International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1–20

https://doi.org/10.70917/ijcisim-2026-0230
Article

Deep learning-supported multi-objective power system scheduling in an energy internet optimal scheduling environment

Qian Liu*

* School of Electronic and Control Engineering, North China Institute of Aerospace Engineering, Langfang, Hebei, 065000, China; liuqian220314@163.com

Abstract: In the context of energy internet, new energy access makes the power system present multi-objective coupling, difficult security constraints, etc., and the traditional model-driven methods have limitations in scheduling efficiency and adaptability. Based on this, this paper proposes a PPO-DDPG model that combines deep reinforcement learning and proximal policy optimization algorithm. The model acquires the dynamic timing data of the power system with deep reinforcement learning and designs constraints on its states, actions, rewards and so on. The PPO algorithm is then introduced to update the network parameters as a way to improve the dynamic scheduling effect on multi-objective power systems. Experiments show that the new energy consumption rate of the PPO-DDPG algorithm after convergence is as high as about 97.5%, which is about 30% higher than that of the PPO algorithm, and the average survival time as well as the reward value are significantly better than that of the existing methods. Therefore, relying on deep learning technology to empower multi-objective power system scheduling is more economical, ensuring that the power system can obtain optimal economic benefits when accessing different types of energy under high uncertainty.

Keywords: deep reinforcement learning; proximal policy optimization algorithm; PPO-DDPG model; multi-objective scheduling; power system

1. Introduction

Energy Internet is an important pillar of the third industrial revolution, with low carbon and clean, safe and efficient as the core, presenting the main characteristics of renewable, distributed, linked up, openness and integration, and all countries are actively promoting the energy Internet strategy [1-3]. Energy Internet applies advanced Internet technology to the energy field, thus realizing an effective mode of distributed energy supply. Compared with a single energy system, the energy Internet can improve the system elasticity, reduce the dependence of a specific energy source and the risk of energy supply; with more scheduling options, it can flexibly maintain a higher energy utilization efficiency; in the form of form, geography, and other factors under the influence of the difference in the price of energy, the energy Internet can also make full use of the form of multi-energy flow to reduce the operating costs [4-8].

And increase the proportion of installed renewable energy power generation in the power system is an effective measure to reduce greenhouse gas emissions and achieve the goal of “carbon peak, carbon neutral” [9-10]. At the same time, wind power, solar and other high proportion of renewable energy power generation access to the grid, the randomness and volatility of its output, increasing the difficulty and complexity of power balance and peak shifting, and increasing the system operating costs [11-13]. Voltage and frequency fluctuations brought about by large-scale renewable energy generation connected to the grid also bring risks to the safe and stable operation of the power grid [14-15]. Promoting the power system scheduling strategy aiming at economy-security-green, realizing the operation mode of wind, light, water, coal and other resources synergizing and complementing, and the flexible interaction of source, grid, load, and storage, has become the intrinsic requirement and

inevitable result of the development of the power system [16-18].

With the deepening of power market reform and the expanding scale of new energy grid integration, multi-objective power system scheduling faces unprecedented challenges. Traditional power system scheduling methods take classical mathematical optimization algorithms such as linear, nonlinear and dynamic programming as the main means. Literature [19] created a risk-constrained economic dispatch model and introduced Pareto-based multi-objective optimization and optimization algorithms based on chaotic swarm search optimizer to achieve the economic dispatch of power system with wind power generation. Literature [20] established a multi-objective multi-population ant colony optimization algorithm for continuous domains, which obtained more accurate Pareto optimal solutions (POs) on ecological niche search methods with Gaussian functions, and therefore presented effects in the environmental economic dispatch (EED) of power systems with security considerations. Literature [21] for the multi-objective EED scheduling of power systems considering integrated natural gas units and variable renewable energy sources, combines two multi-objective optimization algorithms and ideal solution similarity ranking to obtain the optimal PO for the total fuel or power generation cost and pollution emission. Literature [22] employs a multi-layer distributed multi-objective consensus algorithm to efficiently achieve large-scale multi economic dispatch of regional interconnected power systems. Literature [23] proposed a joint dynamic scheduling scheme for power system EED for wind power and wind generation based on an enhanced multi-objective differential evolutionary algorithm to optimize the pollutant emissions and economic costs. Literature [24] evaluated the performance of the multi-objective Salve swarm algorithm in obtaining the optimal solution for power system EED, and the test results of the 6, 10, and 14-unit power systems at the IEEE standard nodes showed that this algorithm outperforms other algorithms. Literature [25] designed a multi-objective power system scheduling method by establishing a spatio-temporal distribution model of pollutants in thermal power plants and a high-dimensional multi-objective optimization model, and introducing a high-dimensional multi-objective optimization algorithm, which reduces the cost and carbon dioxide emissions, and balances the economic and environmental factors effectively. Literature [26] proposed a collaborative constrained multi-objective bi-population evolutionary algorithm based on a non-dominated sequential genetic algorithm in optimal scheduling of integrated power systems for wind power generation, and the bi-population is used to coordinate the POs. Literature [27] provided a multi-objective human learning optimization algorithm for EED of the power system, and the algorithm optimizes the convergence and diversity of the POs in a congestion-distance metric based approach, combining two mechanisms to eliminate the weak solutions to obtain a high quality scheduling solution. Literature [28] proposes a multi-objective group search optimizer which captures important features through a self-learning method of solving spatial features and focuses the search on regions to improve the computational efficiency, which improves high-quality solutions for uncertain power system scheduling.

Although these algorithms have matured in both theoretical foundation and engineering practice, however, their inherent defects are increasingly highlighted in the face of the complexity and variability of modern power systems. The traditional deterministic model is difficult to accurately portray the stochastic fluctuation characteristics of renewable energy generation, and it is also unable to effectively respond to the dynamic change law of load demand [29-30]. In large-scale grid scheduling problems, the number of state variables and control variables grows exponentially, leading to a sharp rise in computational complexity and a serious lack of real-time optimization capability [31-33]. At the same time, the characteristics of equipment start and stop discrete variables and nonlinear constraints of network currents, which are prevalent in the power system, make it difficult to meet the strict requirements of traditional algorithms on the continuous derivability of the objective function, and the accuracy and practicality of the model are limited [34-36]. How to build a multi-objective, intelligent and adaptive scheduling optimization technology system to improve the operational efficiency and reliability of the power system has become a key problem to be solved.

The deep learning technology shows a strong potential in multi-objective power system scheduling by analyzing the historical data of the power system, fusing the data related to the operation of the multi-modal power system, exploring the association between the data, enhancing the system state perception, and realizing the system multi-objective optimization and intelligent decision-making. For example, literature [37] shows that the combination of big data mining techniques, convolutional neural networks, and gated loop units can realize the fusion, feature extraction, and dependency capturing of heterogeneous power data from multiple sources to promote the optimal operation of power systems. Literature [38] learns daily photovoltaic power generation patterns under different weather conditions through deep learning autocoders, explores power generation features, and identifies these patterns in conjunction with the cluster centroid method algorithm, which contributes to optimal energy management and efficient demand-based scheduling.

In the practical application of deep learning technology-enabled power system scheduling, researchers have contributed a variety of scheduling methods. Literature [39] uses generative adversarial networks to output a large amount of virtual power plant scenario data, and introduces deep reinforcement learning (DRL) to analyze these data and historical data to obtain an economic dispatch scheme with robustness. Literature [40], in order to perform dynamic scheduling for a zero-carbon hydropower-PV-pumped storage integrated power system, utilizes information entropy to determine the weights of each scheduling objective, and combines a deep reinforcement learning framework to realize an online dynamic scheduling strategy with a continuous action space. Literature [41] used deep neural network-supported hierarchical learning optimization method for multi-region grid scheduling based on historical data online processing, combined with DRL to optimize large-scale grid scheduling problems. Literature [42] considered the uncertainty of renewable energy sources and improved a real-time security constraint method for economic dispatch of power systems, using a data-driven risk-aware assessment based on DRL and a real-time scheduling strategy architecture based on residual networks to realize real-time scheduling in a secure and economic manner. Literature [43] combined support vector clustering and deep Q-networks with day-ahead scheduling results and real-time renewable energy forecasts as the intermittent and uncertain factors of dynamic scheduling of integrated renewable energy EEDs, and considered the prediction error and penalty cost of energy use, to construct a robust optimization method, which successfully reduces the cost and optimizes the carbon emissions. Literature [44] improved the non-dominated sorted moth optimization algorithm by double deep neural networks and used the algorithm in distributed multi-objective economic dispatch of power systems to improve the computational speed. Literature [45] integrated deep neural network with African vulture optimizer and used a hierarchical distributed approach for economic scheduling of power system with IoT, which improved the scheduling efficiency and saved the cost. Literature [46] used IoT, deep learning, long and short term memory networks and multilayer perceptron to design a supply-demand co-optimization model and a two-way power dispatch optimization architecture for smart grids, which performs power system scheduling under real-time monitoring and dynamic adjustment of reference parameters such as electricity supply demand, generator response efficiency, and electricity price changes. Literature [47] used DRL to construct a two-stage low-carbon economic dispatch model in the context of energy internet, i.e., the output stage of the day-ahead dispatch strategy and the stage of power transmission security limit constraints adjustment, thus realizing the triple objective of security-carbon emission-economy.

Aiming at the uncertainty, security constraints and other problems existing in the current power system scheduling process, this paper proposes to combine deep reinforcement learning with PPO algorithm to establish PPO-DDPG model for multi-objective power system scheduling optimization. The experimental results show that the results obtained by this method in different scheduling scenarios and scheduling strategies are optimal, and the overall algorithm control performance is good, and the scheduling results are more in line with the actual needs of the power system. This study provides an intelligent decision-making solution for new energy power systems with both theoretical rigor and engineering practical value, and shows significant advantages in training stability, multi-objective co-optimization efficiency and power system demand satisfaction rate.

2. Research methodology design

With the development of deep peaking units, flexible loads and other power resources with elastic regulation capability, the flexibility of grid scheduling and operation has been effectively improved, but at the same time, it also increases the complexity of scheduling optimization tasks. Data-driven deep reinforcement learning algorithms have good results in optimizing such problems. However, deep reinforcement learning algorithms usually learn to optimize for specific scheduling tasks, and when the power system changes, the scheduling knowledge learned by the intelligences in the historical tasks is often no longer applicable. Therefore, exploring multi-objective power system scheduling methods in the optimal scheduling environment of the energy Internet has become a must for blocking walls to further enhance the stability of power system scheduling.

2.1. Scheduling model for multi-objective power systems

2.1.1. Objective function design

In the context of the synergistic development of the energy Internet, the power system is constantly accessing many types of new energy power systems, and access to new energy sources will, to a certain extent, increase the load of the power system, making it possible to cause the collapse of the power

system during scheduling [48]. Based on this, this paper establishes the scheduling objective function of the multi-objective power system from the low-carbon economy scheduling objective of minimizing the total cost of the power system as:

$$\min F = \min\left(F_W + F_G + F_{ESS} + F_{co_2}\right) \tag{1}$$

Where F_W is the wind power cost, F_G is the generation cost of conventional thermal power units, F_{ESS} is the operation and maintenance cost of energy storage, and F_{co_2} is the cost of carbon trading.

(1) Wind power cost consists of wind turbine power generation cost and wind abandonment penalty cost. The wind power cost F_W can be expressed as:

$$F_W = F_{W_1} + F_{W_2} \tag{2}$$

Wind turbine power generation cost. The cost of generating electricity during wind power operation refers to the operating costs and investment costs of the wind farm during its entire life cycle discounted to the unit cost of generating electricity, which can be approximated as a linear relationship with the amount of electricity generated, i.e.:

$$F_{W_1} = \sum_{t=1}^{T}\sum_{i=1}^{M} C_1 P_{W,i,t} \tag{3}$$

Where C_1 is the wind turbine generation cost coefficient, and $P_{W,i,t}$ is the actual power generation of the i th wind turbine in unit time t.

Wind abandonment penalty cost. In order to maximize the use of wind power to improve the low carbon system, this paper introduces the wind abandonment penalty cost to limit the wind abandonment behavior, wind abandonment penalty cost can be expressed as:

$$F_{W_2} = \sum_{t=1}^{T}\sum_{i=1}^{M} C_2\left(P_{W,i,t} - P_{S,i,t}\right) \tag{4}$$

Where C_2 is the wind abandonment penalty coefficient, and $P_{S,i,t}$ is the predicted power of the i th wind turbine in unit time t. It should be noted that, in the wind power overestimation, the main task of scheduling is to make up for the power gap by increasing the output of conventional thermal power units, low carbon economy is no longer the main goal of scheduling, so this paper does not consider the wind power overestimation penalty.

(2) The generation cost F_{G} of conventional thermal power units can be approximated as a quadratic function. Then:

$$F_G = \sum_{t=1}^{T}\sum_{j=1}^{N}\left(a_j P_{G,j,t}^2 + b_j P_{G,j,t} + c_j\right) \tag{5}$$

Where a_j, b_j, c_j for the j conventional thermal power unit generation cost coefficient.

(3) The volatility of wind power generation characteristics of the grid can not maintain transmission balance, must be configured with a certain scale of energy storage as a flexible regulation resources, dynamic compensation for wind power generation intermittent, fluctuating characteristics. Energy storage operation and maintenance cost F_{ESS} is:

$$F_{ESS} = \sum_{t=1}^{T}\sum_{k=1}^{K} C_3\left|P_{es,k,t}\right| \tag{6}$$

Where K is the number of energy storage, C_3 is the cost of energy storage operation, $P_{es,k,t}$ is the charging and discharging power of the k th energy storage unit in unit time t, with positive values indicating the charging power and negative values indicating the discharging power.

2.1.2. Constraint establishment

For the constraints of the multi-objective power system dispatch model, they mainly include market clearing constraints, system power balance constraints, unit output constraints, conventional unit climbing constraints, wind turbine operation constraints, line transmission constraints, and positive and negative rotating standby constraints of the system. Specifically as follows:

(1) Market clearing constraints, i.e:

$$\sum_{i=1}^{N} u_i(t)P_i(t) + P_w(t) + P_p(t) = P_L(t) \tag{7}$$

Where $u_i(t)$ is the startup and shutdown state of thermal power unit i at t moment, $P_i(t)$ is the output of thermal power unit i at t moment, $P_w(t)$ and $P_p(t)$ are the wind power and PV scheduling output at t moment in the previous day's plan, and $P_L(t)$ is the previous day's load forecast value.

(2) System power balance constraints, that is:

$$\sum_{i=0}^{N}\left(P_{gi} + P_{wi} - P_{Li}\right) = 0 \tag{8}$$

where N is the total number of nodes in the transmission grid, P_{gi} and P_{wi} are the scheduling plans for node i conventional units g and wind farms w, respectively, both of which are decision variables, and P_{Li} is the system load at node i.

(3) Unit output constraints, i.e.:

$$P_i^{\min} \le P_i(t) \le P_i^{\max} \tag{9}$$

$$0 \le P_{w,t} \le P_{wf,t} \tag{10}$$

$$0 \le P_{p,t} \le P_{pf,t} \tag{11}$$

Where $P_i^{\min}$, $P_i^{\max}$ are the upper and lower limits of thermal power unit output, $P_{wf,t}$, $P_{pf,t}$ are the predicted values of wind and PV outputs at the moment of t, respectively.

(4) Conventional unit climbing constraint, that is:

$$\begin{cases} P_{i,t} - P_{i,t-1} \le U_{i,t-1}R_{u,i} + (1-U_{i,t-1})P_i^{\min} \\ P_{i,t-1} - P_{i,t} \le U_{i,t}R_{d,i} + (1-U_{i,t})P_i^{\max} \end{cases} \tag{12}$$

where $R_{u,i}$ and $R_{d,i}$ are the climbing rate and the slippage rate of thermal power unit i, respectively.

In order to further analyze the impact of the model proposed in this chapter on the dispatch results, three power market operation indicators are defined, namely, the average purchased power tariff $\overline{\rho}$ within a period of clearing cycle, the average nominal power tariff $\overline{f}$ within the same cycle, and the average market share occupied by generating companies i $\overline{\lambda_i}$, respectively:

$$\overline{\rho} = \frac{C_p}{\sum_{t=1}^{T} P_L(t)} \tag{13}$$

$$\overline{f} = \frac{F}{\sum_{t=1}^{T} P_L(t)} \tag{14}$$

$$\overline{\lambda_i} = \frac{\sum_{t=1}^{T} P_i(t)}{\sum_{t=1}^{T} P_L(t)} \tag{15}$$

where T is the number of time periods in the range of cycles under study.

(5) WTG operation constraints, i.e:

$$\Pr\left\{P_{wi} \le p_{wi}, w = 1,2,\cdots,W\right\} \ge \alpha \tag{16}$$

Where p_{wi} is the output of node i wind farm w, the output of W wind farms $p_{i1}, p_{i2}, \cdots, p_{iW}$ obey the Gumbel-Copula joint probability distribution, $\Pr\{\cdot\}$ is the probability value of the inequality, and α is the prior given confidence level.

(6) Line transmission constraints, i.e:

$$-K_l^{\max} \le \sum_{i=0}^{N} h_{li}\left(p_{gi} + p_{wi} - P_{Li}\right) \le K_l^{\max} \ \left(l = 1,2,\cdots,L\right) \tag{17}$$

where h_{li} is the sensitivity coefficient of the injected power of node i to the transmitted power of line l, $K_l^{\max}$ is the transmitted power limit of line l, and L is the total number of transmission lines.

(7) Positive and negative rotating reserve constraints of the system, i.e:

$$\begin{aligned} &\Pr\left\{\sum_{i=1}^{N}\left[P_i^{\max}(t) - P_i(t)\right] \ge U_{SRt}\right\} \ge \beta_1 \\ &\Pr\left\{\sum_{i=1}^{N}\left[P_i(t) - P_i^{\min}(t)\right] \ge D_{SRt}\right\} \ge \beta_2 \end{aligned} \qquad i \in N, t \tag{18}$$

2.2. Sequential Decision Model for Power System Dispatch

2.2.1. Deep reinforcement learning algorithms

Deep Reinforcement Learning (DRL) was proposed to deal with the “dimensionality catastrophe” problem faced by traditional reinforcement learning. Traditional reinforcement learning algorithms have two requirements for the state and action space, which cannot be too large and must be discrete. If they are continuous values, discretization methods can be used, such as sampling. However, the sampled samples are too small to perfectly match the characteristics of the original samples, and too large to be solved by traditional reinforcement learning algorithms due to the limitations of the computer's storage and computing power. Inspired by neural networks, neural networks are used in reinforcement learning to deal with large state space and large action space. Because neural networks have the ability to fit state-value functions and state-action-value functions without completely traversing the entire space. Also neural networks can handle continuous state space and action space problems [49].

Deterministic Deep Policy Gradient (DDPG) algorithm is a deep reinforcement learning algorithm capable of dealing with continuous action space which combines the DPG algorithm and the AC framework. The DDPG algorithm uses two ideas of the DQN, one is the target network and the other is the experience playback. Where the experience playback is basically identical to the DQN algorithm, the target network is a bit different. DDPG contains a total of four networks, which are Actor current network, Actor target network, Critic current network and Critic target network.

actor current network inputs the current state s and outputs a according to the policy as follows:

$$a = \pi_{\varphi}(s) + \mathrm{N} \tag{19}$$

The addition of the noise term N is more conducive to exploration.

Execute a to get a new state s' in this condition with instant reward r. Put them into the experience playback pool. actor target network inputs the next state s' from the experience playback pool and outputs the next action a'. Sample (s, a, r', s') from the experience playback pool.

CRITIC current network is used to evaluate the current Q value, i.e:

$$y = r + \gamma Q_{\theta'}(s', a) \tag{20}$$

where φ and φ' are the actor current network parameters and actor target network parameters, respectively, and θ and θ' are the criterion current network parameters and criterion target network parameters, respectively. criterion target network is used to compute the $Q_{\theta'}(s', a)$ part.

The Bellman error is as in Eq. (21), and D is the empirical playback pool that updates thecritic network.

$$\min(L(\theta)) = E_{(s,a)\sim D}[Q_{\theta}(s, a) - y]^2 \tag{21}$$

Update the parameters of the actor network, i.e:

$$J(\varphi) = E_{(s,a)\sim D}[Q_{\theta}(s, a)] \tag{22}$$

For the target network, DDPG then uses Eq. (23) to update the parameters of the ACTOR network and Eq. (24) to update the parameters of the CRITIC network. This soft update approach allows for slow updates at each step and is more stable than the approach of the DQN algorithm. Namely:

$$\varphi' \leftarrow \tau\varphi + (1-\tau)\varphi' \tag{23}$$

$$\theta' \leftarrow \tau\theta + (1-\tau)\theta' \tag{24}$$

where τ is the soft update factor.

Overall, the DDPG algorithm can not only solve the problem on the space of continuous actions, but also because of its simplicity, it can deal with complex and large-scale problems very well, and it is much faster than the optimal solution sought by the DQN algorithm.

2.2.2. Sequential Decision Modeling

Starting from the connotation of data-driven scheduling strategies, the main purpose of learning optimization is to find an optimal strategy. This strategy gives a reasonable unit plan that meets the scheduling demand according to the directly acquired or observed information, which is the basic process of scheduling, and the sequential decision model is a modeling of this process. From the basic process of scheduling operations, when the scheduling organization obtains the current system state observation, the prediction of a certain time scale in the future, and the scheduling goal, it needs to make the scheduling plan of the corresponding time scale and send it out after passing the safety calibration, and as the main body of the actual system of the regulation after the execution of the plan, the scheduling organization needs to observe the results of the execution and predict the changes that may occur in the corresponding time scale in the future again. After the execution of the plan, the scheduling organization needs to observe the execution results and predict the possible changes in the

corresponding time scale in the future, which can be used as the input information for the next scheduling plan [50].

In order to abstract the sequential decision-making process from a mathematical point of view, relevant concepts in reinforcement learning are introduced. First, when facing a t decision moment scheduling optimization problem, the available information can be constructed as a state as S_t. This state is mainly composed of the set of observations Obs_t of the system at the beginning of the t decision moment, and the set of predictions F_t of the system at the $t+\Delta T$ decision moment. Assuming that a complete scheduling task has a time scale of T, the state S_t can be written as tuple form, i.e:

$$S_t = \left(t, Obs_t, F_t\right), \qquad t \in \{0, \Delta T, 2\Delta T, \ldots T\} \tag{25}$$

The scheduling organization will give action adjustments based on the state S_t corresponding to the possible changes in the power system, i.e., A_t. Assuming that we take the unit plan that needs to be issued as a means of system action adjustment, such as the existence of N_u adjustable units in the system, the action adjustment carried out for the n th unit can be recorded as a_t^n, which in turn allows us to write the action A_t in the form of a tuple, viz:

$$A_t = \left(a_t^1, a_t^2, \ldots, a_t^{N_u}\right), \qquad t \in \{0, \Delta T, 2\Delta T, \ldots T\} \tag{26}$$

The scheduling policy is a parameterized function mapping the state S_t to the action A_t, where π is used to denote the scheduling policy, then the scheduling plan solving process can be abstractly described as:

$$S_t \Rightarrow \tilde{\pi} \Rightarrow A_t \tag{27}$$

When the scheduling organization decides the corresponding system adjustment action A_t based on the current system state S_t, the relevant constituent units of the power system will enter the execution phase after receiving the adjustment instruction, and enter the next decision-making phase and repeat the whole decision-making process after the corresponding execution time. Since the execution phase can be regarded as the process of implementing the action instructions and changing the state of the system, we can utilize the concept of environment mapping in reinforcement learning to make an abstract representation of this process. Here the environment mapping is denoted by $\tilde{E}$, then the complete sequential decision process can be represented as:

$$S_t \Rightarrow \tilde{\pi} \Rightarrow A_t \Rightarrow \tilde{E} \Rightarrow S_{t+\Delta T} \Rightarrow \tilde{\pi} \Rightarrow A_{t+\Delta T} \ldots \Rightarrow S_T \tag{28}$$

Figure 1 shows the construction process of multi-objective power system sequential decision-making model. Entering the decision-making stage at the t moment, the real-time operation of the system is firstly observed from the operating environment of the power system, and the current real-time unit output, real-time cross-section current and real-time bus load are taken as the components of the set of observed values. At the same time, the new energy forecast for the future $t+\Delta T$ moment and the bus load forecast are combined and constitute the forecast value set. The two together form the current state S_t of the decision process. The scheduling policy maps the state S_t to the action A_t, and A_t refers to the real-time scheduling plan that is sent to the relevant execution units in the operating environment of the power system. After the power system executes the corresponding action A_t, the reward function evaluates the overall operating effect of the adjustment guided by the action A_t while the system is in the state S_t, and the change of the gradient of this feedback is the main direction of updating the parameters of the scheduling strategy. When the execution phase is over, the next decision-making phase begins at the moment of $t+\Delta T$.

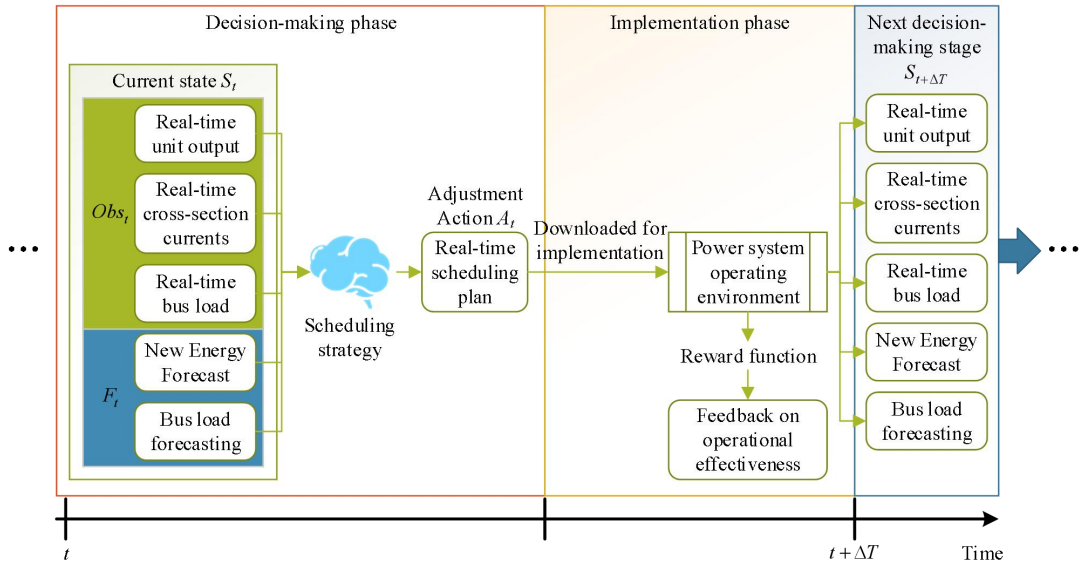


Figure 1. Sequential decision-making process

2.3. Optimized decision-making strategy design for power dispatching

2.3.1. Optimization Algorithm for Near-End Policies

The Proximal Policy Optimization (PPO) algorithm is one of the more advanced algorithms with significant advantages among all current policy gradient methods based on actor-critic structure. In recent years, it has been frequently used to handle discrete control or continuous control tasks. By interacting with the test environment, the PPO intelligences collect state, action and reward information in order to optimize the scheduling scheme of the power system for the purpose of economic scheduling of multi-objective power systems [51].

The actor network therein is mainly used to be responsible for learning the stochastic action strategy $\pi_\theta(a_t \mid s_t)$, and then selecting and executing the containerized transit operation action a_t based on the current environment state s_t at each time step t. The critic network in this case works by taking the current state characteristics s_t at each time step t as inputs, and then generating outputs and using them as estimates of the state value function $v(s)$. In the actor-critic structure, the dominance function $v(s)$ needs to be computed, which helps to reduce the variance, improve the learning efficiency and increase the stability of the learning.

The strategy needs to be run for T time steps during training to collect samples and compute the dominance function $\hat{A}_t$, which is given below:

$$\hat{A}_t = -v(s_t) + r_t + \gamma r_{t+1} + \cdots \gamma^{T-t+1} r_{T-1} + \gamma^{T-t} v(s_T) \tag{29}$$

The truncated agent goal $L_{CLIP}(\theta)$ and the entropy goal $L_E(\theta)$ are used to be in charge of updating the policy π_θ, and their respective formulas are given below:

$$L_{CLIP}(\theta) = \hat{\mathbb{E}}\left[\min\left\{r_t(\theta)\hat{A}_t, \text{clip}\left(r_t(\theta), 1-\text{蝌} + \right)\hat{A}_t\right\}\right] \tag{30}$$

$$L_E(\theta) = \hat{\mathbb{E}}\left[\text{Entropy}\left(\pi_\theta(a_t \mid s_t)\right)\right] \tag{31}$$

where $r_t(\theta)$ denotes the probability ratio between the old and new strategies, defined as $r_t(\theta) = \dfrac{\pi_\theta(a_t \mid s_t)}{\pi_{\theta old}(a_t \mid s_t)}$, $\in$ is a hyperparameter that eliminates the part of the incentive for the strategy

$r_t(\theta)$ to go beyond the interval $[1-蠑+]$. Thecritic network is trained by minimizing the mean square error (MSE) objective.

$$L_{MSE}(\theta)=\hat{\mathrm{a}}\left[\mathrm{MSE}\left(\mathrm{R}\left(a_t\mid s_t\right),\hat{v}_\phi\left(s_t\right)\right)\right] \tag{32}$$

where $\hat{v}_\phi\left(s_t\right)$ is the output of the critic network at time step t based on the current environment state s_t, and $\mathrm{R}\left(a_t\mid s_t\right)$ is the gain obtained by the actor network at time step t performing the action a_t based on the current environment state s_t to perform the action a_t. Combine the above three objectives of $L_{CLIP}(\theta),L_E(\theta)$ and $L_{MSE}(\theta)$ in order to obtain the loss function and (approximate) minimize this loss function at each iteration, i.e.:

$$Loss(\theta,\phi)=c_pL_{CLIP}(\theta)-c_vL_{MSE}(\theta)+c_eL_E(\theta) \tag{33}$$

where c_p is the coefficient of the strategy loss, c_v is the coefficient of the value function, and c_e is the coefficient of entropy.

2.3.2. PPO algorithm training process

Fig. 2 shows the network update method of the PPO algorithm, which combines the PPO algorithm with the DDPG so as to realize the economic scheduling of the multi-objective power system, and its specific training process is as follows:

(1) Input the environment information s_t into the new strategy network to obtain two values, and then construct a normal distribution by using these two values as the mean and variance, respectively, and sample the actions through this normal distribution. An action value is input into the environment to obtain a reward r and the next state s_{t+1}, and (s_t,a,r,s_{t+1}) is stored as a piece of scheduling experience. Then input s_{t+1} into the new policy network and loop the previous step until a certain number of scheduling experiences are stored.

(2) Input s_{t+1} obtained in the previous step into the evaluation network and compute the action value function $V_\pi(s)$.

(3) Input the stored scheduling experience in step 1 into the evaluation network, calculate the corresponding state value function $Q_\pi(s,a)$ according to Eq. Calculate the advantage function $A_\pi(s,a)$, and use the advantage function as the objective function to update the evaluation network parameters by back propagation.

(4) Input all stored state combinations into the new and old strategy networks to obtain the mean and standard deviation to construct the new and old strategy distributions, respectively. Input all stored combinations of actions into the new and old strategy distributions to obtain the probability of each action. Calculate the ratio of the corresponding probabilities.

(5) Calculate the objective function of the action network according to Eq. Then perform backpropagation to update the new strategy network.

(6) Repeat steps 4-5. After a certain step, the loop ends. Update the old strategy network using the new strategy network parameters.

(7) Repeat steps 1-6 training until convergence.

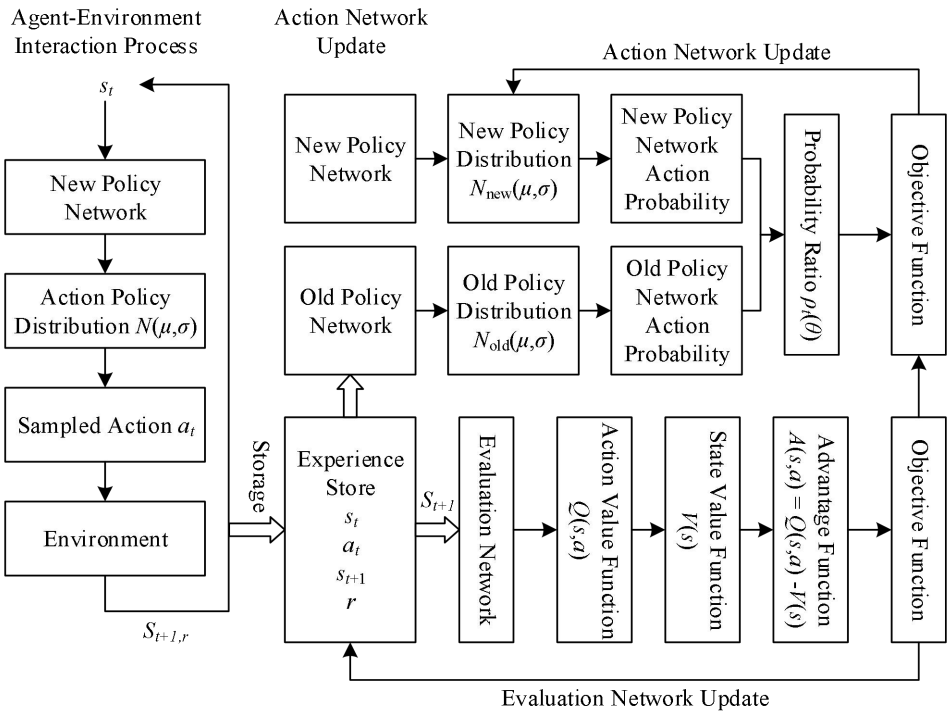


Figure 2. PPO algorithm network update mode

3. Analysis of experimental results

A new type of power system is a power system with the basic premise of ensuring energy and power security, with green power consumption as the main goal, with a strong smart grid as the hub platform, and with the support of source-network-load-storage interaction and multi-energy complementarity, which has the basic features of green and low-carbon, safe and controllable, open and interactive, digitally empowered, and economically efficient. With the goal of "carbon peak - carbon neutral", the proportion of new energy in the power energy supply is gradually increasing, and a new type of power system with a gradually increasing proportion of new energy will be formed. The wide access and rapid development of new energy make the stochasticity and uncertainty of the new power system increase significantly, which brings great challenges to the traditional scheduling optimization methods.

3.1. Power system dispatching simulation experiment setup

3.1.1. Scheduling simulation node system

In order to verify the effectiveness of the multi-objective power system scheduling model proposed in this paper in the objectives of reducing carbon emissions, lowering operating costs, and ensuring the security of power transmission, a simplified power system simulation scenario is established. The simulation system summarizes and simplifies the output of several decentralized wind farms to form a wind farm, and similarly summarizes and simplifies the output of decentralized photovoltaic power plants to form a photovoltaic power plant, an energy storage system, and two carbon capture power plants equipped with four thermal power units, which will not be shut down as long as they are activated during a scheduling cycle, in order to simplify the problem of eliminating the effect of the start and stop time of thermal power units.

The whole system is divided into two parts, a microgrid containing the electrical loads, which are also aggregated from several load sources, and the main grid, where power is exchanged between the two grids through a power transmission cross section. The wind and photovoltaic power generation is the total amount of power generated by wind farms and photovoltaic stations, respectively. The wind power data, photovoltaic data and electrical load data are taken from the European Power Data website, with the same time granularity and units, and the feature dimensions remain unchanged.

3.1.2. Experimental platform for scheduling simulation

In this paper, the Grid2op power system scheduling simulator is used as the training and evaluation platform for the model. In order to highlight the scheduling effect, this paper sets up the Do nothing baseline, i.e., just maintaining the initial grid topology during the multi-objective power system scheduling process and doing nothing else. Specifically, the Grid2op simulator defines the state space by quantitatively defining the scheduling of relevant elements in the power system, including transformer connections, generator outputs, line openings, and so on.

The state space includes discrete or continuous attribute values for each element, such as whether the load is satisfied, whether an overload condition occurs, generator output, etc. Also, the simulator defines as an action space the actions to be taken on each element in that state in order to achieve a stable state of the system, such as disconnecting or connecting the lines. In addition, the simulator also models the possible states of different elements in reality, such as the need for maintenance, cooling time, etc., in order to more realistically reproduce the process of multi-objective power system scheduling.

3.2. Experimental results of power system dispatching simulation

3.2.1. Comparative analysis of different scenarios

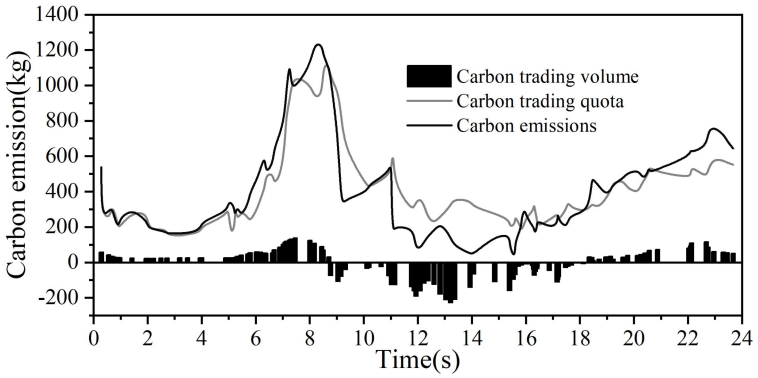
For the multi-objective power system scheduling model proposed in this paper, three different scenarios are set up in this paper, and the summer and winter test days are selected for testing, so as to compare the costs and carbon emissions of each scenario, analyze the economics of the multi-objective power system scheduling model, and provide support for exploring the low-carbon green economy scheduling.

Scenario A: Carbon trading mechanism is not considered in the integrated energy system scheduling, and the optimization objective does not consider carbon trading cost. Scenario B: In the integrated energy system scheduling consider pricing carbon trading mechanism, the optimization objective considers carbon trading cost. Scenario C: Consider stepped carbon trading mechanism in integrated energy system scheduling, and the optimization objective considers carbon trading cost.

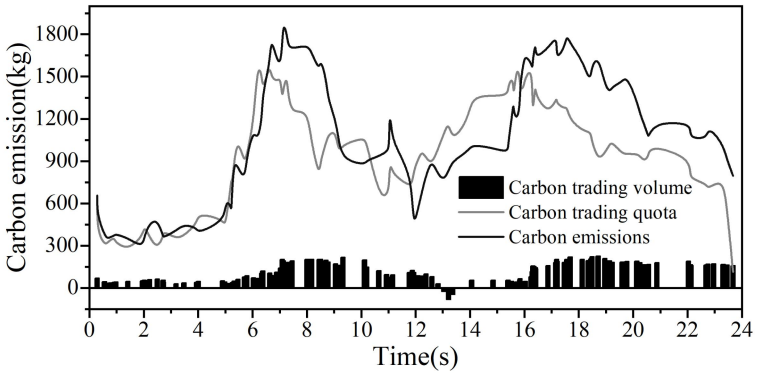
Based on the above three scenarios, the carbon trading scheduling results of the multi-objective power system are obtained as shown in Fig. 3, where Fig. 3(a)~(b) shows the carbon trading scheduling results in summer and winter, respectively. Table 1 shows the operation results of different scenarios.

As can be seen from Fig. 3, when the PV output is larger at noon, the system obtains more carbon allowances, and the actual carbon emissions at this time are smaller than the carbon allowances obtained. When the PV output is small and no output, the actual carbon emissions are larger than the carbon allowances obtained. The winter test day carbon emissions trading amount is 12.18t, the purchase of carbon allowances cost 3695.42 yuan, the summer test day carbon emissions trading amount is -5.05t, the sale of carbon allowances profit 1095.83 yuan.

From the experimental results in Table 1, it can be seen that the total cost of Scenario C in summer and winter is 22076.95 yuan and 73219.73 yuan, respectively, which is 16.70% less and 3.40% more than Scenario A. The total cost of Scenario C in summer and winter is 22076.95 yuan and 73219.73 yuan, which is 16.70% less and 3.40% more than Scenario A. The total cost of Scenario C in winter is 22076.95 yuan. The slightly lower operating cost of Scenario A in winter is due to the fact that the carbon emission constraint is not considered, i.e., at the cost of emitting 6.39 t more carbon to the environment. In summer, the algorithm proposed in this paper makes the system cost lower because the carbon emission is less than the carbon quota and the carbon trading amount is negative. The carbon emissions of Scenario C in summer and winter are 37.26t and 90.18t respectively, which are reduced by 6.48% and 6.62% compared with Scenario A, and 1.77% and 2.42% than Scenario B respectively. It can be seen that stepped carbon trading has a more obvious effect on reducing carbon emissions from the integrated energy system, and the effect is better than that of priced carbon trading. This, to a certain extent, provides scheduling strategy guidance for the low-carbon green economic dispatch of multi-objective power systems.



(a) Summer



(b) Winter

Figure 3. Carbon trading scheduling results of the power system

Table 1. Results of different scenarios

Scene		Total cost/yuan	Carbon trading cost/yuan	Carbon emissions /t
A	Summer	26503.38	-	39.84
	Winter	70815.27	-	96.57
B	Summer	22392.52	-936.48	37.93
	Winter	74348.69	2602.73	92.42
C	Summer	22076.95	-989.57	37.26
	Winter	73219.73	2454.69	90.18

3.2.2. Comparison of different scheduling strategies

In order to verify the effectiveness of the algorithm in this paper, the algorithm is written in the MATLAB simulation platform, and the MATPOWER software package is used to realize the tidal flow calculation process. In order to obtain a more optimal scheduling strategy, three scheduling strategies are set up in this paper. Strategy 1: Traditional centralized optimal scheduling strategy, i.e., main grid, distribution grid and microgrid are optimally scheduled together. Strategy 2: Decentralized coordinated optimal scheduling strategy, i.e., decentralized coordination of traditional power generation as a power source in the main grid, distribution grid, and microgrid. Strategy 3: Decentralized coordinated optimal scheduling strategy, which gives full consideration to traditional generation and renewable energy generation.

After separate tests, the scheduling results of the three strategies are shown in Table 2. As can be seen from the table, the cost difference between Strategy 1 and Strategy 2 is not significant regardless of the main network, distribution network, and microgrid, mainly because both strategies do not consider the participation of renewable energy generation, but only use the traditional generation to meet the load operation mode. Strategy 3 is the result with the participation of renewable energy generation, and the total cost is lower (73,576.37 yuan) due to the relatively low cost of renewable

energy generation, which consumes fewer conventional power sources in the main grid, distribution network and microgrid.

Table 2. Comparison of results from different scheduling strategies

Strategy	Cost (yuan)		
	Main net	**Distribution network**	**Microgrid**
1	51273.64	37512.46	23127.61
2	48135.79	35463.81	21516.89
3	35768.27	22165.39	15642.71

By further analyzing the operation process of the three different strategies, it is found that for the Type 1 strategy, the load in the distribution network is mainly satisfied by the main grid power supply. Similarly, for the type 2 strategy although distributed solution is used, renewable energy is not considered in the scheduling process, but micro and small gas engines are used, thus the load in the distribution network is mainly provided by the main grid power and micro and small gas engines. For the 3rd strategy, due to setting the penalty factor in the deep reinforcement learning method, i.e., the penalty factor is larger when discarding wind and PV, thus forcing the scheduling process to always use renewable energy generation, thus making the overall cost lower.

3.3. Scheduling results for the PPO-DDPG algorithm

3.3.1. Algorithm Training Change Curve

In this paper, the PPO algorithm is combined with DDPG with the aim of achieving optimal scheduling of multi-objective power systems. For the PPO algorithm, the dynamic economic dispatch is firstly transformed into a multi-stage Markov decision process by defining states, actions, reward functions, etc. In order to realize the PPO algorithm, the AC framework needs to be built in combination with deep neural networks. In the AC framework, this paper constructs a five-layer fully connected neural network as an action network to realize the output of the scheduling strategy. Every cycle of 12 scheduling cycles, the neural network is updated once.

In this paper, the dynamic economic dispatch model of multi-objective power system is constructed by using Python language, based on PARL, a reinforcement learning framework developed by Baidu Flying Paddle Paddle. For the offline tidal current calculation part, this paper utilizes C++ language programming to implement it. The model was trained 10000 times for 120,000 scheduling cycles, which took 10 hours and achieved good convergence results. Figure 4 shows the reward function and the penalty variation curve. The light gray curve in the figure shows the change in the average value of the reward function over 12 scheduling cycles, the black curve shows the change in the moving average of the reward function, and the gray curve demonstrates the change in the total penalty value.

As can be seen from the figure, during the training process, the PPO algorithm goes through a process of exploration followed by convergence. At the beginning of the model training, the actions attempted by the intelligent body often violate the system constraints, so the intelligent body is penalized (left side of the gray curve), resulting in a negative reward function for the environment feedback (left side of the black curve). As the number of interactions between the intelligent body and the environment increases, the intelligent body gradually explores and tries the scheduling strategies that do not violate the system constraints, and the neural network continuously updates the parameters according to the tried strategies, so that the total penalty value is getting smaller and smaller (the right side of the gray curve), and the reward function changes from a negative value to a positive value. At this point, the scheduling strategies that have satisfied the system constraints will continue to explore in the direction of economic optimization, and the value of the reward function can be seen growing in the figure. By the late stage of training, the intelligent body has mastered the scheduling strategy and can respond to different scheduling scenarios to give scheduling solutions. In the training curve, there is an up and down fluctuation in the reward function. This is due to the fact that in different scheduling scenarios, the load level and new energy output are different, and the economic cost itself is different. In addition, the unique exploration mechanism of reinforcement learning makes it possible to randomly explore the scheduling scheme that may bring rewards, but may also bring penalties, so the reward function will oscillate to some extent, which is a normal phenomenon. On the whole, the reward function in the training process shows a growing trend, and the training effect of the intelligent body gradually becomes better.

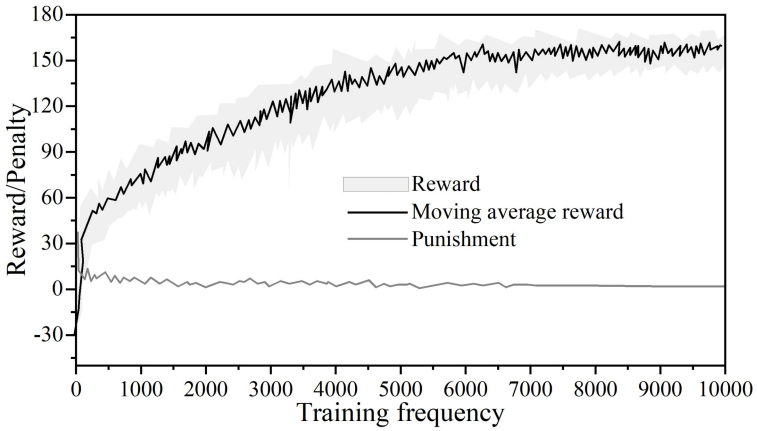
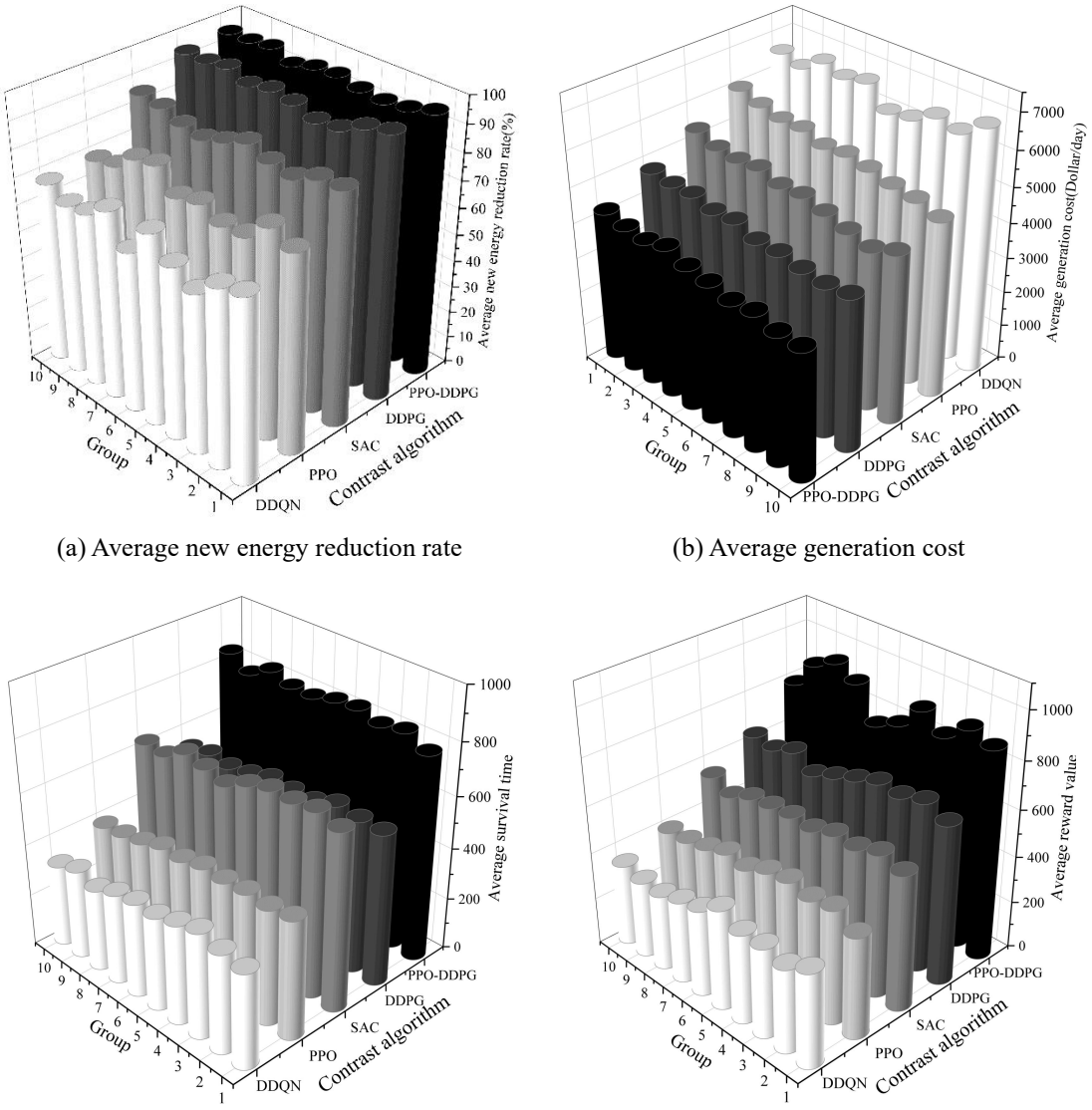


Figure 4. The changing curve of reward function and penalty

3.3.2. Comparison of algorithmic control performance

After the PPO algorithm is trained, based on the simulation example data given in the previous paper, 10 groups of scenarios are extracted from it and given to each intelligent body as inputs, and finally the mean values of the robustness and security indexes of the algorithmic intelligences' responses are calculated in each test group. In this paper, four algorithms, SAC, DQN, DDPG, and PPO, are selected as the comparison of PPO-DDPG algorithm in this paper, and the performance comparison indexes are designed for the economic optimization scheduling problem, which mainly include the average new energy consumption rate, the average power generation cost, the average survival time step, and the average reward value. Figure 5 shows the control performance comparison results of different algorithms, in which Figures 5(a)~(d) show the comparison results of average new energy consumption rate, average generation cost, average survival time step, and average reward value, respectively.

As can be seen from the figure, compared with the DDQN algorithm and the PPO algorithm, the SAC algorithm obviously has more superior performance in the training process. In terms of new energy consumption rate index, the new energy consumption rate of DDPG algorithm and PPO-DDPG algorithm after convergence is as high as about 95.5-97.5% or so, which is about 30% or more higher than that of the simple PPO algorithm, which indicates that the PPO-DDPG algorithm has the strongest performance in terms of environmental protection. In terms of the generation cost of the new power system, the PPO-DDPG algorithm can save about 1800 yuan /day after convergence compared to the PPO algorithm (the exact amount of savings is not necessarily accurate, and mainly serves as a comparison), which indicates that the PPO-DDPG algorithm has the strongest performance in terms of economy. In terms of the number of survival time steps, the PPO-DDPG algorithm survives slightly more than the DDPG algorithm by about 120 time steps after convergence, which is much higher than the DDQN algorithm and the PPO algorithm, which indicates the existence of a superiority of the PPO-DDPG algorithm in terms of operational robustness. In terms of reward value, the PPO-DDPG algorithm obtains the highest reward value after convergence, while the DDPG algorithm obtains a slightly higher reward value than the SAC algorithm after convergence, and all the reward values obtained by the PPO-DDPG algorithm are much higher than those of the DDQN algorithm and the PPO algorithm, which indicates that there exists a superiority in terms of the overall performance of the PPO-DDPG algorithm.



(a) Average new energy reduction rate

(b) Average generation cost

(c) Average survival time

(d) Average reward value

Figure 5. The control of different algorithms is compared

3.3.3. Analysis of optimized operation results

To further illustrate the effectiveness of the algorithm, the trained model is saved, and then the load data of commercial, residential and work areas in a park within a day are used as the test samples, and the operation results of commercial, residential and work areas are obtained as shown in Figs. 6, 7 and 8, respectively.

As shown in Figure 6, between 0:00 and 7:00, electricity prices are in the valley, meaning that costs are lower. The demand for electricity gradually increases during this period and the energy storage gradually increases, indicating that the storage system is actively purchasing electricity and storing energy. Between 7 and 9 o'clock, we see a decrease in power demand and a corresponding increase in storage status as the PV starts to come into play, suggesting that utilizing PV for energy storage is the preferred option. The power demand reaches two peaks between 10:00 and 14:00, and between 17:00 and 19:00, and the tariff curve rises during these hours, when the energy storage system releases energy to meet the high demand and reduce the cost of purchased power. During the other periods, especially when power demand falls, the energy storage state curve is lower than the other periods, indicating that the energy storage system discharges energy to utilize the cheaper power stored. This graph shows that the use of energy storage systems can smooth the power demand, especially during peak hours by utilizing storage discharge to reduce power purchases during peak hours at high electricity prices. At the same time, purchasing power and storing it during low electricity prices can reduce overall operating costs.

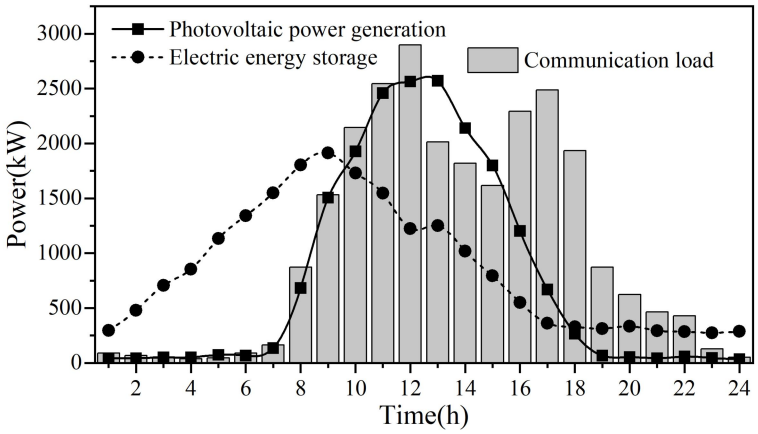


Figure 6. Commercial District Operational Results Chart

In Fig. 7, electricity prices are in the valley from early morning to 7 a.m., and the energy storage system charges during this period. The electricity price and demand is peak from 9 a.m. to 15 p.m., and the energy storage system discharges during this period to minimize power purchases under high electricity prices. During the evening hours when electricity prices and demand are lower, the energy storage system is again operated for charging. During the daytime, especially before the peak hours, PV is the main source of power supply, when the energy storage battery prioritizes the use of PV generation for charging. When the PV generation is not sufficient to meet the demand, the storage system purchases power from the main grid and starts the gas turbine to supplement the supply. During periods of reduced demand, additional PV power, if available, may be sold or used for energy storage.

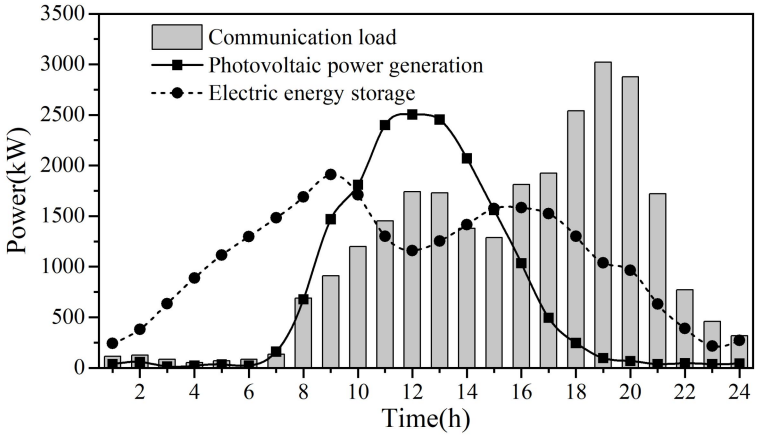


Figure 7. Residential Area Operational Results Chart

In Figure 8, the system purchases cheap electricity to charge during the nighttime to early morning hours. As electricity prices begin to rise in the morning, the system stops purchasing electricity, relies on PV generation and retains storage power. During midday and evening when electricity prices and demand peak, the system may use storage discharge to meet demand and reduce costs. During hours when demand drops but electricity prices remain high, gas turbines are brought into service to provide additional power and avoid purchasing high-priced electricity. With such dynamic regulation, the system can more effectively balance supply and demand, optimizing operating costs while improving energy efficiency.

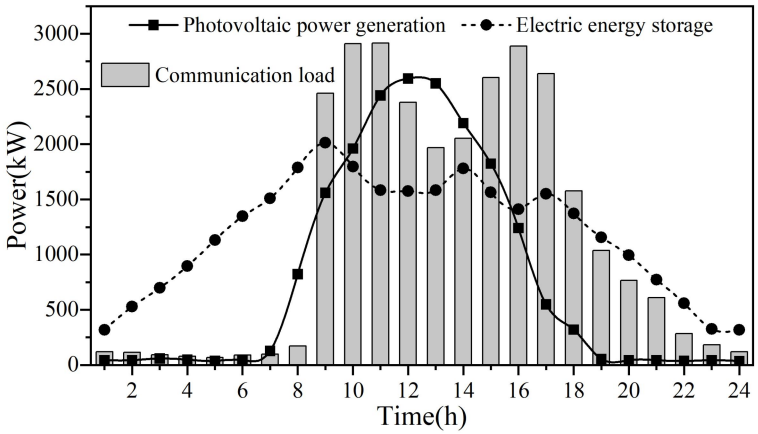


Figure 8. Work Area Operational Results Chart

4. Conclusion

Based on the optimal scheduling environment of the energy internet, a proximal policy optimization algorithm (PPO-DDPG) combined with deep reinforcement learning is proposed to address the problems of uncertainty, multi-objective conflicts and constraints faced in power system scheduling. The method constructs a dynamic state encoder by introducing a deep reinforcement learning model, which improves the ability of the strategy to perceive the time-varying characteristics of the system, and designs a multi-objective hierarchical rewarding mechanism, which realizes the adaptive trade-off between wind power consumption and operation cost under the premise of ensuring power balance and security. Combined with the experimental results, it can be seen that the PPO-DDPG algorithm significantly outperforms the existing methods in terms of new energy consumption rate, average operating cost, and average reward value in multi-objective power system scheduling, and it has good comprehensive scheduling performance and prospects for popularization and application.

Future research will further combine federal learning and multi-intelligent body framework, explore the mechanism of multi-region cooperative scheduling and cross-site policy migration, and introduce the attention mechanism to improve the state characterization ability and decision-making interpretability, so as to promote the actual landing and highly reliable deployment of reinforcement learning technology in intelligent power systems.

References

1. Wang, K., Yu, J., Yu, Y., Qian, Y., Zeng, D., Guo, S., ... & Wu, J. (2017). A survey on energy internet: Architecture, approach, and emerging technologies. IEEE systems journal, 12(3), 2403-2416.
2. Kong, X., Zhao, X., Wang, C., Duan, Q., Sha, G., & Liu, L. (2022). Promote the international development of Energy Internet technology standards based on key competition mode. Sustainable Cities and Society, 86, 104151.
3. Feng, C., & Liao, X. (2020). An overview of "energy+ internet" in China. Journal of Cleaner Production, 258, 120630.
4. Wang, K., Hu, X., Li, H., Li, P., Zeng, D., & Guo, S. (2017). A survey on energy internet communications for sustainability. IEEE Transactions on Sustainable Computing, 2(3), 231-254.
5. Li, J., Shang, Z., Qiang, R., Pang, J., Guo, H., Wang, J., & Niu, H. (2021). Energy Internet Security Risk Evaluation Index System. In IOP Conference Series: Earth and Environmental Science (Vol. 645, No. 1, p. 012045). IOP Publishing.
6. Cheng, L., & Yu, T. (2019). Smart dispatching for energy internet with complex cyber-physical-social systems: A parallel dispatch perspective. International Journal of Energy Research, 43(8), 3080-3133.
7. Shi, X., Wen, G., Cao, J., & Yu, X. (2018). Model predictive power dispatch and control with price-elastic load in energy internet. IEEE Transactions on Industrial Informatics, 15(3), 1775-1787.
8. Mohammadi, M., Kavousi-Fard, A., Dabbaghjamanesh, M., Farughian, A., & Khosravi, A. (2021). Effective management of energy internet in renewable hybrid microgrids: A secured data driven resilient architecture. IEEE Transactions on Industrial Informatics, 18(3), 1896-1904.
9. Mostafaeipour, A., Bidokhti, A., Fakhrzad, M. B., Sadegheih, A., & Mehrjerdi, Y. Z. (2022). A new model for the use of renewable electricity to reduce carbon dioxide emissions. Energy, 238, 121602.
10. Mahmud, U. P. A. L., Alam, K. H. O. R. S. H. E. D., Mostakim, M. A., & Khan, M. S. I. (2018). AI-driven micro solar power grid systems for remote communities: Enhancing renewable energy efficiency and reducing carbon emissions. Distributed Learning and Broad Applications in Scientific Research, 4.
11. Nassar, I. A., Hossam, K., & Abdella, M. M. (2019). Economic and environmental benefits of increasing the renewable energy sources in the power system. Energy Reports, 5, 1082-1088.

12. Bessa, R., Moreira, C., Silva, B., & Matos, M. (2019). Handling renewable energy variability and uncertainty in power system operation. Advances in Energy Systems: The Large-scale Renewable Energy Integration Challenge, 1-26.
13. Cole, W., Gates, N., & Mai, T. (2021). Exploring the cost implications of increased renewable energy for the US power system. The Electricity Journal, 34(5), 106957.
14. Holjevac, N., Baškarad, T., Đaković, J., Krpan, M., Zidar, M., & Kuzle, I. (2021). Challenges of high renewable energy sources integration in power systems—the case of croatia. Energies, 14(4), 1047.
15. Mararakanye, N., & Bekker, B. (2019). Renewable energy integration impacts within the context of generator type, penetration level and grid characteristics. Renewable and Sustainable Energy Reviews, 108, 441-451.
16. Li, X., Li, Y., Tan, Q., & Zhang, S. (2025). Bi-objective robust dispatch model for virtual power plant with security-economy-green equilibrium in the new-type power system. Energy Conversion and Management, 344, 120253.
17. Chen, X. U. E., Jing, R. E. N., Peng, W. A. N. G., Xin, Z. H. O. U., & Ya, L. I. U. (2021, March). An optimal dispatch method for high proportion new energy power grid based on source-network-load-storage interaction. In 2021 4th International Conference on Electron Device and Mechanical Engineering (ICEDME) (pp. 119-122). IEEE.
18. Guo, Y., Ming, B., Huang, Q., Wang, Y., Zheng, X., & Zhang, W. (2022). Risk-averse day-ahead generation scheduling of hydro–wind–photovoltaic complementary systems considering the steady requirement of power delivery. Applied Energy, 309, 118467.
19. Li, Y. Z., Li, K. C., Wang, P., Liu, Y., Lin, X. N., Gooi, H. B., ... & Luo, Y. (2017). Risk constrained economic dispatch with integration of wind power by multi-objective optimization approach. Energy, 126, 810-820.
20. Zhou, J., Wang, C., Li, Y., Wang, P., Li, C., Lu, P., & Mo, L. (2017). A multi-objective multi-population ant colony optimization for economic emission dispatch considering power system security. Applied Mathematical Modelling, 45, 684-704.
21. Omar, A. I., Ali, Z. M., Al-Gabalawy, M., Abdel Aleem, S. H., & Al-Dhaifallah, M. (2020). Multi-objective environmental economic dispatch of an electricity system considering integrated natural gas units and variable renewable energy sources. Mathematics, 8(7), 1100.
22. Yin, L., & Sun, Z. (2021). Multi-layer distributed multi-objective consensus algorithm for multi-objective economic dispatch of large-scale multi-area interconnected power systems. Applied Energy, 300, 117391.
23. Bai, Y., Wu, X., & Xia, A. (2021). An enhanced multi-objective differential evolution algorithm for dynamic environmental economic dispatch of power system with wind power. Energy Science & Engineering, 9(3), 316-329.
24. Hassan, M. H., Kamel, S., Domínguez-García, J. L., & El-Naggar, M. F. (2022). MSSA-DEED: A multi-objective salp swarm algorithm for solving dynamic economic emission dispatch problems. Sustainability, 14(15), 9785.
25. Dai, H., Huang, G., & Zeng, H. (2023). Multi-objective optimal dispatch strategy for power systems with Spatio-temporal distribution of air pollutants. Sustainable Cities and Society, 98, 104801.
26. Zhang, Z., Zhang, H., Tian, Y., Li, C., & Yue, D. (2024). Cooperative constrained multi-objective dual-population evolutionary algorithm for optimal dispatching of wind-power integrated power system. Swarm and Evolutionary Computation, 87, 101525.
27. Cheng, C., Fang, Y., Wang, J., & Peng, C. (2025). A new multi-objective human learning algorithm for environmental-economic dispatch of power systems. Electric Power Systems Research, 246, 111687.
28. Deng, B., Li, M. S., Ji, T. Y., & Wu, Q. H. (2025). Learning-based stochastic multi-objective optimizer for uncertain power system scheduling. Applied Soft Computing, 113402.
29. Jiang, H., Du, E., He, B., Zhang, N., Wang, P., Li, F., & Ji, J. (2023). Analysis and modeling of seasonal characteristics of renewable energy generation. Renewable Energy, 219, 119414.
30. Yan, J., Qu, T., Han, S., Liu, Y., Lei, X., & Wang, H. (2020). Reviews on characteristic of renewables: Evaluating the variability and complementarity. International transactions on electrical energy systems, 30(7), e12281.
31. Mouassa, S., & Bouktir, T. (2019). Multi-objective ant lion optimization algorithm to solve large-scale multi-objective optimal reactive power dispatch problem. COMPEL-The international journal for computation and mathematics in electrical and electronic engineering, 38(1), 304-324.
32. Hossain, M. T., Hossain, M. A., & Adnan, M. A. (2024, December). A Confidentiality-Preserving Distributed Linear Programming Model for Solving Large-Scale Economic Dispatch Problems. In Proceedings of the 11th International Conference on Networking, Systems, and Security (pp. 8-15).
33. Nghitevelekwa, K., & Bansal, R. C. (2018). A review of generation dispatch with large-scale photovoltaic systems. Renewable and sustainable energy reviews, 81, 615-624.
34. Pape, M., & Kazerani, M. (2020). Turbine startup and shutdown in wind farms featuring partial power processing converters. IEEE Open Access Journal of Power and Energy, 7, 254-264.
35. Scarabaggio, P., Carli, R., & Dotoli, M. (2022). Noncooperative equilibrium-seeking in distributed energy systems under ac power flow nonlinear constraints. IEEE transactions on control of network systems, 9(4), 1731-1742.
36. Yuan, G., & Yang, W. (2019). Study on optimization of economic dispatching of electric power system based on Hybrid Intelligent Algorithms (PSO and AFSA). Energy, 183, 926-935.
37. Liu, Z., Qian, R., Jin, X., Zhao, H., Li, H., Hu, D., & Hu, H. (2024). Multi-source heterogeneous data fusion

technology for electric power based on big data mining. Journal of Computational Methods in Sciences and Engineering, 24(6), 3366-3380.
38. Miraftabzadeh, S. M., Longo, M., & Brenna, M. (2023). Knowledge extraction from PV power generation with deep learning autoencoder and clustering-based algorithms. IEEE Access, 11, 69227-69240.
39. Fang, D., Guan, X., Hu, B., Peng, Y., Chen, M., & Hwang, K. (2020). Deep reinforcement learning for scenario-based robust economic dispatch strategy in internet of energy. IEEE internet of things journal, 8(12), 9654-9663.
40. Yang, J., Liu, J., Xiang, Y., Zhang, S., & Liu, J. (2022). Data-driven optimal dynamic dispatch for hydro-PV-PHS integrated power systems using deep reinforcement learning approach. CSEE Journal of Power and Energy Systems, 9(3), 846-858.
41. Tang, H., Lv, K., Bak-Jensen, B., Pillai, J. R., & Wang, Z. (2022). Deep neural network-based hierarchical learning method for dispatch control of multi-regional power grid. Neural Computing and Applications, 34(7), 5063-5079.
42. Hu, J., Ye, Y., Tang, Y., & Strbac, G. (2023). Towards risk-aware real-time security constrained economic dispatch: A tailored deep reinforcement learning approach. IEEE Transactions on Power Systems, 39(2), 3972-3986.
43. Ma, A., Li, Z., Shen, F., Peng, X., Liu, Y., Zhong, W., & Qian, F. (2025). Multi-objective dispatch of integrated renewable power systems leveraging robust optimization in deep reinforcement learning. Computers & Chemical Engineering, 109173.
44. Yin, L., & Ding, W. (2025). Dual deep neural networks-accelerated non-dominated sorting moth flame optimizer for distributed multi-objective economic dispatch. Expert Systems with Applications, 259, 125259.
45. Yin, L., Ye, Y., & Zhang, X. (2025). Deep Learning-based Approach for Accelerated Economic Dispatch in Hierarchical Distributed Power Systems with Internet of Things. IEEE Internet of Things Journal.
46. Huo, X., & Wang, X. (2025). A Deep Learning-Driven Bidirectional Power Dispatch Optimization Framework for Smart Grids Using IoT Sensing Data. Informatica, 49(26).
47. Zhu, R., Guan, X., Zheng, J., Wang, N., Jiang, H., Cui, C., & Ohtsuki, T. (2023). DRL based low carbon economic dispatch by considering power transmission safety limitations in internet of energy. Internet of Things, 24, 100979.
48. Jiaying Wang,Xiaoqian Meng,Xuan Yang,Haibing Yin & Pingkai Fang. (2025). Multi-objective energy-efficient power system scheduling using Stochastic State Space Model and reinforcement learning. Sustainable Computing: Informatics and Systems,48,101224-101224.https://doi.org/10.1016/J.SUSCOM.2025.101224.
49. Haifeng Zhang,Yifu Zhang,Jiajun Zhang,Xiangdong Meng & Jiazu Sun. (2025). Resilient dispatching optimization of power system driven by deep reinforcement learning model. Discover Artificial Intelligence,5(1),189-189.https://doi.org/10.1007/S44163-025-00451-1.
50. Aoqun Ma,Zhi Li,Feifei Shen,Xin Peng,Yurong Liu,Weimin Zhong & Feng Qian. (2025). Multi-objective dispatch of integrated renewable power systems leveraging robust optimization in deep reinforcement learning. Computers and Chemical Engineering,201,109173-109173.https://doi.org/10.1016/J.COMPCHEMENG.2025. 109173.
51. Yang Jingxian,Liu Junyong,Qiu Gao,Liu Jichun,Jawad Shafqat & Zhang Shuai. (2023). A spatio-temporality-enabled parallel multi-agent-based real-time dynamic dispatch for hydro-PV-PHS integrated power system. Energy,278(PB),https://doi.org/10.1016/J.ENERGY.2023.127915.