent community based on multi-group co-evolution gen. Power Demand Side Management, 2019, 21(2):10-14. (in Chinese).

[19]. *B. Ammar, A. M. Mohd, S. S. Izwan, A. A. S. Ghassan,* Q-Learning vertical handover scheme in two-tier LTE-A networks. International Journal of Electrical and Computer Engineering (IJECE), 2020, 10(6).

[20]. *S. Wang et al.,* A Data-Driven Multi-Agent Autonomous Voltage Control Framework Using Deep Reinforcement Learning, in IEEE Transactions on Power Systems, vol. 35, no. 6, pp. 4644-4654, Nov. 2020.

[21]. *G. Yang, F. Zhang, C. Gong, et al.,* Application of a deep deterministic policy gradient algorithm for energy-aimed timetable rescheduling problem. Energies, 2019, 12(18):3461.

[22]. *J. Duan et al.,* Deep-Reinforcement-Learning-Based Autonomous Voltage Control for Power Grid Operations, in IEEE Transactions on Power Systems, vol. 35, no. 1, pp. 814-817, Jan. 2020.

[23]. *Y. Zhou et al.,* A Data-driven Method for Fast AC Optimal Power Flow Solutions via Deep Reinforcement Learning, in Journal of Modern Power Systems and Clean Energy, vol. 8, no. 6, pp. 1128-1139, November 2020.

[24]. *Q. Li, Y. Qiao, Y. J. Zhang,* Continuous Reactive Power Optimization of Distribution Network Using Deep Reinforcement Learning. Power System Technology, 2020, 44(04):1473-1480. (in Chinese).

[25]. *J. Li, T. Yu, H. Zhu, F. Li, D. Lin, Z. Li,* Multi-Agent Deep Reinforcement Learning for Sectional AGC Dispatch, in IEEE Access, vol. 8, pp. 158067-158081, 2020.

[26]. *R. R. Hossain, Q. Huang and R. Huang,* Graph Convolutional Network-Based Topology Embedded Deep Reinforcement Learning for Voltage Stability Control, in IEEE Transactions on Power Systems, vol. 36, no. 5, pp. 4848-4851, Sept. 2021.

[27]. *Y. Guo and H. Gao,* Data-Driven Online System Equivalent for Self-Adaptive Droop Voltage Control of Wind Power Plants. IEEE Transactions on Energy Conversion, vol. 35, no. 1, pp. 302-305, March 2020.

[28]. *L. Xi, L. Yu, Y. M. Fu, et al.,* Automatic Generation Control Based on Deep Reinforcement Learning with Exploration Awareness. Proceedings of the CSEE, 2019, 39(14):4150-4162. (in Chinese).

[29]. *Y. X. Chen, X. S. Zhang, L. X. Guo, et al.,* Optimal Carbon-energy Combined Flow in Power System Based on Multi-agent Transfer Reinforcement Learning, 2019, 45(03):863-872. (in Chinese).

APPENDIX A

Table A1

Improve the hyperparameters and descriptions of DDPG algorithm settings

Parameter	Meaning description
Tau	Smoothing coefficients of target networks in Actor and Critic networks
α	Learning rate of evaluation network in Actor network and Critic network
Batch_size	Number of extracts from the experience pool per training

Capacity	Size of experience pool
σ	Controls the initial value of the variance of the exploration range
γ	Discount factor
Episode	Rounds of agent training
Step	One iteration in each round of agent training

APPENDIX B

Table B1

Related definitions of DDPGC

Node number	a	b	c
1	3	0.043	20
2	3	0.25	20

Table B2

DDPG hyperparameter settings

Parameter	Figure
Tau	0.000001
α	0.000001
Batch_size	64
Capacity	100000
σ	2

Table B3

Hyperparameter setting of DDPG migration method

Parameter	Figure
Tau	0.0001
α	0.0001
Batch_size	16
Capacity	1000
σ	1
γ	0.95